Knowledge Engineering for Interoperable
Electronic Health Records
Daniela Sofia Rijo Oliveira
UMinho | 2022

Universidade do Minho
Escola de Engenharia

Daniela Sofia Rijo Oliveira

Knowledge Engineering for Interoperable Electronic Health Records

March 2022

Universidade do Minho
Escola de Engenharia

Daniela Sofia Rijo Oliveira

Knowledge Engineering for Interoperable Electronic Health Records

Doctoral Thesis
Doctorate in Biomedical Engineering

Work supervised by
José Manuel Ferreira Machado
António Carlos da Silva Abelha

March 2022

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the Universidade do Minho.

Braga, 10th March 2022

(Location) (Date)

Assinado por : **DANIELA SOFIA RIJO DE OLIVEIRA**
Num. de Identificação: BI143636324
Data: 2022.03.10 16:10:12 +0000

(Daniela Sofia Rijo Oliveira)

As for the future, your task is not to foresee it, but to enable it.

Antoine de Saint Exupery

Acknowledgements

It's true! I have achieved another very important milestone in my life and I have so much to be thankful for!

To begin with, I would like to express my profound gratitude to Professor José Machado for his continuous encouragement during my academic journey and my research project, as well as for his patience, enthusiasm, and vast knowledge-sharing. His counsel supported me throughout the research and writing phases of this thesis.

In addition, I'd like to thank my second supervisor, Professor António Abelha, for his constant support throughout this journey, as well as for all of the constructive discussions we've had over the last four years, both professionally and personally, which have helped me to grow as a professional and as a person.

Apart from my advisers, I would like to express my gratitude to the PCE team at CHUP institution for four years of collaborative work, particularly to nurse Nuno Abreu and doctor Pedro Leuschner. Without their daily cooperation and clinical expertise, conducting this research would be much more difficult.

My sincere gratitude also goes out to my KEG lab mates for their suggestions, assistance, inspiration, and good times shared.

My heartfelt and deep gratitude to all my family and my closest friends for their unconditional support throughout this thesis, which culminated four years of rigorous effort. Mom and dad, I am so sorry for my obvious absence as I worked on my thesis, and I want to express my gratitude for all the encouraging words and motivation that you always have provided me.

And to my boyfriend, Rui, I want to express my deep gratitude for his unconditional love, support, patience, and encouragement, without which I would have given up on my thesis long ago. You have been amazing, and I will now clear all clutter from the kitchen table and chairs, as you have requested multiple times!

Abstract

Knowledge Engineering for Interoperable Electronic Health Records

The demand for more efficient healthcare delivery has prompted hospital institutions to become increasingly concerned with the evolution of their information systems over time, pursuing increased efficiency and interoperability among them on a global scale. Due to the large degree of variety in healthcare, an increasing number of methods and standards for standardized information and establishing critical levels of interoperability have emerged. The need of using standards to collect as much structured data as possible has become increasingly evident as a result of unstructured data's inability to assist new knowledge discovery and the development of clinical decision support systems.

The abundance of healthcare software, typically as the consequence of lengthy and complex development, has aided some systems while hindering others. Nonetheless, the goal of this study is to create an artefact that enables the building of new healthcare information systems while preserving clinical and ethical validations, based on a novel paradigm. This new paradigm is guided by the openEHR standard, which reduces the concept of component-oriented software development and optimizes data structuring and centralization procedures in an ever-changing clinical environment.

It was essential to conduct a needs assessment of various Portuguese healthcare institutions using research and requirements gathering techniques in order to design and develop an artefact that would fill those gaps. The research results were extremely promising, resulting in the development of a new demographic and clinical systems, as well as a new platform for managing the structures necessary to ensure the system's operability. Additionally, the case studies conducted revealed that, although at different stages of maturity, the system was successfully implemented and accepted by the users.

Keywords: Electronic Health Record, Healthcare Information Systems, Interoperability, Multi-Agent Systems, OpenEHR

Resumo

Engenharia de Conhecimento para Registos Clínicos Electrónicos Interoperáveis

A necessidade de uma prestação de cuidados de saúde cada vez mais eficiente levou as instituições hospitalares a preocuparem-se cada vez mais com a evolução dos seus Sistemas de Informação ao longo do tempo, procurando a máxima eficiência e interoperabilidade possivel, em todo o mundo. Devido à sua elevada heterogeneidade, têm surgido enumeros métodos e normas para a normalização de informação e estabelecimento de níveis cruciais de interoperabilidade na saúde. A importância da sua utilização revela-se na máxima recolha possível de informação estruturada, devido à inutilização de dados não estruturados face à descoberta de novo conhecimento e ao desenvolvimento de sistemas de apoio à decisão clínica.

A abundância de *software* de cuidados de saúde, normalmente provenientes de um desenvolvimento demoroso e complexo, tem ajudado alguns sistemas ao mesmo tempo que tem dificultado outros. Assim sendo, o objetivo deste estudo é desenvolver um artefacto que permita a construção de novos sistemas de informação de saúde, preservando as validações clínicas e éticas, com base num novo paradigma. Este novo paradigma baseia-se no *standard* openEHR, reduzindo assim o conceito de desenvolvimento de *software* orientado ao componente, otimizando ainda a estruturação e processos de centralização de dados num ambiente clínico em constante mudança.

Foi essencial avaliar as necessidades existentes em diferentes instituições de saúde portuguesas através da utilização de técnicas de investigação e levantamento de requisitos, para projetar e desenvolver um artefacto que preenchesse tais lacunas. Os resultados foram extremamente promissores, resultando num novo sistema demográfico e clínico, bem como numa nova plataforma de gestão de estruturas necessárias para garantir a coerência de todo o sistema. Os casos de estudo efetuados revelaram que, embora em diferentes fases de maturação, o sistema foi implementado com sucesso e aceite pelos utilizadores.

Palavras-chave: Interoperabilidade, OpenEHR, Registo Clínico Eletrónico, Sistemas de Informação na Saúde, Sistemas Multiagentes

Contents

List of Figures

List of Tables

Listings

Acronyms

ADL	Archetype Definition Language
AI	Artificial Intelligence
AM	Archetype Model
API	Application Programming Interface
AQL	Archetype Query Language
ASAC	Archetype-based metric Software Architecture Coupling
BI	Business Intelligence
BPMN	Business Process Model and Notation
CDA	Clinical Document Architecture
CDM	Clinical Decision Model
CDS	Clinical Decision Support
CEN	European Committee for Standardization
CHD	Computable Health Data
CHTS	*Centro Hospitalar do Tâmega e Sousa*
CHUP	*Centro Hospitalar Universitário do Porto*
CIM	Clinical Information Model
CKM	Clinical Knowledge Manager
CRISP-DM	Cross Industry Standard Process for Data Mining
CS	Computer Science
CW	Collaborative Work
DAI	Distributed Artificial Intelligence
DES	Data Entry System
DHSA	Decoupled Healthcare Software Architecture
DICOM	Digital Imaging and Communications in Medicine

DIM	Demographic Information Model
DL	Deep Learning
DLM	Decision Logic Module
DM	Data Mining
DMM	Data Mining Models
DMT	Data Mining Techniques
DS	Data Science
DSR	Design Science Research
DSS	Decision Support System
DT	Decision Tree
DW	Data Warehouse
EC	European Commission
ECG	Electrocardiogram
EHR	Electronic Health Record
EMR	Electronic Medical Record
ETL	Extraction, Transformation and Loading
FHIR	Fast Healthcare Interoperability Resources Specification
GDL	Guideline Definition Language
GECCO	German Corona Consensus Dataset
GUI	Graphical User Interface
HIMSS	Healthcare Information and Management Systems Society
HIS	Healthcare Information Systems
HIT	Healthcare Information Technology
HL7	Health Level Seven
HP	Healthcare Professionals
HTTP	Hypertext Transfer Protocol
ICD	International Classification of Diseases
IEEE	Institute of Electrical and Electronics Engineers
IES	Intelligent Embedded System

IM	Information Model
IS	Information System
ISO	International Organization for Standardization
ISO EN 13606	International Organization for Standardization European Norm 13606
IT	Information Technology
ITIL	Information Technology Infrastructure Library
JDT	JSON Data Template
JSON	JavaScript Object Notation
KDD	Knowledge-Discovery in Databases
KPI	Key Indicators of Performance
LOINC	Logical Observation Identifiers Names and Codes
LS	Legacy System
MAS	Multi-Agent System
MHR	Manual Health Record
MI	Medical Informatics
ML	Machine Learning
MS	Mathematical Statistics
MSS	Management Support System
NB	Naive Bayes
NB-K	Naive Bayes - Kernel
NER	Named Entity Recognition
NFC	Near Field Communication
NLP	Natural Language Processing
NoSQL	Non-relational database
OLTP	OnLine Transaction Processing
OPT	Operational Templates
ORM	Object-relational Mapper

PACS Picture Archiving Communication System
PoC Proof of Concept
PROC Process model

QBE Query By Example

RDBMS Relational database management system
REST Representational State Transfer
RF Random Forest
RIM Reference Information Model
RIS *Rede Informática da Saúde*
RM Reference Model
RNU *Rede Nacional de Utentes*
RQ Research Question
RSE *Registo de Saúde Eletrónico*
RT Random Tree

SGTD *Sistema de Gestão do Transporte de Doentes*
SIGLIC *Sistema Integrado de Gestão de Inscritos para Cirurgia*
SM Service Model
SNOMED CT Systematized Nomenclature of Medicine Clinical Terms
SNS *Sistema Nacional de Saúde*
SOAP Simple Object Access Protocol
SODA Simple Oracle Document Access
SQL Structured Query Language
SWOT Strengths, Weaknesses, Opportunities, and Threats

TAM Technology Acceptance Model
TAM2 Technology Acceptance Model II
TAM3 Technology Acceptance Model III
TRA Theory of Reasoned Action

UC Ubiquitous Computing
UI User Interface

ULSCB *Unidade Local de Saúde de Castelo Branco*
ULSG *Unidade Local de Saúde da Guarda*
ULSNA *Unidade Local de Saúde do Norte Alentejo*
UML Unified Modeling Language
UMLS Unified Medical Language System

WAP Wireless Application Protocol
WHO World Health Organization

XML Extensible Markup Language

Chapter 1

Introduction

The research project described in this document is presented in the form of a doctoral thesis entitled "Knowledge Engineering for Interoperable Electronic Health Records" within the scope of the Doctoral Program in Biomedical Engineering at the University of Minho. This investigation was conducted out at the Algoritmi Research Center, a research unit in the University of Minho's School of Engineering, as well as in Portuguese healthcare institutions. As a result, this introduction chapter comprises a description of the project's scope and context (Section 1.1), an explanation of the purpose behind it (Section 1.2), a list of objectives to be met as a result of its fulfillment (Section 1.3) and a document structure description (Section 1.4).

1.1 Context and Scope

Clinical data in Healthcare Institutions are highly important and useful for several purposes. Data are filled and stored on many Healthcare Information Systems (HIS) developed for different use cases. Only with an optimized cooperation and interoperability between multiple HIS, it is possible to ensure the continuous improvement and maintenance of the Electronic Health Record (EHR).

Over the past few years, a significant amount of work has been developed to improve the communication between different Information System (IS) across multiple domains, lowering costs and usual human errors, improving the efficiency, performance and security of systems in institutions and companies [50]. Thereby, there was an evolution of the IS in different periods starting in the late 1960s. In the first Era, one of the concepts that growth was the Management Support System (MSS) in order to support the organizations on decision-making processes. Afterwards, in the 1970s - second Era - these MSS became more independent, providing a daily assistance to their business and generating the concept of Decision Support System (DSS). Moreover, in the 80s Artificial Intelligence (AI) has been integrated in the Information Technology (IT) environment marking also the creation of more advanced and self-operated IS capable to support more complex decision-making processes, increasing previous results. Thenceforward, concepts such as Ubiquitous Computing (UC), Wireless Application Protocol (WAP) and Near Field Communication (NFC) appeared to transform the world of technology [104]. All these concepts were applied in the most diverse areas, particularly in the clinical environment, estimating that the adoption of IT in healthcare in order to result in a potential health advantages and also possible cost savings, creating the Healthcare Information Technology (HIT) concept. Since its definitive implementation, the EHR has the potential to save over 81 billion dollars per year. In addition, the efficiency and security improvements associated with the care plans and the patients are enormous [11].

Aside from the significant evolution of the Manual Health Record (MHR) to the EHR, the clinical information digitalization has resulted in creation of extravagant amounts of data every day, stored in dedicated repositories, and afterwards used or not. In this context, multidisciplinary teams in the fields of IS and Medical Informatics (MI) have arisen in healthcare institutions, particularly the biomedical engineer role specializing in MI has proven to be crucial in data control, developing tools to use them and managing healthcare institution's HIS, in order to provide a complete symbiosis and interoperability between all systems [15]. Therefore, the research in this area has opened new horizons in recent years, through innovative techniques and solutions to reach more and more in terms of the HIS efficient use, employing the new clinical standards used worldwide and adopting new methodologies for its ergonomic use.

The continuous marathon for data structuring contributes to the data within plain text gaining meaning and being transformed into relevant knowledge in a specific domain. In the healthcare field, a lot of information present in diaries and clinical reports written in an unstructured way, can often be lost knowledge about a specific clinical case. Furthermore, nowadays many clinical encounters and procedures still occur in an isolated way, making it difficult to access to a more patient information in real time, which can negatively influence the decision-making [87]. To prevent this from happening, a preventive and collaborative approach to develop the used HIS by healthcare professionals is necessary.

Thus, this manuscript is intended to describe the research project about a new interoperable clinical system based on open specifications and clinical models suggested by the standard called openEHR. This project was developed in collaboration with multidisciplinary teams from Portuguese healthcare institutions, made up from IT specialists to doctors and nurses. The main goal of this project is to develop an interoperable and secure system capable of having a hybrid use. The paradigm for its conception was to ***develop a system that can be modeled and adaptable to the user***, ensuring the data structuring and its preservation always in the same way. The novel system ensures the clinicians liberty to modeling the several openEHR structures and introduce them in a new platform responsible for the interpretation of these structures and its conversion to a standard format comprehended by new artefact. Also, the system is responsible for the parallel structures creation and maintenance such as clinical terminologies subsets or internal relevant data to append to the openEHR structures, establishing semantic interoperability. In addition, the recovery, maintenance and transformation of data is carried out by a retrieval system, capable of creating, organizing and maintaining a demographic model and a clinical model, separately, in a documental way, providing the interoperability with all others existing Legacy System (LS). These features will provide a security level to the system because one clinical record is not directly linked to a particular patient. The next chapters will demonstrate how these characteristics and modules were achieved, developed and implemented. I believe that this work and all the projects carried out in this specific area will give more efficiency, dynamism and autonomy to the use of HIS and, consequently, to the clinical practice.

1.2 Motivation

Within healthcare institutions, information exchange and interpretation in the same way among heterogeneous HIS is something that is vital for promoting the high level of quality and safety that the clinical act requires. In this effort, the main goal is to develop a new interoperable, adaptable, collaborative, and hybrid HIS that will make it possible to continuously enhance the healthcare services given while also discovering new relevant knowledge to support decision-making on a continual basis. From the perspective of work management, this unique approach will be able to stimulate collaboration between IT specialists and Healthcare Professionals (HP). This will be achievable because the system will not be completely reliant on information technology and physicians will have the option to represent their clinical knowledge in a consistent and evolutionary manner. From the perspective of information management, each EHR will be intelligently generated and constructed over the course of a patient's lifespan and will be available for consultation at any time during that time. This new artefact will also be scalable and responsive at the system architecture level, giving all information in the same way for all other LS and news platforms that wish to make use of this new artefact. Apart from that, from the standpoint of the user, clinicians will be able to dynamically select how they wish to see the patient's information structured in real time. In terms of record unification and globalization, the use of internationally recognized standards is becoming increasingly important in order to achieve the highest levels of inter- and intra-hospital interoperability possible, both nationally and internationally, in order to achieve the highest levels of interoperability possible. In this case, data normalization aids in the better interpretation of information among different HP types, departments, and institutions. By adopting internationally recognized and continually evolving standards, institutions are able to develop with them and continue to improve as both care providers and organizations.

1.3 Objectives

An open *Healthcare Information System* artefact is the purpose of this PhD thesis, which will be designed and developed as part of this project. As a result, this ground-breaking solution demonstrates how to quickly and easily develop and install a new clinical system while retaining the existing LS and exchanging information in a secure and clear manner. It's the goal of this doctoral thesis to design and construct a new open HIS **artefact**, which will be detailed later in this thesis. Consequently, this novel technique describes how to easily create and implement a new HIS while simultaneously maintaining the other current systems and exchanging information in an interpretable and secure manner, as described in this thesis. According to the preceding sections, it is intended to propose a new archetype that responds to important issues encountered by IT in the medical industry. As a result, the primary Research Question (RQ) that informs this doctoral thesis is as follows:

How to define the structuring, data exchange and knowledge discovery processes in a healthcare context, creating an interoperable Electronic Health Record (EHR)?

Based on the results of an initial requirements survey conducted out in some Portuguese healthcare institutions, it was discovered that there were gaps and inconsistencies in the structure, as well as non-universal content in data. It was also discovered that free text fields were used very frequently, which translates as being out of date for use, transformation, and data exchange. Aside from that, there are several IS, as well as a large variety of databases and versions, all of which require daily maintenance and monitoring by IT professionals. Following the major RQ, this research project can be defined by the following main objectives that must be met in order for it to be successfully completed:

1. A state of the art literature evaluation within the scope of this doctoral thesis;
2. Research Problems with HIS in Portuguese Healthcare Institutions;
3. Identifying the best research methodology and technologies to be used;
4. Proposal system's architecture design;
5. Relational and Non-relational database design;
6. Business Intelligence (BI) indicators and Machine Learning (ML) algorithms definition;
7. System development based on aforementioned points;
8. Developed archetype simulation, deployment, assessment, and implementation;
9. Artefact Proof of Concept (PoC);

10. System sharing to the target customers and the scientific community.

The six-step project represented in Figure 17 was designed to achieve all of the objectives listed above, and each stage has a number of questions that must be addressed.

Figure 1: The project's six distinct stages representation.

According to the above schema, each stage contemplates some secondary RQ that will be described below and answered over the next few chapters.

- **Stage I:**

  RQ1: What is the importance of structured data at the HIS level?

  RQ2: What are the reasons for using globally recognized standards?

  RQ3: Is the openEHR recommended approach appropriate for designing and building interoperable healthcare systems?

- **Stage II:**

  RQ4: What are the most suitable methodologies to conduct this research?

  RQ5: What are the most appropriate technologies for the development of this artefact?

  RQ6: What will be the best architecture for the novel system?

- **Stage III:**

  RQ7: What are the major challenges encountered throughout the development of the proposed solution? How were these obstacles overcome?

- **Stage IV:**

  RQ8: What are the biggest challenges associated with testing and implementing the proposed solution? How were these overcome?

- **Stage V:**

  RQ9: Throughout the usability testing of the system, what were the biggest difficulties experienced by users?

RQ10: Do the case studies intend to continue using the proposed system in the future?

- **Stage VI:**

RQ11: How will the proposed system improve the quality of health care provided?

RQ12: When compared to other similar existing projects, why is the research study considered important, distinctive, and innovative?

RQ13: How can this doctorate thesis inspire future research projects?

The system architecture depicted in Figure 2 can be used to synthesize this doctoral thesis project. Its features have as their focal point a Intelligent Embedded System (IES), which is integrated with a management platform denoted by the *AIDAEHR*.

New openEHR structures are uploaded through this platform, also the data refsets can be created with the *refsetBuilder* module, in order to be consumed by the *formBuilder* module. This is accomplished through the use of the *refsetBuilder* module, which enables for the development of both static and dynamic data subsets. Static subsets can be imported via *csv* files or built from other sources, whereas dynamic subsets are based on queries to specific databases, which results in the generation of a URL that can be utilized immediately.

Also, the *formBuilder* module receives the openEHR structures, which have been transformed into a correctly notated format in order to be consumed, and transforms them into a dynamic Graphical User Interface (GUI). The edit mode allows the user to make changes to the form fields, and to connect the refset URL created in the previous module with the form fields that have been modified. These standard forms are now available for any applications running within the environment.

The IES is comprised of a MAS that retrieves the required data while taking clinical governance standards into consideration, and then maps this data into the openEHR structure. In the field of addiction, MAS is responsible for the development and maintenance of the non-relational openEHR repository.

The BI component provides access to a wide range of information in order to summarize and provide physicians and IT professionals with the knowledge they require about the system. Aside from that, the ML portion is in charge of organizing the EHR records through the use of unsupervised learning techniques.

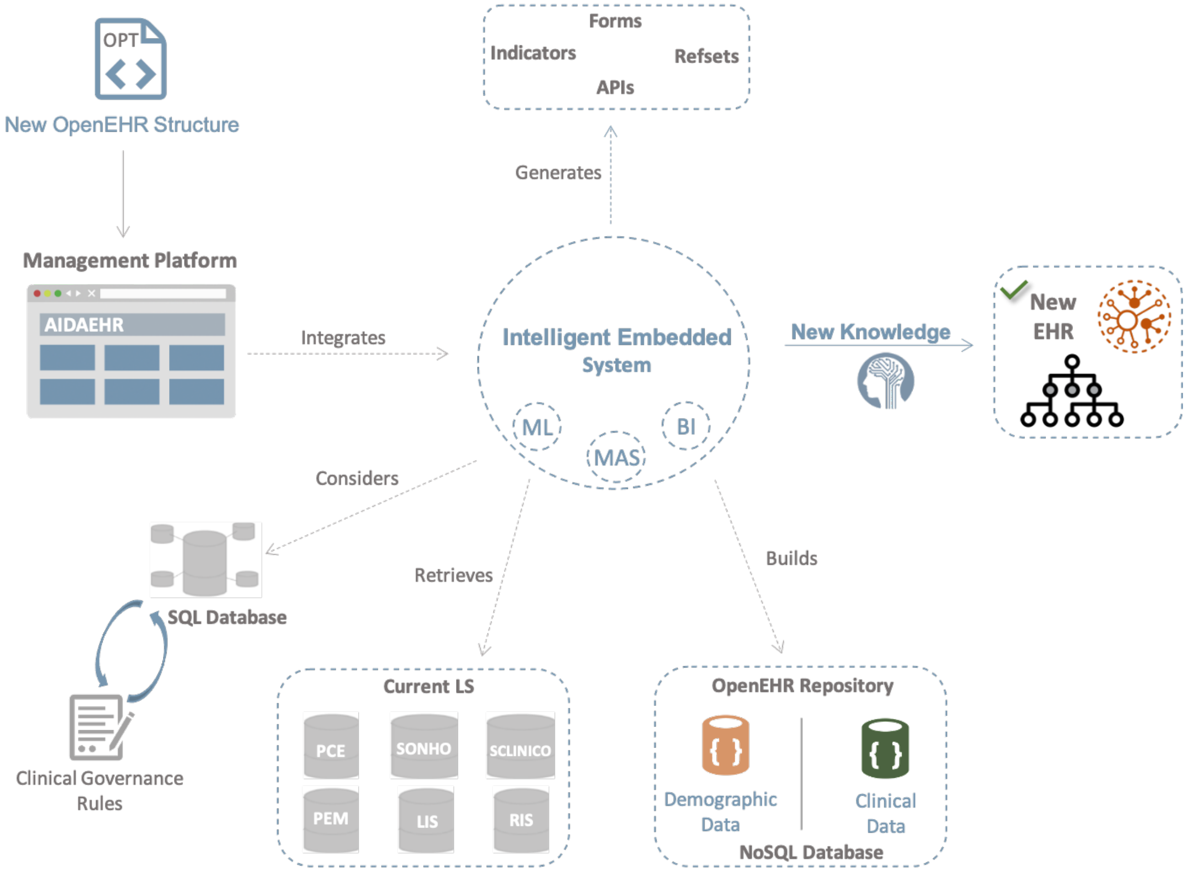

Figure 2: Doctoral thesis overall structure proposal.

1.4 Dissertation Organization

This manuscript is divided into six chapters, which are listed below:

- **Introduction**: the first chapter of this document begins with a description of the research project's focus and context, its purpose, a description of the main objectives to be achieved through its implementation, and the thesis structure. This chapter describes the overall approach used to conduct this research as well as the set of essential methods and procedures used, such as the Design Science Research (DSR) and Proof of Concept (PoC) strategies. There is also information about the research techniques employed in this thesis.

- **Background**: the most relevant theoretical and scientific concepts to this investigation are succinctly described and discussed, including the use of HIS and the evolution of the EHRs, the importance of international standards to promote interoperability between IT systems, the application of Multi-Agent systems in the healthcare industry, and the key concepts of Data Science for discovering new knowledge.

- **The Development of a Interoperable Healthcare System**: the new artefact is proposed and developed in this chapter, including the main technologies chosen and the architecture built for each system component.

- **Case Studies**: the use cases that were used to test the consistency and usability of the new developed system are presented in this section. Different approaches are explained, using the same system such as emergency clinical data migration to openEHR, an openEHR modelling structure approach to accelerate the COVID-19 patient's treatment, and a predictive model for COVID-19 diagnosis based on patient symptoms. In addiction, each use case has its own Strengths, Weaknesses, Opportunities, and Threats (SWOT) analysis and conclusions.

- **Discussion and Conclusions**: It concludes by highlighting and present the significant contributions made through the conclusion of this Doctoral thesis. The next steps in terms of potential future work are also described.

Chapter 2

Research Methodologies

2.1 Introduction

Developing unique techniques for defining research phases and procedures in the most effective and demanding manner is essential for the completion of a research project. Research subjects should be developed and clarified when the nature of the research has been determined. They should be backed by a thorough and critical examination of the literature. It is therefore necessary to develop a thorough research strategy that is based on the types of philosophies and techniques that will be used. It will be possible to submit the research project idea if it has been adequately supported and analyzed, both qualitatively and statistically, using primary and secondary data [100].

In order to conduct a critical literature review, it was determined that it was necessary to first look for keywords, and then, bibliographically, look for works carried out in the engineering field that fit the project's primary objectives and objectives as stated in the project. Primary sources, such as reports, articles, or academic dissertations of previously completed work, were examined in this follow-up, but from a more secondary viewpoint, some magazines and books. It was also necessary to conduct some direct online research in order to explain several topics. However, this can only be accomplished when the research approach to be utilized has been established, which in this case study will be of the deductive type: Through the use of a well-designed research strategy, a conceptual theory or hypothesis artefact will be built, and finally tested using data [120].

The investigation process results in the development of knowledge and skills in a certain domain, in this case, in engineering area, through the acquisition of new knowledge and skill. A philosophical perspective ontology is the branch of philosophy concerned with the nature of reality, presenting questions about how the universe works and the commitments implicit in the researcher's specific point of view on the subject. The ontological philosophy followed in this research is the Pragmatism that rejects any disagreements about specific ideas of truth and reality. The investigation is carried out in accordance with the researcher's ethical principles, and the findings are used to emphasize the importance of such an investigation in the future.

2.2 Design Science Research Methodology

In *Research Methods for Business Students*, the pragmatism is defined as an *external, numerous, view chosen to best ease responding of research topic* [120]. In order to answer the main research question and the subsequent ones, it was important to project the Research Design that would be used. Because it was applied to science and engineering, the Design Science Research (DSR) technique was adopted, which is the most extensively used method today for generating solutions and approaches in the IT field. When a problem is identified, this methodology supports the researcher in defining his investigation, resulting in an artefact that is then reviewed and, if it is the final solution to the problem, it is regarded as complete and can be communicated and published [75]. The methodology employs a cyclical approach, so it can be due several times until the best answer is discovered dependent on the generated artefact analysis.

According to Figure 3, the authors of *A Design Science Research Methodology for Information Systems Research* ilustrate the DSR methodology [94].

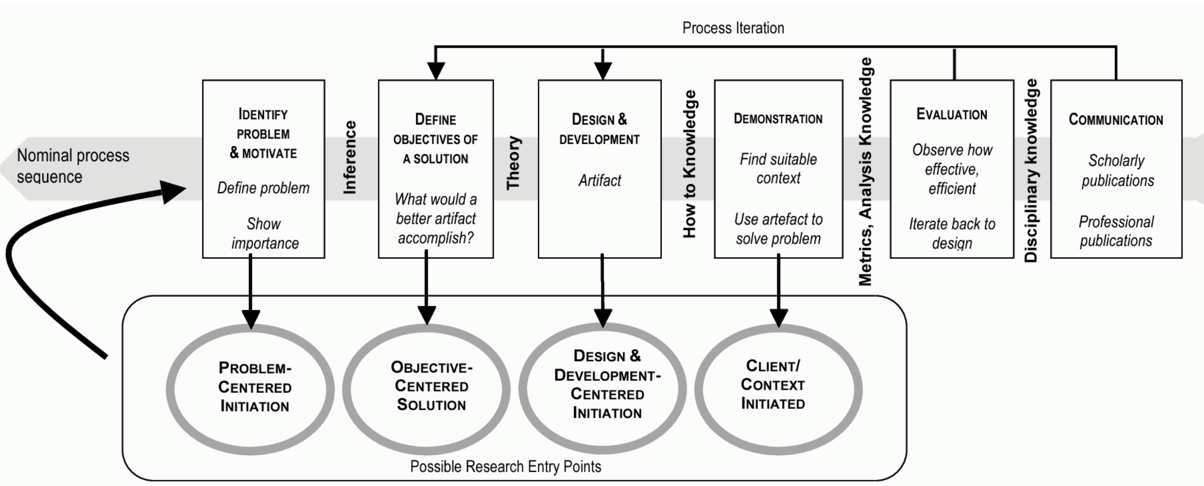

Figure 3: The DSR methodology's processing model. Adapted from [94].

Different research strategies were employed in accordance with the DSR methodology and their respective natures. Exploratory in nature, this research is looking for a fresh viewpoint and method to address inefficiencies that have been identified in a specific context. Through the application of the Case Study strategy, inefficiencies were discovered through the exploration of existing information exchange and storage mechanisms in healthcare institutions, and specifically in the CHUP portuguese institution, as well as the survey of their gaps, which resulted in the proposal of a new approach to healthcare information exchange and storage mechanisms.

Semi-structured interviews and focus groups were conducted in order to gather information for this evaluation. Questionnaires, content analysis, and the LS critical study were the techinques used in the process. A mixed research technique was adopted as a result, with quantitative and qualitative methodologies being used both jointly and independently. Additionally, the project is defined by its longitudinal nature, in terms of its temporal horizon.

2.3 Case Study Approach

The case study research approach, ethnography, and a cross-sectional methodology were employed in this project to conduct an in-depth and extensive research of several subjects of study (the cases), as well as their associated contextual conditions, at different moments. As a result, the goal is to examine several HIS in order to acquire a broad set of information, with the goal of producing and constructing a universal artefact that is relevant to the greatest number of similar healthcare environments as possible.

In a case study, the goal is to address a specific research topic. To achieve the best answer to the study topic, the information obtained is collected and analyzed. As a result, most of the users are never satisfied with a single study. Typically, a case study is focused on the search for similar examples and the interpretation of issues and solutions used, or it uses these cases to collect data necessary to solve the problem that occurs during the research [12, 45, 126].

Case studies can incorporate both qualitative and quantitative research methodologies. This study employs mixed methods research, in which many research techniques are utilized independently or in combination to collect and analyze data. As a result, quantitative and qualitative research methodologies of various types are used to collect and analyze quantitative and qualitative data. Although this, this research is primarily based on qualitative data. Primarily, primary data are collected through participant observation, focus groups, semi-structured interviews, and self-administered or interviewer-administered questionnaires, or surveys. Additionally, secondary data were gathered through archival research, primarily from primary (dissertations and conference proceedings) and secondary (journals and books) sources. Thus, data collection includes portuguese healthcare institutions and its associated contextual conditions, the target audience's characteristics, and scientific publications from research databases and search engines. Following that, using grounded theory, it was feasible to generate new theories and models based on the quantitative and qualitative data acquired and analyzed, mostly of an archetype. As such, it is supported by an inductive method, as the construction of those theories and models can only occur after data collection and analysis, which are then compared to the literature. However, this scientific project employed a deductive technique as well, as it is exceedingly difficult to perform a solid and credible study that heavily relies on case studies alone through an inductive strategy. The inductive approach carries a great risk. Thus, the preparation of this doctoral thesis is dependent on a literature review of the state of the art . As a result, hypotheses are formed based on existing theories, and a research method is developed to evaluate those hypotheses using a deductive approach.

2.4 Proof of Concept

The PoC evaluation methodology is a practical paradigm that may be used to prove or validate an existing concept or theory for real-world application [103]. The PoC applicability is commonly reported as one of the most important steps in the process of designing, developing, implementing, and proposing a new artefact of a particular solution in the IT area [101]. This method includes iterative usability testing with different user types and a cross-sectional study to validate the system's feasibility and utility, in order to measure the impact archetype.

Thereby, the PoC is an attempt to prove the validity of a concept, a hypothesis, or a method. Most often, it is employed in computer science, where the final result of a project is the creation of a new item. Additionally, a PoC is carried out to demonstrate the validity of scenarios that have been developed [13]. These methods allow us to demonstrate that an artefact is both efficient and compliant. The technical and functional qualities of the system were initially evaluated and the outcomes were evaluated by the Perceived Usefulness, Perceived Ease of Use, Behavioral Intention and Use Behaviour, using the Technology Acceptance Model (TAM) method [19]. Thus, the main goal is to determine whether the system actually benefits organizations before the definitive implementation, through a small number of users or a subset of the built system.

The acceptability, feasibility, and utility of this research project were justified in this doctorate thesis using the SWOT analysis and TAM evaluation methodologies, described below.

2.4.1 SWOT Analysis

The SWOT analysis evaluated project's main strengths, weaknesses, opportunities, and threats. This strategic planning method evaluates internal and external environments of a product or organization, justifying decisions and alternatives to its weak and strong points, opportunities, and threats [44, 93].

Favorable factors contribute favorably to the achievement of the initially set goals, whereas unfavorable factors have the opposite effect and are viewed as barriers. These two types of elements might be of internal nature, which an organization or product can self-control, or external nature, which is uncontrolled from their perspective. Throughout the case study, the SWOT analysis evaluated the prototype of new artefact, with the goal of gathering the discovered strengths and weaknesses, as well as opportunities and threats [46]. Thereby, in the one hand, the use of SWOT analysis, increase the potential of environment by highlighting the strengths of a specific solution, also, in the another hand, try to minimize the risks and weaknesses associated with it. It should be used by any business that desires to compete in its market. Figure 4 is a matrix representation of the SWOT analysis concept.

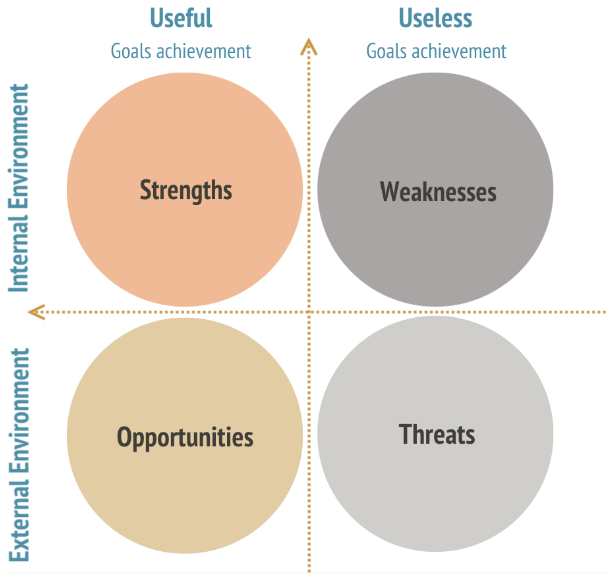

Figure 4: Matrix illustration representing the SWOT analysis's key concepts.

2.4.2 Technology Acceptance Model

Following the creation of a new product, it is critical to identify and establish circumstances for the product's acceptance by its target audience. Several theoretical models have been created in this area to assess the usability, acceptability, and usability of a technology. Davis' TAM stands out among them. This model was introduced in mid-1985, based on the Theory of Reasoned Action (TRA) paradigm, and quickly gained prominence for its effectiveness and coherence in forecasting the influence of technology on the community [40]. The TAM model is concerned with IT adoption in work environments, which is why it is frequently employed in the context of information technology and work-related technologies, and also in the medical field to evaluate several HIS. Briefly, TAM establishes a strong foundation for mapping the effect of external variables on an individual's intrinsic factors. Also, these external factors can be measured through variables associated with affective and cognitive factors [29, 74, 85, 117].

As a result of TAM increasing criticism, it has been expanded to include additional factors that influence behavior intention, such as social effect, age, and gender. In addition to incorporating external social variables that influence behavioral intentions to use new technology, Technology Acceptance Model II (TAM2) is an extension of Venkatesh and Davis' original TAM from 2000, which incorporates external social variables that impact behavioral intentions to use new technology. These social variables are the subjective norm, voluntariness and image. Also, the cognitive instrumental processes level is based on job relevance, output quality, result demonstrability and perceived ease of use. In 2008, Venkatesh and

Bala developed a more detailed version of the concept, originating the TAM3. Figure 5 demonstrates the evolution model over the time and major factors added to improve the new IT artefacts evalutation.

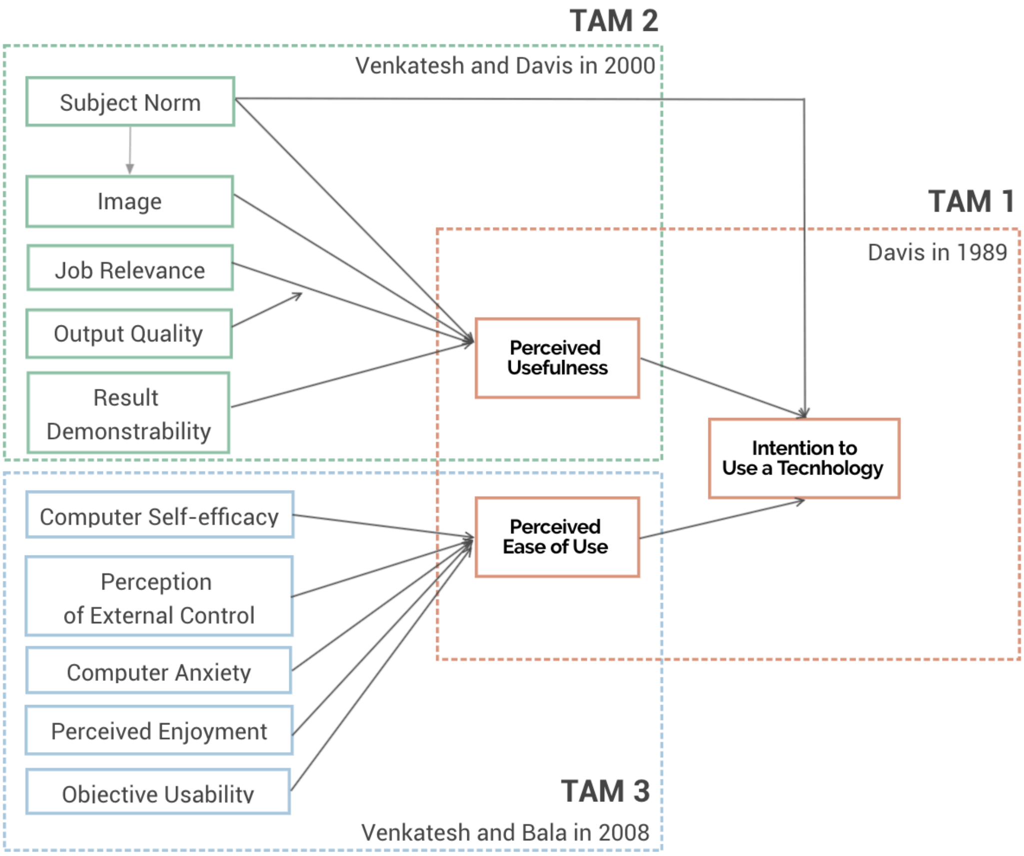

Figure 5: The Evolution of TAM Methodology. Based on [28, 116, 117].

In *"Technology acceptance model 3 and a research agenda on interventions"*, the authors claim that the TAM methodology reveals the users' behavioral intentions for a certain IT system are based on:

- Perceived Usefulness: The degree to which the user believes that using the new system can improve their work performance, for example, completing the tasks more quickly. It major determinants are:
  - Subjective norm: A users's perception of how important the people think that he should use the system.
  - Image: How much a user believes that using a new technology will elevate their social status.
  - Job relevance: How much a user believes the target system applicable to his work.
  - Output quality: How much a user feels the system performs her job well.
  - Result demonstrability: Individual's conviction that the new system outcomes are tangible, observable and sharable.

- Perceived Ease of Use: The degree to which a user has spend mental or physical effort in order to use the new technology. This evaluation is based on the next determinant factors:
  - Computer Self-Efficacy: How much a user feels able to do a task/job using a computer.
  - Perception of External Control: A user's belief in the system's organizational and technological resources.
  - Computer Anxiety: Level of apprehension or fear when confronted with the idea of using computers.
  - Perceived Enjoyment: How much the action of using a specific system is seen to be enjoyable, regardless of any performance consequences.
  - Objective Usability: Level of the effort required to accomplish certain tasks compared across systems.

In conclusion, the TAM3 was used in this research project to evaluate the proposed system's acceptance through the production of questionnaires for both target audience segments.

2.4.3 Research Techniques

The above mentioned methodologies were implemented in this research project, through the use of some research techniques, in order to facilitate and complete the research and data collection process. At the beginning of the project, and so that the problem formalization stage was possible, visits to the real environment of the study were made.

In order to understand the difficulties experienced on a daily basis by healthcare professionals, interviews were conducted, through face-to-face and non-face meetings, and in a more comprehensive way, an initial questionnaire was carried out to obtain a considerable sample, resulting in a better evaluation of the current situation.

2.4.3.1 Literature Review

An exhaustive bibliographic review was conducted in search of similar projects already done and published over time. This step is very important in all research projects, not only for a better definition of the problem under study, but also to obtain greater accuracy about the state of developments on the subject in question. According to Teresa Cardoso *et al.*, all researchers should thoroughly analyze the work of others who have done so, and only after this, go on their own adventure. A good review leads to an easier delineation of the case study, and the discovery of new lines of investigation, avoiding possible repetition and proposing future objectives [17]. According to [14], the driving forces for a rigorous literature review in a research project are:

- Delineation of the research problem in relation to other projects already done in the area under study;
- Definition of new research directions that have not yet been explored;
- Acquisition of new methodological perspectives, extracting new ideas about methods already used, according to the literature;
- Prevention of unsuccessful approaches, thus avoiding the use of methodologies that a priori do not contradict the expected results;
- Suggestion of new topics for future research, through other studies, adapting them to the case study.

2.4.3.2 Interviews

Interviews are one of the most important techniques in the research process. Direct contact with future users of the artefact makes the researcher able to assume their needs and conclude how important their contribution will be. The researcher can feel the reality and clarify his objectives, the structure he intends to follow throughout the interview and its purpose, gaining the trust of the interviewee to ensure his collaboration in the future [68]. For the correct use of this technique, some assumptions must be respected:

- The interviewer must be able to solidify the relationship with the interviewee, gaining his/her trust;
- The interviewer must ensure that the interviewee is as ”at ease”as possible during the interview, making the procedure as less painful and more dynamic as possible;
- The interviewer must be able to ask questions whose answers can not be influenced in any way, thus demonstrating a neutral position during the dialogue.

Through the various visits made to the institutions under study, opportunities were created to hold dialogues relevant to the study. On the clinical side, some doctors and nurses with clinical and technical knowledge were the main interviewees. On the technical side, IT experts from the institutions' information systems made a great contribution with their interviews. Throughout these interviews, the interviewees reported their greatest daily difficulties and introduced improvement interventions that, in their opinion, would be opportune in the technical and clinical acts and could be included in the project. In order to reach a larger target population, the next research technique was put into practice in a simple and objective manner.

2.4.3.3 Surveys

The survey technique is similar to the interview with the difference that the interviewer does not have direct contact with the interviewees, being written anonymously. The researcher should be careful in the type of questions asked, and these should be clear, coherent and as neutral as possible, so as not to influence the interviewee's answer [123]. In the first phase of the project, a survey with 8 questions was made, which can be seen in attachment (Appendix A), thus helping to identify the problem. The group of interviewees were IT members of some Portuguese healthcare institutions, with a final result of 23 answered questionnaires. The questions were constructed with the maximum of care and objectivity, with the objective of understanding the main flaws that the interviewees already felt in their workplace. All answers were treated in the Tableau software in order to clearly achieve key indicators from the data obtained.

Despite the fact that each questionnaire was filled out anonymously, the first question is about the health facility where the respondent works. The purpose of this question is to understand which healthcare institutions participated in this study. Figure 6 shows that the institutions with the highest participation rate, with the same number of responses, were CHUP and *Centro Hospitalar do Tâmega e Sousa* (CHTS). They were followed by the *Unidade Local de Saúde do Norte Alentejo* (ULSNA),*Unidade Local de Saúde da Guarda* (ULSG) and *Unidade Local de Saúde de Castelo Branco* (ULSCB), in descending order by number of responses.

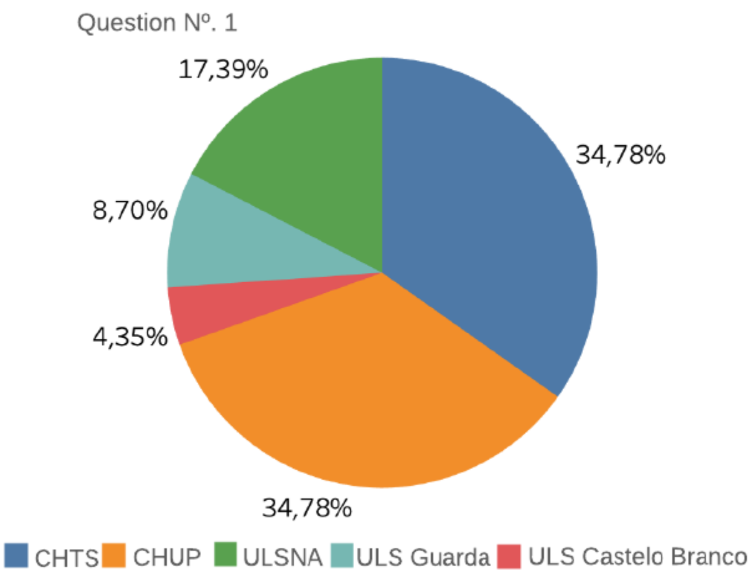

Figure 6: Answers of question 1.

The second and third questions inquired about the quantity of databases, as well as the different types and versions of them that respondents deal with on a daily work. The results of the two questions are shown in figure 7, which shows the large number of databases that the respondents handle on a daily basis, ranging between 8 to 10. A similar number of responders report handling two to four databases per day. In terms of database versions and types, the majority of respondents said they handle two to four different data environments.

These results reflect the high level of data decentralization existing in the HIS of the healthcare institutions surveyed. This is a significant barrier to system interoperability, causing some institutions to look for higher-level integration solutions.

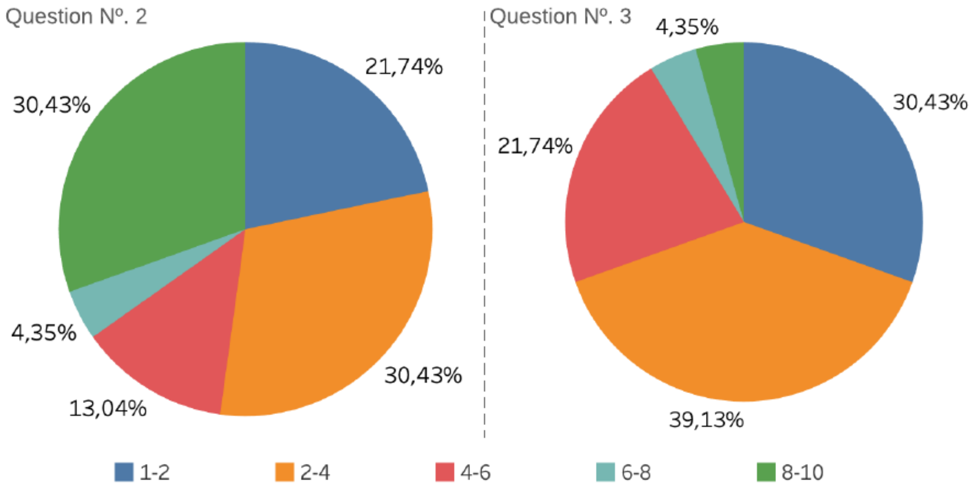

Figure 7: Answers of questions 2 and 3.

The answers to the previous questions are reflected in the responses to the 7th question, present in figure 8, with a large number of databases to maintain reflecting a large number of existing HIS. More than 50% of those respondents said they deal with 8 to 10 distinct HIS on a daily basis. This reveals the need for high attention from healthcare professionals to different platforms during the clinical act, and the high support work of IT experts.

As noted previously, the difficulties faced by professionals over time have necessitated the use of integrative platforms. The major goal of these integrative systems is to put the patient at the center of the clinical act by providing timely access to the needed medical information. This increasingly reinforcess the need for open and standardized solutions that establish the semantic and syntactic interoperability necessary for optimal symbiosis within and between hospital institutions.

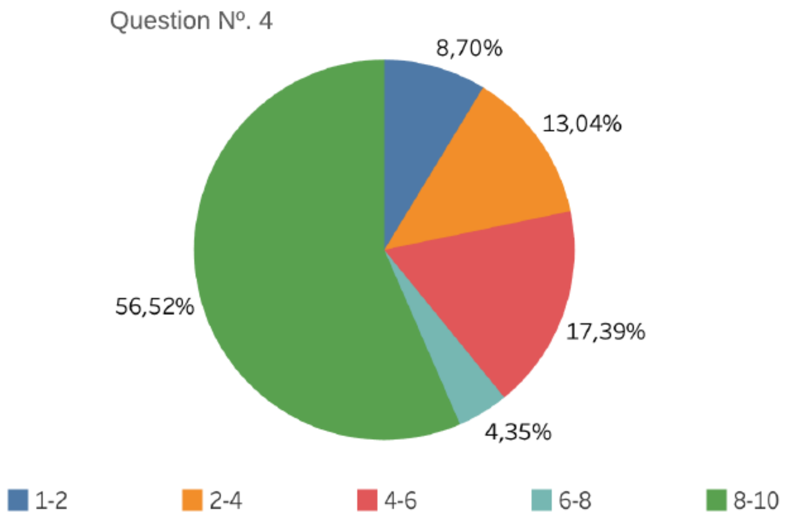

Figure 8: Answers of question 4.

The 5th question was made to evaluate whether the surveyed institutions already use established clinical standards. As indicated in figure 9, approximately 74% of respondents reported that their institutions already use standards to ensure HIS interoperability.

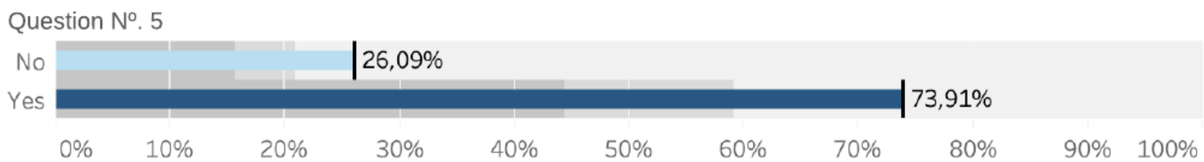

Figure 9: Answers of question 5.

As a result of the previous question, respondents were asked to record the standards their institution uses in the 6th open-ended question. The open-ended responses were duly processed and quantitatively analyzed, as presented in figure 10.

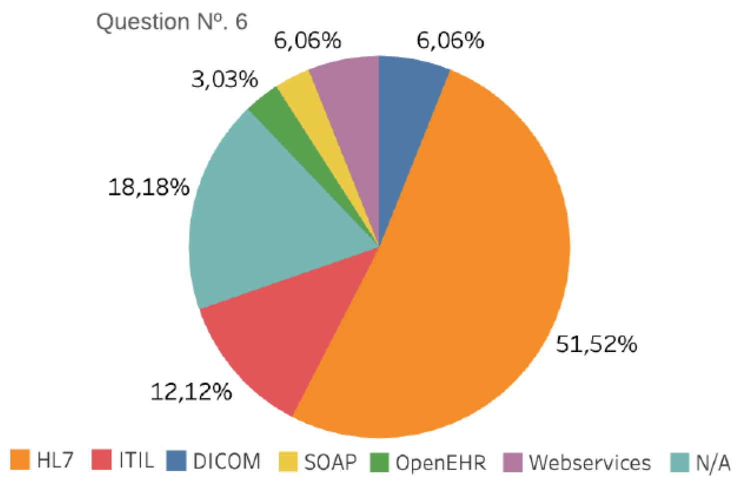

Figure 10: Answers of question 6.

As can be seen in the piechart in figure 10, the standard most used by the surveyed institutions is Health Level Seven (HL7), defined and described in section 3.3 presented below as one of the most accepted and implemented worldwide since 1987. The second most mentioned standard was the Information Technology Infrastructure Library (ITIL) framework, which has as its main goal to provide a consistent structure to manage and integrate different services, providing guidelines that help to achieve a good quality system environment and overcome difficulties that may arise in the development of IT systems. Also, many international companies such as Microsoft and IBM have already benefited from the potential of this framework for their IT services management.

The Digital Imaging and Communications in Medicine (DICOM) standard was also mentioned by some of the respondents. Its low number of answers may be due to the fact that most respondents take for granted that their institution uses the PACS system. This tool allows the digital archiving of medical images and their distribution in any sector of the healthcare institution, through the DICOM and HL7 protocol. Therefore, most respondents did not refer to the DICOM standard due to its integration already existing in the imaging system of their institution. The same number of respondents suggested webservices as being clinical standards used in their institution. Although, this answer does not correspond to the goal of the question number 6. "Webservices"are understood as a set of backend services developed for system integration, without any kind of clinical standardization. The same happens with the answer "SOAP"(Simple Object Access Protocol (SOAP)), which is a communication protocol with specific requirements whose response is returned in Extensible Markup Language (XML) documents.

Finally, a very small percentage of the survey respondents mentioned openEHR as one of the standards used in their institution. The goal of this question was to understand, at a national level, the impact that openEHR already has and how the artefact presented in this doctoral thesis might help in this regard.

The 7th and 8th questions try to discover the degree of intra-hospital and inter-hospital interoperability perceived by the surveyed health institutions. The responses were examined qualitatively, as can be seen in figure 11, with the majority of participants rating both scenarios as "Satisfactory"in terms of interoperability.

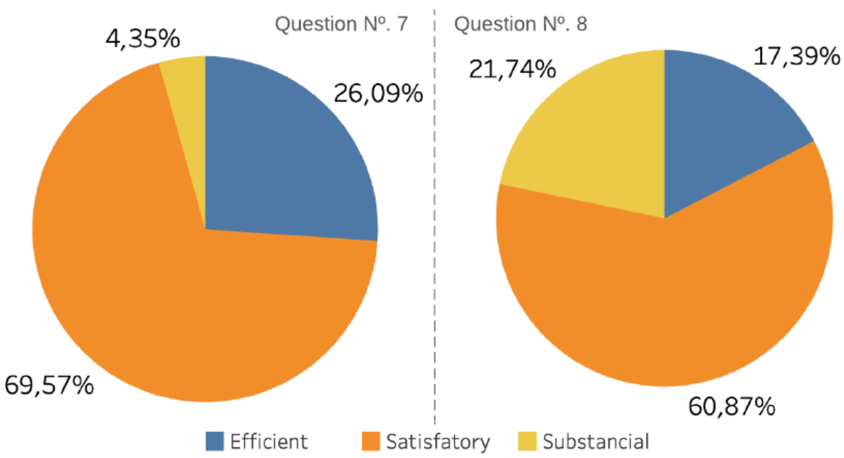

Figure 11: Answers of questions 7 and 8.

The questionnaire's 9th and last question is open-ended and aims to understand the involved institutions' current IT needs or gaps. The majority of the responses show a lack of integration between the HIS of *Sistema Nacional de Saúde* (SNS) and the public institutions, specifically the *Sistema Integrado de Gestão de Inscritos para Cirurgia* (SIGLIC), *Rede Nacional de Utentes* (RNU), *Sistema de Gestão do Transporte de Doentes* (SGTD), and Alert P1 systems. The necessity for interoperability between public and private health facilities was also emphasized. This gap has been felt for along time, owing to many existing barriers, such as data protection and private hospitals' lack of access to public *Registo de Saúde Eletrónico* (RSE) and *Rede Informática da Saúde* (RIS) systems.

2.5 Conclusion

This chapter was written at the beginning of the second stage of this research project. During the "Prototype Design"phase, it is primarily necessary to delineate a research methodology to conduct the present investigation. A search for the main methodologies used in IT area was conducted. The "Research Methodologies"chapter demonstrated that DSR approach is the most used methodology in this engineering field, and for that reason, these methodologies were chosen. Also, some research techniques were mentioned throughout this chapter and an initial survey was analyzed. This analysis was made through the key indicators that demonstrated the necessity of improving interoperability within and between institutions. RQ4 was addressed in the second stage of this project, and it was answered throughout this chapter.

Chapter 3

Background

This chapter discusses the state of the art relevant to this doctoral thesis. Thus, following a brief introduction, the HIS and EHR potential to support healthcare provision is highlighted in Section 3.1. Following that, Section 3.2 promotes interoperability as a crucial point. In Section 3.3, we discuss the potential offered by global standards practical application. Section 3.4 introduces the MAS approach in order to ensure synchronization and data exchange. Finally, Data Science (DS) importance is outlined in Section 3.5, as an efficient way to support decision-making and to develop more innovative goods and services.

3.1 Healthcare Information Systems and Electronic Health Records

Large amounts of information are produced and daily consumed in the healthcare industry. With the exponential growth of clinical data in recent years, there was an emerging need for its better organization, in order to automate, collect and analyze it whenever necessary. As a result, in the mid-1990s, hospitals began allowing IT to intervene in the clinical process at multiple levels. These IT systems enabled a more effective and efficient hospital management, from the development of clinical reports to the registration of clinical exams, among other functionalities. All of these advancements and advantages are facilitated and coordinated by IS.

To make a precise clinical decision, it's necessary to have access to patient information, including retrospective and prospective data. It's important that all of this information be immediately accessible, not only during the health care provision, but also to support the clinical research or any other activity that may contributes scientifically to continuous improvement in healthcare area.

HIS and the not-for-profit HIMSS both emerged in the mid-1960s and have had a significant impact on healthcare and clinical practices, from the administrative data management to the clinical data management and communication [67]. Thereby, hospital information management depends on the use of specialized information systems applied to the medical domain. The IT professionals, insurance and pharmaceutical companies, and governments use a variety of IS in the healthcare area, such as Data Entry System (DES) and DSS.

A HIS has a complex integration of several solutions inside a healthcare facility. The collecting, processing, and reporting of data are critical aspects that place the interoperability at forefront in a hospital environment. So, according this:

A HIS is a collection of data, procedures, people, and information technology that work together to collect, store, process, and provide the information needed for health care.

Every day, as a result of technological progress, new and modern systems arise with unique approaches to provide decision makers with enhanced healthcare data. So, the term ***e-Health*** is now used as a natural mode and represents not only the electronics concept, but also a set of words that are also important, such as efficacy, enhancing quality, evidence-based, empowerment and encouragement as well as education, enabling, extending, ethical and equity. The expression also means easy-to-use, entertaining, and even exciting [36].

However, the main concern is how to establish communication between these systems and the existing LS. Without this mindset, the separated mounts of information will continue to exist and grow inside healthcare organizations and countries [82]. Also, the main HIS characteristics are [122]:

- Has a socio-technical hospital subsystem
- Contribute to a high-quality and efficient care provision
- Simplify administrative workflows
- Process several information different types
- Affect the patient's health status directly, avoiding medical errors
- Deal with privacy and confidentiality concerns

The data that circulates inside HIS can be clinical, demographic, or administrative in nature. On the one hand, an EMR contains relevant data regarding a specific clinical act involving a particular patient, and on the other hand, the collection of all these data constitutes the EHR.

In order to design and implement a new IS in a healthcare institution, significant differences between EMR and EHR must be taken into account, as shown in Figure 12.

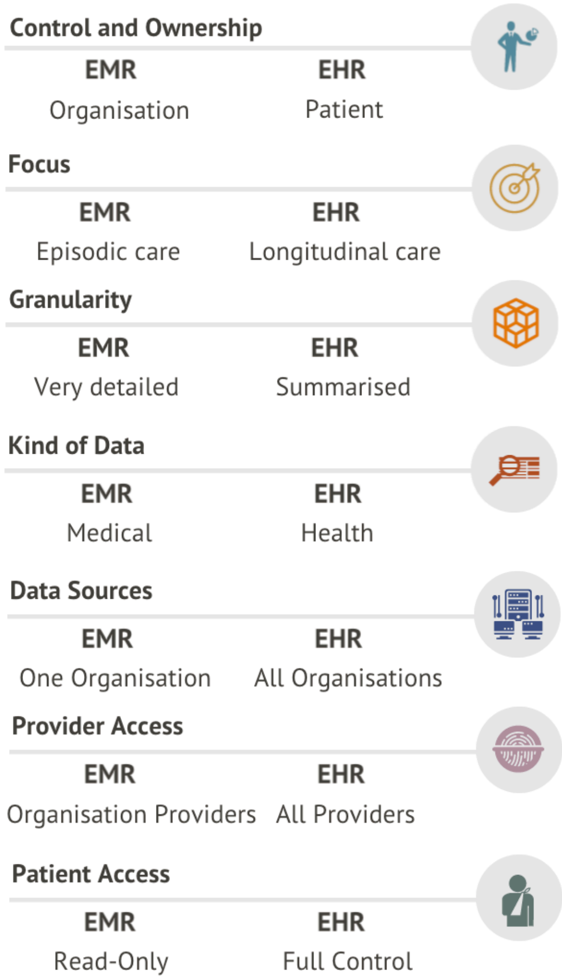

Figure 12: Key differences between EMR and EHR. Adapted from [114].

Within a healthcare institution, several units and departments collect pertinent and necessary information that will assist the professionals in their decisions. The EHR system collects, organizes, analyzes, and processes this data in order to presenting essential information in a simple, intelligible and accessible way to the users. Thus, the EHR is the basis for an effective management of a wide variety of information types, including the patient's medical history, demographics and personal metrics, clinical diaries and hospitalizations, as well as problem lists, prescribed and administered medication, allergies, clinical analyses and exams, and vital signs measurements [2, 79, 95, 96].

The conversion of MHR to digital format has been ongoing since the 1970s. Although it has considerably improved clinical access to patient data and reduced drastically the number of recording and clinical errors, the way that information is entered and stored in the EHR differs significantly, which can cause confusion among users [30]. Many platforms for physicians to interact with EHR have been developed, but the underlying information model of how clinical, demographic, and administrative data should be recorded and stored has been under the responsibility of IT professionals. As a result, depending on the systems structure of each institution, different methodologies and technologies were adopted. One of the most challenging aspects of dealing with the many LS is that the most of them do not employ clinical information standards. The software providers generate its own data models and is expected that this data be exchanged properly between different systems in other institutions, which rarely happens [47].

To ensure the viability and functionality of the entire ecosystem, it is vital to promote interoperability between all LS and to implement globally recognized standards. Thus, all patient information will be centered in only one place, with all updated data from other systems, and will be accessed and retrieved easily and quickly. Additionally, all existing data can be analyzed to construct alert or recommendation systems. Furthermore, implementing an EHR-oriented system enables the creation of a big database for doing research on a several human health-related topics [30, 31].

The EHR characteristics indicated in Figure 13 are based on the next principles: it's important to have clear and structured information, and exchanging health information between different LS must be easy [59].

Figure 13: The widely recognized EHR characteristics.

In [97] the authors recognized the lack of standards and interoperability for dealing with big data scenarios, compromising data storage and, as a consequence, knowledge discovery. As a consequence, there is a need for innovative approaches to defining new clinical information workflows, that is, how the information will be recorded, exchanged, and retrieved in the EHR. In order for these approaches to be viable, HIS must share information with a symbiotic and interoperable way.

3.2 Interoperability

It's crucial to highlight that openness and shared properties are key success factors for *e-Health* solutions. Technologies in a medical operating protocol can only handle a limited number of needs. Many times, a solution is part of a collection of supporting resources for other protocols, and this is the core of the issue. The environment has been populated with hastily generated new technological inputs with no regard for mutual interaction. As a result, each solution's time-to-market is limited. In practice, however, the protocol's implementation must be feasible and simple. In medical terms, this large number of solutions makes it difficult for healthcare professionals to immediately integrate a patient's situation. So, Interoperability is captivating because to its potential to interact harmoniously with a wide range of technologies, ensuring that all modules work at maximum efficiency and communicate with each other easily and consistently [66].

Table 1: Interoperability definitions by HIMSS, ISO, IEEE, CEN organizations.

Organization	Definition
HIMSS	*"It's the ability of different information systems, devices and applications to access, exchange, integrate and cooperatively use data in a coordinated manner, within and across organizational (...) boundaries, to provide timely and seamless portability of information and optimize the health of individuals and populations globally."* [58]
ISO	*"Interoperability is the ability of independent systems to exchange meaningful information and initiate actions from each other, in order to operate together to mutual benefit."* [61]
IEEE	*"The ability of two or more IT systems to exchange information and to make mutual use of the information that has been exchanged."* [56]
CEN	*"A state which exists between two application entities when, with regard to a specific task, one application entity can accept data from the other and perform that task in an appropriate and satisfactory manner without the need for extra operator intervention."* [27]

The abovementioned institutions define *e-Health* interoperability as a technical contribution that operates predominantly at the software and data layers of the overall system architecture, consequently providing benefits to actors during some of their collaborative acts. Also, its definition must go beyond this, addressing organizational, cultural, ethical, economic, financial and legal aspects. Each of these factors is a information significant source that can be critical in the collaborations development between different HIS [66, 81].

However, in mid-2019, HIMSS updated its interoperability definition, adding a fourth level to fullfill the supramentionated gaps. In Mari Greenberger's point of view as a senior director of informatics at HIMSS, the organization is trying to achieve global interoperability by adding an additional "organizational"level to address the need for a robust interoperability infrastructure, taking into account non-technical aspects that contribute to the success of interoperability.

According to HIMSS, the interoperability can be classified into four distinct levels (See Figure 14):

- **Foundational:** Establishes the inter-connectivity requirements for a system or application to securely communicate with and receive data from another system or application.
- **Structural:** It defines the syntax for data exchange, i.e., the format, syntax, and organization, including at the data field level interpretation.

- **Semantic:** Provides the ability for two or more systems to understand the meaning in same way of the models and data encoding exchanged, including the use of data elements derived from publicly available value sets and encoding vocabularies.
- **Organizational:** Includes governmental, political, social, legal, and organizational aspects to facilitate communication and safe use of the systems. These components enable shared consent, trust, and integrated end-user processes and workflows.

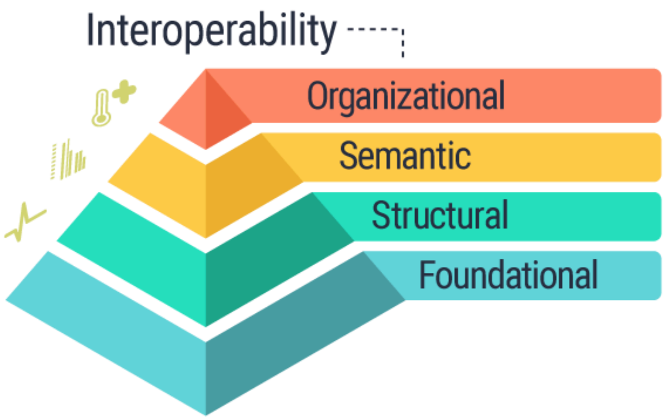

Figure 14: HIMSS new definition for interoperability. Adapted from [51].

While current solutions and technologies are capable of solving the majority of basic and structural interoperability issues, semantic interoperability is still the main concern, with a special research development focus. The necessity to make health information understandable and automatically computerizable by external IS increases the challenge of establishing the semantic interoperability level. From the viewpoint of the European Commission (EC), semantic interoperability is an essential factor to materialize the benefits of EHR, allowing to improve the quality and safety of health care provision, public health and its management [24].

A knowledge-oriented HIS development, including ontologies and clinical terminologies, is emerging to represent and share complex medical meaning. This development must ensure a viable, maintainable and adaptable patient-centric EHR system. Some requirements must be fulfilled in order to achieve this:

- Reliable and secure environment for healthcare data exchange between systems;
- Update EHR data with clinical terminologies and information templates;
- Update EHR data with context and explanations understood by patients;
- Ensure data quality and consistency for public health and research.

3.3 Healthcare data standards

A result of international coordinated efforts has resulted in the ongoing development of standards and guidelines for defining an EHR as one or more open access information repositories by different HIS. The CEN, HL7, ISO, and the openEHR Foundation are non-profit organizations dedicated to the international frameworks and standards development, including terminologies, EHR specifications, and information models, in order to support the exchange, integration, and retrieval of electronic health information [34].

As already mentioned, the most difficult level to achieve and guarantee is the semantic interoperability level. According to EC, semantic interoperability will only be feasible if there are agreements on standards, information models, terminologies, and semantic definitions used for data sharing. In addition, social, cultural and legal aspects within each organization, region or country will influence the implementation of semantically interoperable systems. It is recommended that such a high level of semantic interoperability be initially implemented and practiced only in specific, priority clinical areas of high relevance to patient safety. Current attempts to standardize the capture, representation and communication of clinical data span three artefact layers represented in Figure 15 [118].

Figure 15: Artefact layers for developing systems with semantic interoperability.

On the one hand, the RM can guarantee data in a complete, homogeneous and secure way. On the other hand, a clinical data structure definition allows to build precise clinical information models with a unique meaning and use. To complement this, clinical terminologies are precious, ensuring the unified meanings through specific vocabularies. So, the main question that arises here is *"What is the importance of standards in the clinical environment?"*.

To answer this question, we have to take into account that for a long time, HIS were designed to satisfy certain functional requirements defined by specific users. As a consequence, over time, many of these existing systems became limited in different contexts of use, or even obsolete in the face of new user requirements. Besides that, many of these older systems tended to record only the minimum information required for each clinical act. Consequently, additionally valuable information for diagnosis and clinical decision support was continuously lost.

As represented in Figure 16, clinical standards should give flexibility to data structures, so that new relevant information can be added. Thus, EHR standards define generic data models that can be linked to specific clinical concepts and information structures. In addition, standards should ensure a complete representation of clinical act context information, such as the participants and location of clinical act, which may influence the clinical data interpretation contained in the EHR. Also, clinical data are complemented by information necessary to support audit or control tasks [84].

Figure 16: Key principles for the standards development.

In *"Clinical data interoperability based on archetype transformation"*, based on their extensive literature review and international guidelines, the authors recommend using standards with *duo* architecture to ensure semantic interoperability through the usage of a basic unit named **Archetypes**. These are the minimum units of information that HIS can transmit with one another. An archetypes combination gives rise to complex **templates** capable of demonstrating sequential clinical workflows. ISO 13606, openEHR, and HL7 are examples of dual architecture standards that are used globally [26].

Also, the *"Towards Semantic Interoperability for Electronic Health Records: Domain Knowledge Governance for openEHR Archetypes"* research classifies the EHR as Patient-centred, longitudinal, extensive e prospective, and must retain all the care provided to the patient and not just an isolated episode, therefore, its character is longitudinal and comprehensive, being present not only past events, but also future perspectives and preventive planning. Therefore, the sharing of clinical data between different healthcare professionals and institutions, through different HIS, has to be efficient and fast. For this, clinical data will have to be of high quality, reliable and flexible. Thus, highly standardized structures will have to be developed, as well as each institution's governance rules for managing them. As a viable solution, the authors propose the use of the openEHR standard. They argue that various levels of interoperability can be achieved using it [43]. The most popular healthcare standards are presented in the next Table 2.

Table 2: The most used standards in healthcare IT.

Purpose	**Standards**
Content	HL7 v2, HL7 CDA
Data Exchanging	HL7, FHIR, DICOM, PACS
Terminologies	ICD, SNOMED CT, LOINC
Information Modelling	HL7 RIM, ISO EN 13606, openEHR

3.3.1 HL7

The HL7 organization emerged in 1987 in the USA with the purpose of creating a complete framework for electronic health information exchange that supports clinical practice, administration, delivery, and evaluation of healthcare services [3]. Based on its expertise, the organization classifies the next messaging standards user types and how they affect the standards development and use [48].

- Clinical interface specialists: Users responsible for creating applications that allow clinical data exchange between applications or healthcare providers;
- Government and politically entities: Users which have the ability to adopt or mandate a messaging standard;
- Medical informatics: Users that want to build or adopt a clinical ontology, implementing a data model, terminology and clinical workflows.

HL7 v2 was published in 1991 and is the most widely used standard for healthcare in the world, with presence in more than 35 countries. However, this version is not well normalized and there is no explicit attempt to define the process. So, in order to fullfill this lacks, HL7 organization have join their efforts to supporting the HL7 v2 in parallel with new HL7 v3.

The HL7 v3 first published was in 2003 and defines as a set of specifications based on HL7's Reference Information Model (RIM), supporting healthcare workflows following a model driven methodology type. The RIM consists of five main concepts: Act, Act Relationship, Participation, Roles and Entities. So, in this model, every event is an Act, i.e., procedures, medications, and observations, and a set of Acts are related through Act Relationships. The Act context is defined by Participation, which is, i.e., the author, location, performer, subject. Thus, each participant has a specific Role to play, such as patient, employee, provider, etc, but these roles can only be played by a specific participant type, namely Entity, i.e., persons, organizations, etc.

In 2005, the HL7 v3 Clinical Document Architecture (CDA) was released as a clinical document in terms of persistence, context, authentication, management and Human readability. CDA is based on XML structure and can represent an admission, discharge summary, imaging or pathology report, etc.

The first version of Fast Healthcare Interoperability Resources Specification (FHIR) was merged in mid-2011 and its focus is on implementability, ensuring a fast and easy implementation, with multiple libraries available and can co-exist with other HL7 versions. In addiction, FHIR has a strong foundation in web standards (XML, Json, HTTP, OAuth,etc) and support RESTful architectures [57, 62].

3.3.2 DICOM and PACS

DICOM emerged in 1993 as a standard that allows storing, retrieving, printing, processing, transmitting, and displaying clinical digital images and their information between imaging medical devices and workstations. Depending on the imaging exam type performed, there are diffent protocols for radiography, ultrasonography, etc. Also, DICOM fields can reach sizes in the order of GB, so it is necessary to compress the images with a lossless compression algorithm. When recording, these files are consequently saved in the PACS.

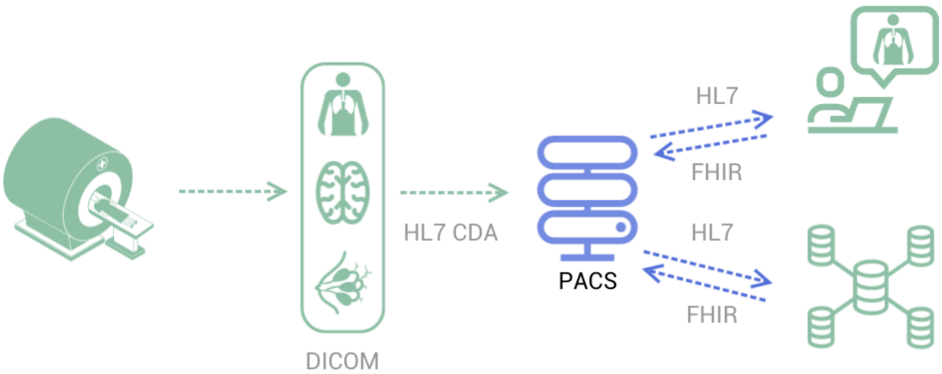

Figure 17: PACS storage system and standards involved in data exchanging.

The DICOM standard is considered to be very useful and complete for storing and exchanging medical images, ensuring good interoperability of images between different medical devices, the organization, and its healthcare professionals [18].

3.3.3 ICD, LOINC and SNOMED CT

The adoption and continuous improvement of the EHR has raised the issue of coding or not coding the information. The search and use of the information stored in a system in free text format has a large margin of error and loss of information due not only to the existence of abbreviated terms, but also to the occurrence of spelling errors in the record. Thus, it has become indispensable to use coding in order to be able to retrieve the data and group then when necessary. The use of codes ensures that a concept, even if represented by a synonym, acronym, or keyword, is uniform for all clinicians who interpret it. Thus, greater consistency and reliability in data entry is achieved, allowing clear and uniform communication between different healthcare providers. The standards listed below are some of the most often used in HIS as a terminology around the world. They are used to assist the input of medical data, while other standards are used to classify and retrieve data from a system.

3.3.3.1 ICD

ICD is a standard for classifying and coding information through a set of diagnosis and procedure codes, enabling mortality and morbidity statistics. It is used in various types of health records, mainly death certificates and hospital files, thus becoming the international standard diagnosis classification system. The World Health Organization (WHO) is in charge of publishing and revising the ICD standard, and the most recent stable version is ICD-10 CM/PCS. The 43rd World Health Assembly approved ICD-10 in mid-1990, and it has been in use in World Health Assembly member countries since 1994 [21]. The Figure 18 represents one of its 22 chapters.

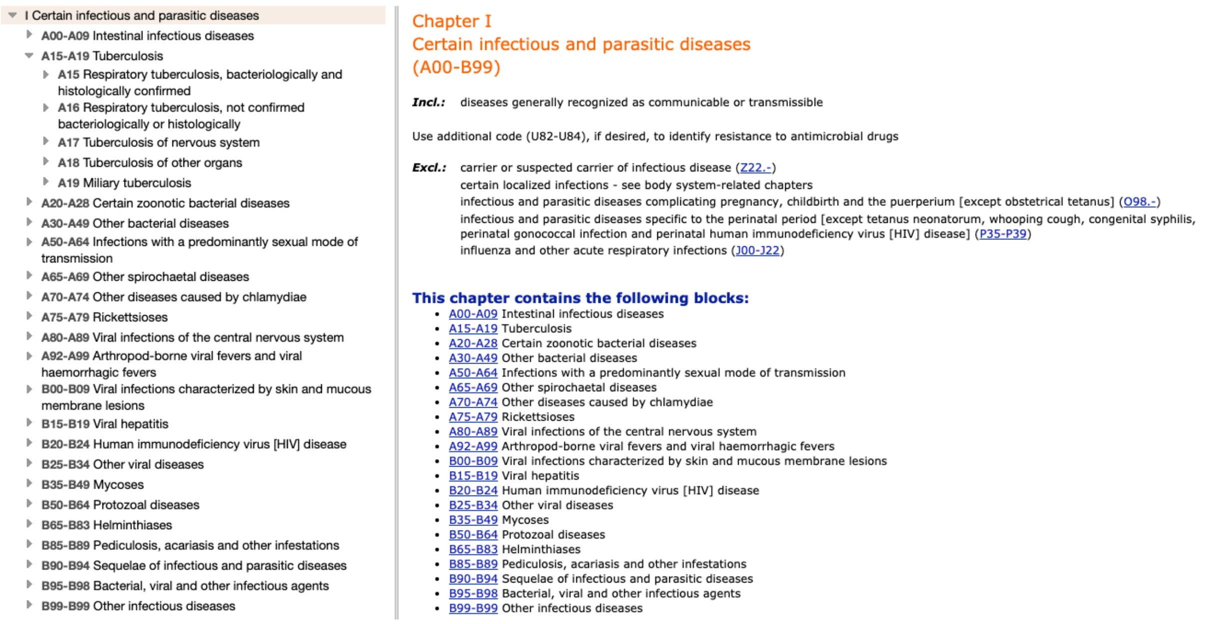

Figure 18: ICD-10 codes for infectious and parasitic diseases [55].

ICD-10 has almost 68,000 clinical diagnostic codes and 87,000 operations and procedures codes, providing many combinations and branchings. In the first step, the standard is divided into ICD-10-*Clinical Classification* for diagnosis codification and into ICD-10-*Procedure Coding System* which is used for procedures codification [49]. Also, the most recent ICD version was lauched in 2019. ICD-11 was presented to the member states and will come into effect in 2022 [10].

3.3.3.2 LOINC

Logical Observation Identifiers Names and Codes (LOINC) is a worldwide coding system for categorizing clinical tests and observations that allows data to be shared for clinical care, research, and outcomes management, among other tasks. This standard defines medical measurements such as vital signs, Electrocardiogram (ECG), hematology results, among others. LOINC codes distinguish a given observation across six dimensions called Parts, as showed in Figure 19.

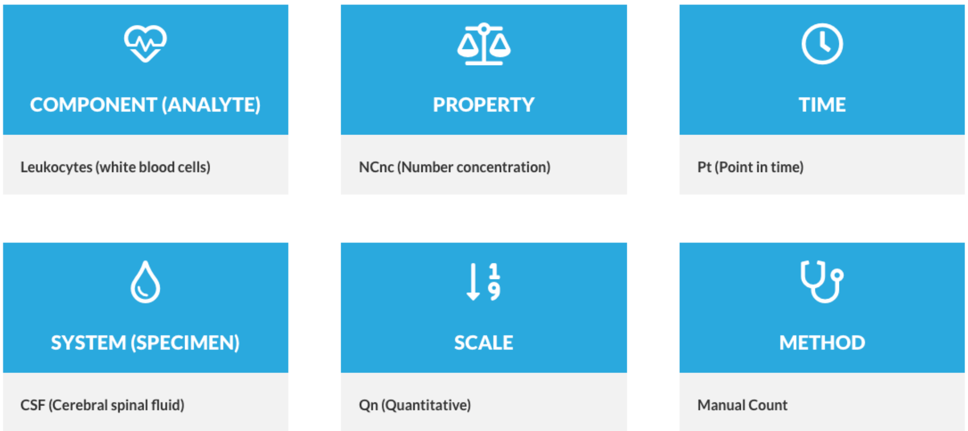

Figure 19: Manual count of white blood cells in cerebral spinal fluid specimen, represented by LOINC code **806-0** [71].

In the example shown above, leukocytes are the component to analyse, with the NCnc (Number concentration) as a property and the Pt (Point in time) as a timing. Also, the CSF (Cerebral spinal fluid) is the sample to analyse quantitatively through a manual count method.

The LOINC is maintained by *Regenstrief Institute* and its latest version (2.70) has almost 97,000 codes divided by Laboratory type (58,464 codes), Clinical type (25,301 codes), Attachments type (1,157 codes) and Survey type (11,313 codes) [63]. In 2018, LOINC already had 60,000 registered users across 170 countries and it has been translated into 18 variants of 12 languages.

3.3.3.3 SNOMED CT

Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) is a most complete and multilingual clinical terminology in use worldwide, can provide a multiple benefits for healthcare systems worldwide [107]. It is owned, maintained, and distributed by SNOMED International, an international not-for-profit organization in the United Kingdom, and is in use in more than eighty countries.

The SNOMED CT terminology is composed of concepts, terms, and relationships that enable accurate representation of clinical information based on logic organized into hierarchies for the entire healthcare area. Its use provides interoperable coded data that enhances the implementation of clinical practice, facilitating decision support [108].

In 2020, the SNOMED CT catalog had 352,567 concepts about diagnosis and clinical findings (signs and symptoms), such as surgical, therapeutic and diagnostic procedures, and also other relevant information which supports clinical knowledge: body structures, organisms, medications, specimens, among others, as shown in next Figure 20 [106].

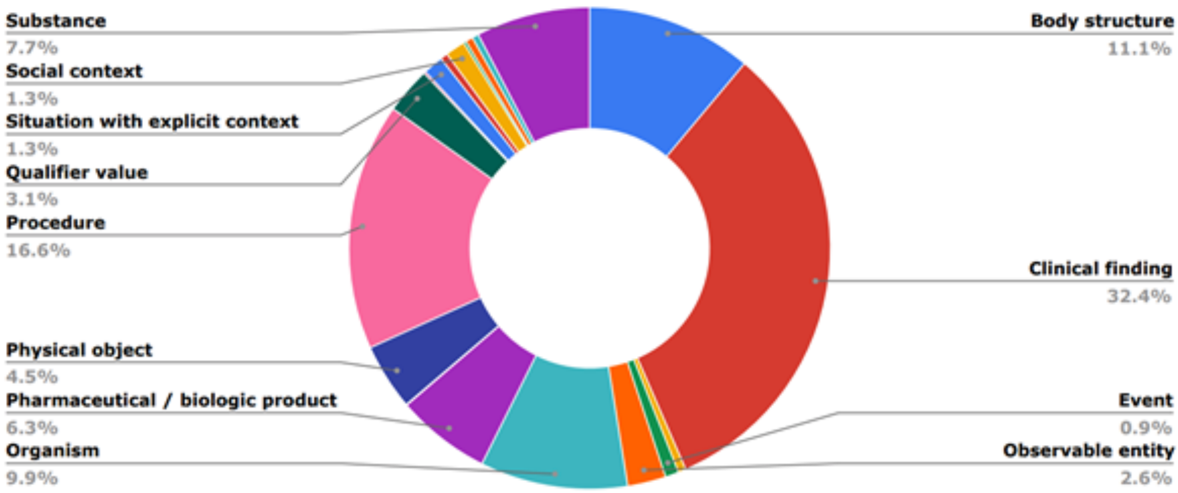

Figure 20: The hierarchical division of concepts (June 2019). From [106].

The hierarchical structure of terminology is divided into categories at different levels, with 19 categories at the first level. Each category is sub-classified into categories at the second level, and these, in turn, are sub-classified into child concepts, and so on until the concepts with the greatest granularity are achieved. This conceptual model allows multiple inheritance, where a concept can belong to several categories [16, 127].

In *"Evaluating Suitability of SNOMED CT in Structured Searches for COVID-19 Studies"*, the author demonstrates that the SNOMED CT concepts effectively capture typical data items important to COVID-19 research. The researchers used the German Corona Consensus Dataset (GECCO) dataset as a knowledge basis and the scoring system ISO/TS 21564 to prove it. Also, the writers recommend the use of a hybrid approach using several terminologies and the implementation of mapping rules to improve mapping quality and intercoder reliability [119].

3.3.4 ISO EN 13606 and OpenEHR

All clinical data needs context information, i.e., each clinical act must be accompanied by additional information such as how and by whom the data was recorded, and where it was recorded. This additional data is very useful for improving the interpretation of clinical data, supporting the organization of tasks, supporting audits, and legislative issues.

Currently, one of the major concern of healthcare IT area is information modelling, describing how information should be organized and structured. Previously, most IT solutions and their underlying information model for storing clinical information have been left exclusively to the developers. Thus, over the years, issues of linking different information from different systems or, e.g., the difficulty of interoperating multiple devices into a single database have arisen [53]. It was then increasingly emergent to develop standards that provide functional and consistent information models across different situations.

Nowadays, the most advanced EHR architectures and standards are based on dual model approach architecture such as ISO EN 13606 and openEHR, defining by the Reference Model (RM) and Archetype Model (AM). Thus, important generic concepts should be defined [77]:

- **Reference Model (RM)**: A set of entities that act as the generic building blocks of an EHR and also contain non-volatile EHR features, defining clinical information.
- **Archetype Model (AM)**: It presents the clinical concepts in a combined, structured way, restricted to the entities contained in the RM.
- **Archetypes**: *Tool* for building clinical compliance in a consistent way in dual model approaches, and they are considered essential for providing completely interoperable EHRs. These structures define clinical concepts through the combinations of entities presented in RM, defining clinical knowledge.
- **Templates**: Combination of two or more archetypes, representing a full clinical act.

3.3.4.1 ISO EN 13606

ISO EN 13606 was originally designed by CEN to achieve the semantic interoperability level, specifying information models and vocabularies relevants. The norm main purpose is to define a strong and stable architecture for communicating part or all of a single-subject EHR between HIS systems, or between HIS and a centralized data repository. Figure 21 represents the RM main classes of ISO EN 13606.

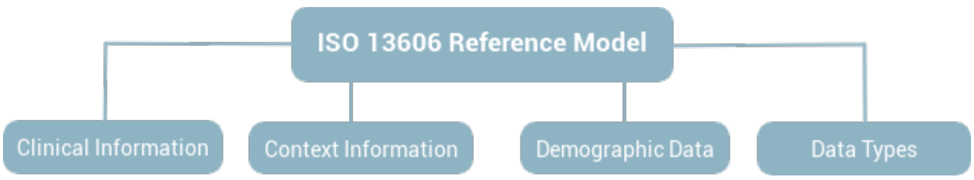

Figure 21: Set of classes of ISO EN 13606 RM [60].

Each main class has a set of classes. Thereby, the clinical information presented in EHR - folder, composition, section, entry, cluster, and element as the base unit. The main subclass of clinical information class is the EHR extract. The context information class contains additional information and annotations as well as the versioning system. Furthermore, the demographic data class supports the institution in identifying the entities and subjects and their roles. Finally, the data types class is the basis of any record, subdivided into basic types (Integer, Real, Boolean, Instance Identifier, Multimedia), code and text types (Coded simple, coded value, concept descriptor), quantity types (Physical quantity, Ratio of numerator, Ordered values) and also, the date and time types (date, timestamp, duration). ISO EN 13606 was initially published in 2008 and was reviewed in 2019. Thus, Figure 22 represents the latest RM version when it is possible to see each component and its inheritance in more detail, through the UML diagram [86].

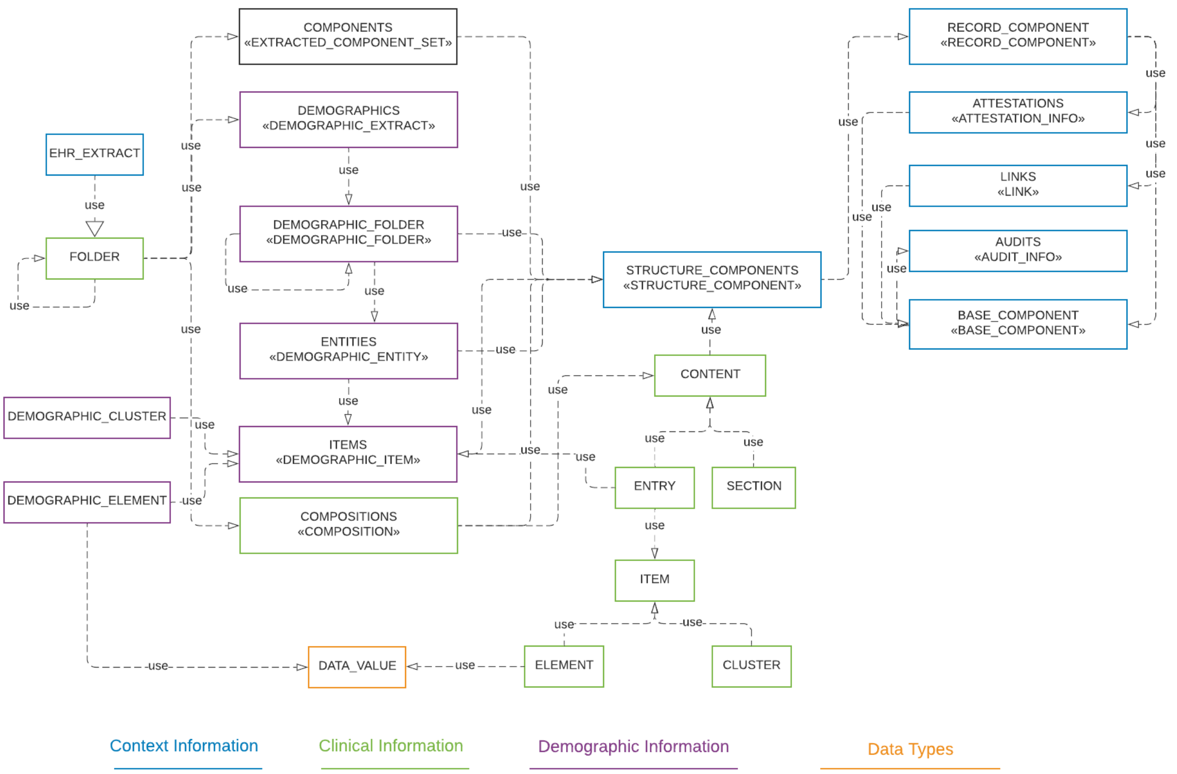

Figure 22: ISO EN 13606 RM class diagram.

Compared with openEHR RM, the ISO EN 13606 RM is more simpler. ISO EN 13606 also represents the information of an EHR, however it does not provide support for other relevant modules, such as the representation of local information templates and a query language for the EHR. Therefore, this norm was designed with a specific propose: Extracting EHRs between HIS.

3.3.4.2 OpenEHR

OpenEHR is a non-profit foundation with the University College London and Ocean Informatics company as a founders. Besides that, openEHR has a huge international community with a main goal: create comprehensive and interoperable clinical EHRs.

OpenEHR presents itself as an open industry standard for e-health specifications, models, and software. The community has a specification program in order to ensure quality in healthcare information, support industrial technology, integrate legal requirements, and manage the impact of change in healthcare institutions. To accomplish these goals, various quality criteria were established, including having a clear scope, ensuring modularity, and performing a rigorous design. Furthermore, ensure that the specifications are coherent, understandable, and unambiguous. Moreover, it should be simple to compute and implement, with a consistent change-management system [52].

All of the openEHR requirements, such as object models, languages, and Application Programming Interface (API)s, are organized by component in the diagram presented in Figure 23. This view gives a quick snapshot of the entire set of openEHR requirements. Dependencies exist exclusively from higher to lower parts.The CNF and ITS components are separated because they are semantically different from the primary specifications, but are important artefacts for further use by software engineering [7].

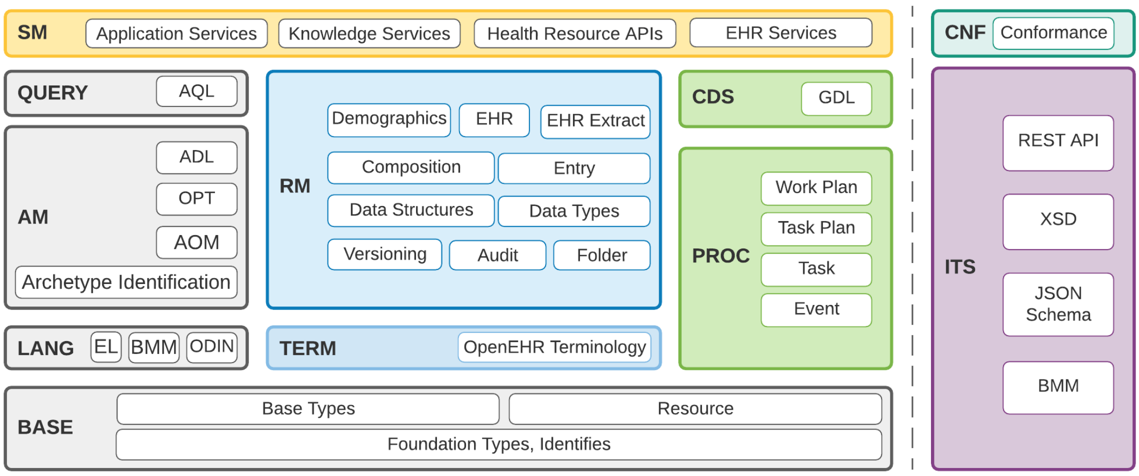

Figure 23: OpenEHR Specification - global view.

The core of openEHR standard are the artefacts of clinical knowledge named Archetypes, defined in AM, which are formal specifications of clinical information and are the basis for defining, discussing and present the medical knowledge [42]. In order to fulfill RM sub components, the BASE component is made up of fundamental and identifier types. Also, some terms can be coded by TERM component. In terms of clinical decisions, the Clinical Decision Support (CDS) module ensures the decision logic structuring through the use of Guideline Definition Language (GDL). The openEHR Decision Logic Module (DLM) is a formalization for expressing decision logic and rules used by process-oriented healthcare information systems, active forms, and other purposes. Decision logic processes are often nested inside a larger environment of plans, guidelines, and data sources, such as the EHR. Task Plans and GDL v3 standards both demand an mechanism for expressing rules and specifying input variables in openEHR ecosystem.

These needs are satisfied by the adaptable use of a single type of multi-section module known as a DLM [90]. In terms of APIs and interfaces, the Service Model (SM) is a set of specifications for developing concrete service interfaces to openEHR system components. It is a formal and abstract definition of the platform interfaces, so that the formal semantics of the interface can be stated and separated from the implementation technology [91].

The *duo* architecture of openEHR is composed by Archetype Model (AM) and Reference Model (RM). Clinical concepts are combined and structured on Archetype Model (AM) layer, based on RM classes. As well as ISO EN 13606 AM, the archetypes are the elementary concept base in Archetype Definition Language (ADL) format as a abstract sintax, and its combination generates template structures. The possibility of sharing and reuse structures is an essential advantage to create different and more complex structures (Templates).

OpenEHR has a very solid RM, separating the semantics domain, which makes it possible to involve different types of healthcare professionals, e.g., clinicians and administrators, as well as IT developers in its implementation [8]. Figure 24 represents the main IM packages of openEHR RM.

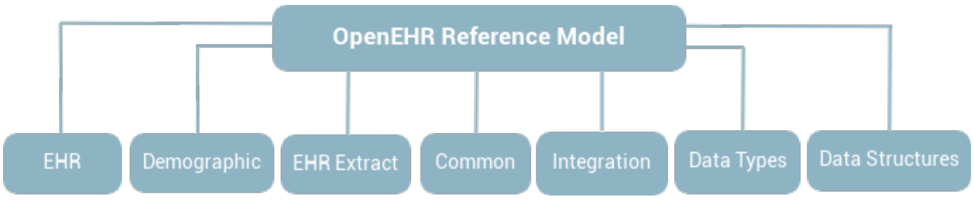

Figure 24: OpenEHR RM Overview [89].

Each openEHR IM package can belong to one of the two following categories:

- ***Domain-related***
  - EHR IM: The central component of openEHR architecture that includes the *EHR* and *COMPOSITION* packages, and defines the content and context semantics of the main EHR concepts: *COMPOSITION*, *SECTION*, and *ENTRY*. These classes are the core components of the EHR and resemble the classes in ISO 13606 standard and quite close to the layers of the same names in HL7 CDA 2.0.
  - EHR Extract IM: It defines how an EHR extract is built from compositions, demographic, and access control information from the EHR. The Extract supports "full openEHR", a simplified form for ISO 13606 integration, and an openEHR/openEHR synchronisation Extract.
  - Integration IM: It represents a tree-like representation of nonlinear legacy or external data by the *GENERIC ENTRY* subtype class, which are used in conjunction with clinical archetypes.

- Demographic IM: It defines generic categories like *PARTY*, *ROLE*, *CAPABILITY*, as well as contact information. This IM allows the separation of patient demographic information, providing more security in the protection of personal data.

- ***Generic***

  - Data Types IM: It includes a range of basic and clinically specific data types required for all kinds of health information. The data types reference model defines the following data type categories: Basic types, text, quantities, cncapsulated data, time_specifications and uri identifiers.
  - Data Structures IM: A set of structures that describe the content that will be consumed by the archetypes, such as single, list, table, tree and history types.
  - Common IM: It includes classes that are repeated in higher-level packages, such as the *LOCATABLE* and *ARCHETYPED* classes that link the RM to the archetypes.

EHR IM is divided into two packages: *EHR* and *COMPOSITION*, with the latter containing a main package, that is named by *ENTRY* and contains the *OBSERVATION*, *EVALUATION*, *INSTRUCTION*, and *ACTION* classes, which are subtypes of a clinical record and correspond to stages of a possible clinical process. They serve as the foundation for entry archetypes such as blood pressure, medication order, procedure, and smokeless tobacco summary among others. All of this information is stored in *DATA STRUCTURES* attributes, and it is also expressed in *DATA TYPES* attributes. Each clinical act is done by someone, and is associated to a specific subject in Demographic IM. This package also defines the actors, its information, and the roles that each intervenient plays in a clinical act, group, or organisation.

EXTRACT package is required to extract generic or specific content, such as clinical, administrative, and demographic data. Each *EXTRACT_REQUEST* must have its own specifications, e.g., *extract_type*, an attribute coded by the openEHR terminology group *extract content type*. Also, its *manifest* attribute inherits the class *EXTRACT_MANIFEST* to define the candidate entities,e.g., through the *EHR_ID*, and optionally top-level items, such as Compositions, to be included in the extract process. In a simple and succinct way, Figure 25 shows openEHR RM and its potential as a clinical standard to be deployed.

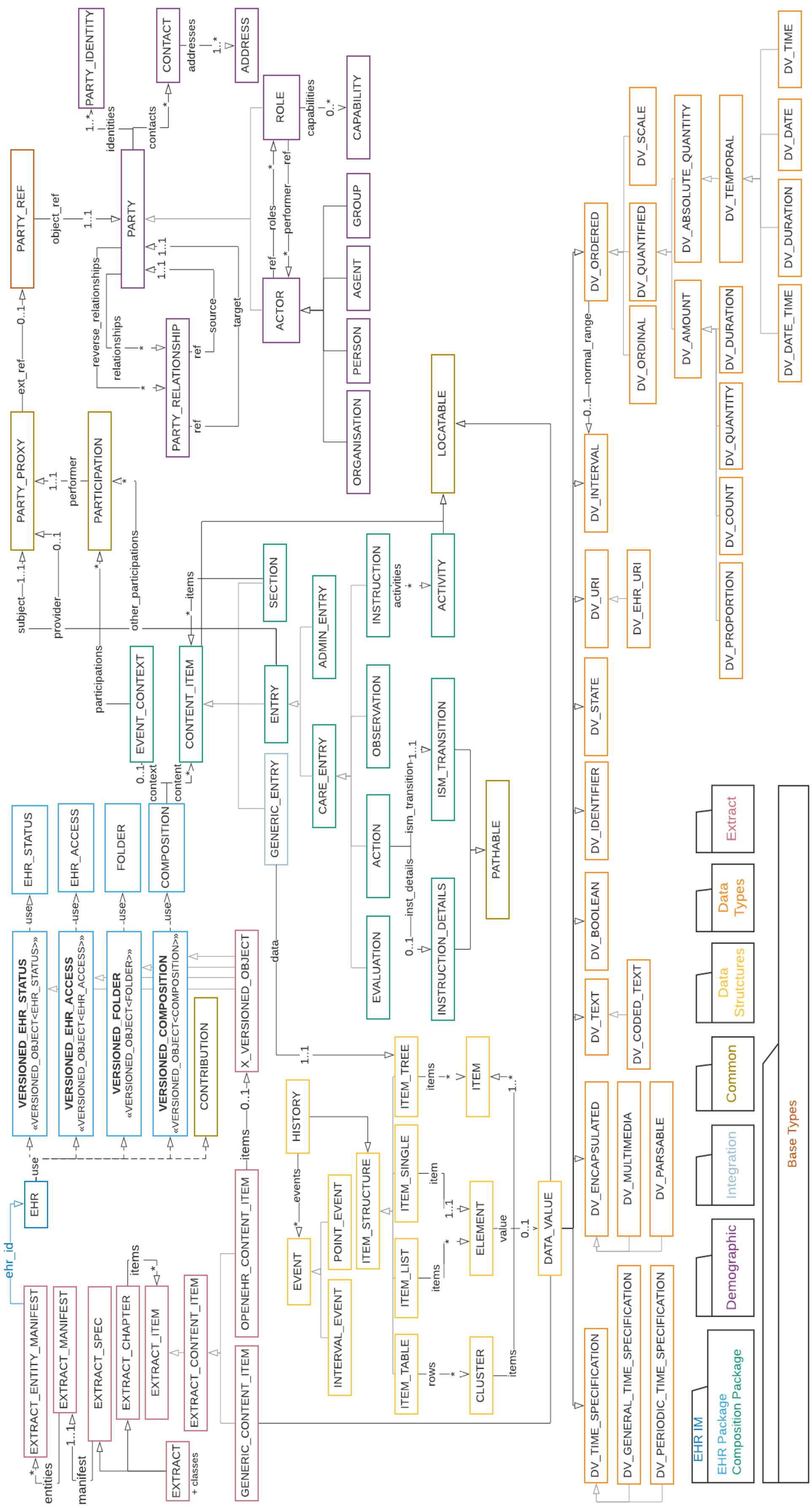

Figure 25: OpenEHR class diagram.

3.3.4.3 Standards implementations in world

The ISO EN 13606 healthcare standard enables the secure exchange and establishes semantic interoperability of EHR data by using archetypes to define the structure and semantics of the content. The *"Extraction of standardized archetyped data from EHR systems based on the Entity-Attribute-Value Model"* study focuses on the specific class of largely research-oriented EHR systems that are internally based on the Entity Value-Attribute Model. The researchers propose a new approach for extracting data from ISO EN 13606 archetypes using a standardized XML mapping system. Validation and evaluation of the new approach was performed using an EHR system used for clinical research at the Medical University of Vienna. Data defined by three different archetypes were used, the source documents were successfully extracted from the archetype-conformant ISO EN 13606 EHR extracts. Thus, the new approach proved that EHR data can be extracted from EHR systems based on Entity Attributes, although in order for this to happen, the internal data model of the EHR systems and the archetypes must overlap for a possible semantic mapping. In addition, systems must be able to export documents in a conventional format, including data and metadata that are mandatory in the archetype and ISO EN 13606 RM [32].

A German research entitled *"Transformation of microbiology data into a standardised data representation using OpenEHR"* presents an openEHR-based application for microbiological data, called by *openMibi*. It uses Archetype Query Language (AQL) queries to retrieve relevant data from the new openEHR repository. The application was built with the Angular framework and requests to the openEHR platform are done through Representational State Transfer (REST) APIs. On each interaction with the application, a new AQL-based REST request is sent, where the query parameters are dynamically modified depending on what was selected [124].

In 2016, Romanian researchers developed a platform for registering children's chronic diseases based on the openEHR specifications, ensuring the security and privacy requirements of data and the Romanian healthcare system. The prototype developed works on two levels, where the first is local and corresponds to a medical unit, while the second is in charge of the connection and communication within the entire Romanian national medical system. In order to ensure maximum semantic interoperability in the communication module, the ISO EN 13606 or HL7 v3 and/or HL7 v3 CDA standards were used. The authors concluded that the prototype was functional and applicable for medical and administrative purposes. However, its integration with LS and its regional and national scalability have not yet been achieved [41].

"A methodology based on openEHR archetypes and software agents for developing e-health applications reusing legacy systems" shows a new methodology for converting LS data through an automated agent-based system. The proposed approach is based on the use of software agents and openEHR archetypes. The new artefact was shown to be able to support interoperability between HIS by using openEHR archetypes to represent the information contained in the EHR clinical extracts. The case study demonstrated that we can move from an LS to an agent-based system seamlessly due to the employment

of openEHR archetypes in order to achieve semantic interoperability. The use and evaluation of the new system by healthcare professionals and patients showed its perceived usefulness and acceptance [83].

Further research can be obtained at the Zotero OpenEHR Group, which has over 300 publications classified into different categories covering implementations, modelling methodologies, and related standards.

3.4 Multi-Agent Systems

In the eyes of Distributed Artificial Intelligence (DAI), a MAS is defined as *"a loosely coupled network of problem-solving entities that work together to find answers to problems that are beyond the individual capabilities or knowledge of each entity"* [33]. In 2007, the Multi-Agent Systems Working Group of IEEE Power Engineering Society undertook a survey of key concepts and approaches in the area of MAS as applied to the field of Power Engineering. The group first defines what it means by Intelligent Agent and, subsequently, what defines a MAS. Based on Wooldridge's definition, which describes an agent as a software or hardware entity situated in a given environment, with the ability to react autonomously to changes in its surroundings, the cluster characterizes an intelligent agent based on three fundamental characteristics - Reactivity, Proactivity and Social Ability. Respectively, an intelligent agent is able to react to eventual changes in its environment, always focusing on a certain goal-oriented behavior, and is also able to interact with other agents, an indispensable feature during data transition. That said, two or more agents, with different objectives, interacting with each other, form a MAS [78].

João Moraes *et al.* defines an agent as an autonomous computational entity that makes decisions and performs its own tasks. For him, a software agent runs continuously and autonomously in a given environment and is capable of intervening without requiring human intervention. At the level of its intelligence, an agent can be classified as cognitive, if it has significant cognitive capabilities to perform its actions, or reactive. Although the same environment contains several agents, each one has its purpose and characteristics, guaranteeing its autonomy [83].

In *"A Decoupled Health Software Architecture using Microservices and OpenEHR Archetypes"*, the authors propose a Decoupled Healthcare Software Architecture (DHSA), based on the development of three components, a tool, and a formal metric. The components called Connector, Container, and Archemicro (Archetype-based microservice), make the DHSA, which is dynamically generated by the Microservice4EHR tool algorithm. For testing and evaluation of the new artefact, an old software used in Brazilian hospitals was migrated to the new DHSA system. Afterwards, the formal Archetype-based metric Software Architecture Coupling (ASAC) was used to define the coupling level of the proposed architecture. As a result, the new approach presented increases the software decoupling index in health by 66.6%. With this result, the researchers concluded that healthcare systems will increasingly benefit from the development of software architectures that keep the system functional even if a component of the system is broken [1, 5, 72].

In the present project, MAS proves to be of high importance and usability, as a new EHR data repository will be created, meeting the ideals of healthcare institutions under study, and relevant data present in the heterogeneous and different HISs will have to be synchronized to new EHR system developed. This synchronization will rely on the action of appropriately programmed intelligent agents. But what is the best type of data repository to be created in healthcare? In *"Ambient Intelligence via Multiagent Systems in the Medical Arena"*, the authors developed a MAS in the healthcare environment, called *GENsis*, translating into an Agency for Integration, Archiving and Dissemination of clinical information. This agency involves proactive intelligent electronic agents, responsible for the communication of different HISs, sending and receiving information, such as clinical reports and prescriptions, exams in image format, among others, managing and storing information in a certain time period. At the level of data storage in healthcare, Relational database management system (RDBMS) have been frequently used by institutions and are defined as a database management system based on a relational model. Through this model, the use of OnLine Transaction Processing (OLTP) tools and the consequent Extraction, Transformation and Loading (ETL) process, allows the efficient use of data, from its analysis to the identification of useful trends, allowing preventive action, thus supporting clinical decision making. In short, healthcare institutions store their data in different Data Warehouse (DW), which comes from several HIS sources. This DW, with its relational structure, allows relevant information to be materialized, transforming it into interpretative Views, such as Dashboards of important indicators of a given healthcare unit [73]. Following this, over the last few years, the concept of BI has been introduced as a set of intelligent techniques used to promote data interpretation, identifying, for example, key indicators, diagnostic patterns or even risk factors.

3.5 Data Science

The exponential growth in the amount of data produced daily, along with its complexity, has triggered the development of techniques and methodologies for them to be synthesized and subsequently analyzed for various purposes. In this sense, since the mid 1970's several researchers have been defining the concept of DS as a discipline that can be employed in the most diverse domains, namely in the performance analysis of large companies. DS consists of a multidisciplinary technological area capable of dealing with problems such as the size, speed and enormous diversity of information sources, in support of decision making. Initially "Data science was the science of dealing with data, [...] while the relationship of data to what they represent is delegated to other fields and sciences [...],"and its concept has evolved over the years into a more comprehensive and complex universe that deals with the most diverse areas, such as data and business analytics. In the Cleveland point of view, DS emerges in the context of Computer Science (CS) and Data Mining (DM) [22]. In this follow-up, the term DS started to be addressed in the scope of AI, specifically in its ML area, thus combining knowledge discovery, CS and Mathematical Statistics (MS).

In *"Advances in Knowledge Discovery and Data Mining"*, the authors define the concept of Knowledge-Discovery in Databases (KDD) as the process of identifying new patterns of interest resulting from data exploration [37]. DS encompasses fields such as CS, MS, DM, and Big Data Analytics in processes such as outlining workflows, defining robust statistical methods, using predictive techniques and algorithms, and innovative computational solutions. On a time scale, BI processes monitor historical information that allows organizations to find patterns and trends for continuous improvements in their management. It encompasses the concepts of DW and DS.

Data science is increasingly being used in healthcare to extract knowledge from noisy unstructured data and convert crucial data from documents into useable formats. In the transition from manual to electronic medical records, the development of various intelligent systems based on supervised and unsupervised ML algorithms has been essential [38, 88]. Unsupervised systems are increasingly focusing on unstructured clinical writings on paper or in a database, which include vital data for the creation of decision support systems [80].

In *Natural language processing of symptoms documented in free-text narratives of electronic health records: a systematic review*, the authors wrote a review article on NLP publications of documented symptoms in free text. Twenty-seven articles were selected based on the predefined criteria and the main results are presented in figure 26 [64].

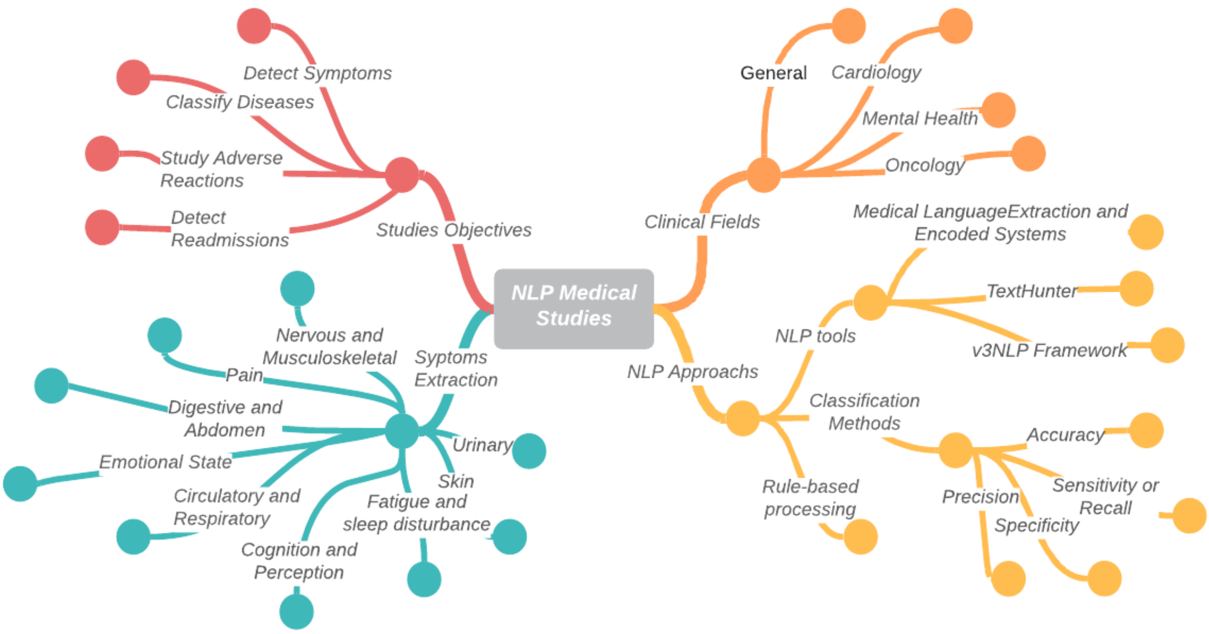

Figure 26: NLP Medical Studies map. Based on [64].

In order to recover relevant clinical data to improve the quality of care provided, the approach published in [125] proposed an openEHR-based pipeline for extracting and standardizing unstructured clinical data through NLP, for pediatric medical histories. The authors developed a pipeline for transforming clinical free-text pediatric records into structured and standardized formats, choosing and modeling appropriate openEHR structures for use as standard clinical information models. They also established mapping rules between the NLP output and the openEHR structures, as well as a German dictionary with key text markers that served as an expert knowledge base for an NLP pipeline. The first method evaluation with 50 manually annotated patient histories achieved 97% precision and 94% recall, which are very good results.

In order to achieve the most reliable results with NLP in a healthcare context, biomedical information resources are crucial, providing several controlled vocabularies with terminological knowledge. Unified Medical Language System (UMLS) is a tool that combines a set of files and software to integrate and distribute biomedical vocabulary and standards to enable interoperability between IS [115].

SNOMED CT, ICD and LOINC terminologies can also be used as sources of clinical knowledge for NLP systems. Additionally, some open-source biological databases are an excellent resource for retrieving biomedical data [20].

3.6 Conclusion

This chapter was written during the first stage of the research project, which was titled "Literature Review". A search for the main concepts relevant to the subject under study was collected, and some of the research work developed was presented. The "Background"chapter also highlighted the relevance of structured information in a simple way, through the complex definition of a HIS, thus answering RQ1. This chapter also discussed the benefits of adopting globally recognized standards and protocols for a consistent development of a new information flow, both structurally, syntactically, and semantically, in order to address RQ2 and RQ3.

Chapter 4

The Development of a Interoperable Healthcare System

4.1 Adopted Technologies and their Main Advantages

Software development methodologies provide a reliable technical reference for developing OpenEHR systems. Their frameworks are part of the software specifications of new EHR systems, constraining the object-oriented RM to be used. Therefore, there is a direct relationship between the modeled structures and the technical implementations of the HISs, implying constant interaction between modeling and software development.

This section provides a brief summary of each technology used to construct the proposed system, as well as the primary benefits that support the decision to adopt it.

4.1.1 Databases

A database is *"an organized collection of structured information, or data, typically stored electronically in a computer system"* [121]. In a nutshell, this technology enables the development of any IS, allowing to create, read, update and delete data in a quick, secure and efficient way. This represents usually one of the main pillars of each and every IT solution, in a way that can be used in many different use cases and purposes in the IS world.

During the technology decision phase of this specific project, it was really important to study the legacy databases that already exist in the healthcare environment and compare the most suitable options regarding the selection of a modern and powerful database and identify all the advantages and disadvantages of the most popular types of databases existing nowadays.

4.1.1.1 Relational VS Non-Relational Databases

The first topic to take into consideration, when deciding which database to use, is definitely the choice between RDBMS and Non-relational database (NoSQL). In regards to RDBMSs, they have been highly used for decades, and are still dominating the market share but this advantage has decreased during the last years due to the growth of the NoSQL databases. This last ones are becoming more and more popular due to its capabilities for unstructured data, that has been extremely needed between data engineers and data scientists for many business cases like, data analytics, artificial intelligence, machine learning and other emerging technologies. Also, the Non-Relational approach have gained popularity in the IT world solving problems such as the large number of joins and auto-joins, frequent change schemas, and slow queries, which can limit the usability the products that the RDBMS supports. Table 3 represents the NoSQL and Structured Query Language (SQL) major differences.

Table 3: Main differences between Non-relational and Relational Databases [65].

NoSQL	**SQL**
Open source	Costs associated
Fast and Flexible	Slower and Structured
Easy Replication	Complex Replication
Horizontal Scalability	Vertical Scalability
Document, key-value, graph based	Table based
Simple queries	Complex queries
Non-centralized	Centralized

As described in Table 3, the NoSQL approach has a horizontal scalability, which facilitate the database segmenting across multiple servers. So, when data storage needs to grow up, it is only necessary buy economic servers and integrate them. In the RDBMS case, scaling up necessitates the purchase of new expensive hardware [69].

However, relational databases, such as Oracle, MySql, SQL Server, among others, are still heavily used in healthcare institutions.This decision should be determined on the requirements of the systems that the database will support, the costs that each institution could pay, and the interoperability with other current LS.

MySQL is an open-source RDBMS maintained by Oracle Corporation. According to Stackoverflow's 2021 report, in the *"Databases"* category, MySQL remains the most widely used SQL language in the world, with 50.18% of the votes [111]. Its data security, scalability, high performance due to the storage structure, and its open source flexibility contribute to MySQL's success.

In *"Native JSON Datatype Support: Maturing SQL and NoSQL convergence in Oracle Database"*, the authors showed how the JavaScript Object Notation (JSON) datatype is fully integrated with the Oracle Database systems and how to transform them into a mature platform to serve both SQL and NoSQL paradigms. They described the JSON format as a schema-flexible data model which can expand to store new attributes without modify the inheret schema. Also, JSON is human readable, fully self-contained and easily consumed by object-oriented programming languages. JSON is an extension of the BLOB and CLOB datatypes, that was only recognized as a datatype in Oracle Database 20c version. Also, this study mentionated the SODA Oracle concept. As a major company, Oracle developed a simple document access API over their database, as a pure JSON document store layer. As a result, the users have a NoSQL-style access layer, based on documents and a key-value approach. Besides that, queries can be performed in json format, called by QBE. This SODA API is available in Java, REST, PL/SQL, Node.js, Python, and C programming languages [70, 105]. The example showed below is a QBE to a specific collection in order to obtain all documents that contains the string *"Transplante Cornea"* in *"identities.party_identity.name"* path.

Listing 1: SODA QBE example.

```
{
 "identities.party_identity.name": {
  "$contains": "Transplante Cornea"
 }
}
```

The development of SODA API has demonstrated continuous adaptability of solutions over time by Oracle. This can be a good solution for institutions and companies whose migration process from relational to non-relational approach can be very painful.

Both SQL and NoSQL approaches were developed in this research project. To support the processing of openEHR structures and all of its derived information, a MySQL RDBMS was developed. Furthermore, in accordance with the technologies employed in the research facility, a NoSQL Oracle method was developed via its new JSON document layer dubbed SODA.

4.1.2 Backend and Frontend Technologies

Created on 2009, Node.js is a scalable network application builder that uses an asynchronous event-driven JavaScript engine. Also, Hypertext Transfer Protocol (HTTP) is a first-class *citizen* in this framework, having been created specifically for streaming and low latency, making it an excellent choice as the basis for a web library or framework [4].

Aside from being simple to learn and very adaptable, Node.js is an ideal choice for application development, owing to its high efficiency (it uses the Google V8 engine to compile JavaScript code) and non-blocking asynchronous behavior, resulting in faster code execution) [76].

Node.js was ranked consecutively in 2019 and 2020 as the most popular technology in the *Other Frameworks, Libraries, and Tools* category in Stackoverflow's *"Developer Survey Results"* [109, 110].

In order to communicate with the different databases developed, it was necessary to use their respective packages to make this possible. To guarantee the security of access to the information, a **middleware** was developed, through the *Express.js* minimalist web framework for Node.js. This middleware performs the validation of a given token so that it is possible to obtain the answer to each request made [35].

Sequelize is a promise-based Node.js Object-relational Mapper (ORM) for databases such as Postgres, MySQL, MariaDB, SQLite and Microsoft SQL Server. It supports transactions, relations, eager and lazy loading, and read replication. Using an ORM it's possible to achieve better database synchronization, helping with data manipulation and querying through the use of objects from the database. Also, through the use of ORM, the SQL queries are optimized, making them easy to reuse and maintain [102].

In terms of frontend technology, the React.js JavaScript library was chosen for this project, in order to develop the openEHR governance web platform. With 40.14% of the votes in Stackoverflow's 2021 report, in the *"Web Frameworks"* category, this framework was the most commonly used. The ability to create massive Web applications that can change over time without reloading the entire page is the library's key feature. Some packages were used throughout the development of the web platform, such as User Interface (UI) design packages (antd and tailwindcss), HTTP clients (axios and fetch), navigational packages (react-router-dom), among others [6, 9, 39, 98, 99, 112].

4.2 The System Architecture developed

Figure 27 represents all project architecture components and their respective tecnhologies and programming languages used.

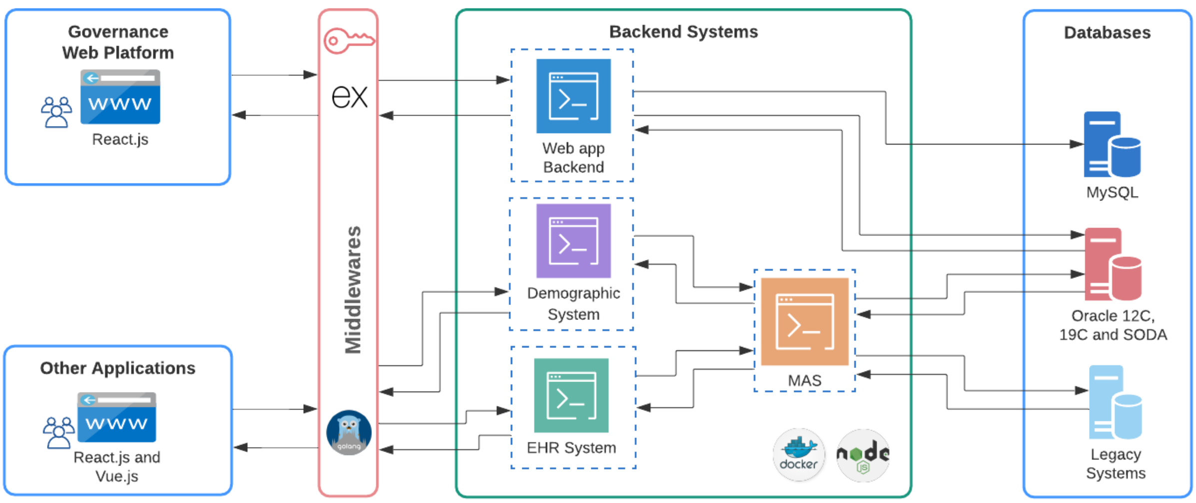

Figure 27: Project architecture and implemented tecnhologies.

First, an openEHR framework governance platform and all its underlying components were developed. The UI was developed in React.js and improved over the last 4 years through constant feedback from users. To make this possible, it was necessary to develop a backend module in Node.js. Between the interface and the backoffice system, a middleware system was developed with Express.js to ensure the security of obtaining data. In turn, this data was distributed by the backoffice methods to a MySQL RDBMS repository and to the Oracle repository. The MySQL repository was developed to ensure a backup of the openEHR frameworks developed over time. All actions performed through the governance platform are reflected only in the Oracle repository. The following subchapters will detail the main developments in each subcomponent of the system architecture presented above.

4.2.1 Web Application Backend

The backend developed for the governance web platform consists of several methods. Sets of methods were developed to handle user registration and authentication, governance request management, templates, forms, data refsets, DLM structures, and work backlog. However, one of the main and most complex components developed is the interpretation of the OPT structures and their consequent manipulations.

4.2.1.1 OPT to JDT Conversion

In order to implement the modelled openEHR structures, it was necessary to develop an algorithm to convert the *.opt* extension that contain the OPT developed into a clearer and easier to understand format. For a better understanding of the structure of a template (file with the extension *.opt*), its main structure is shown in Figure 28 .

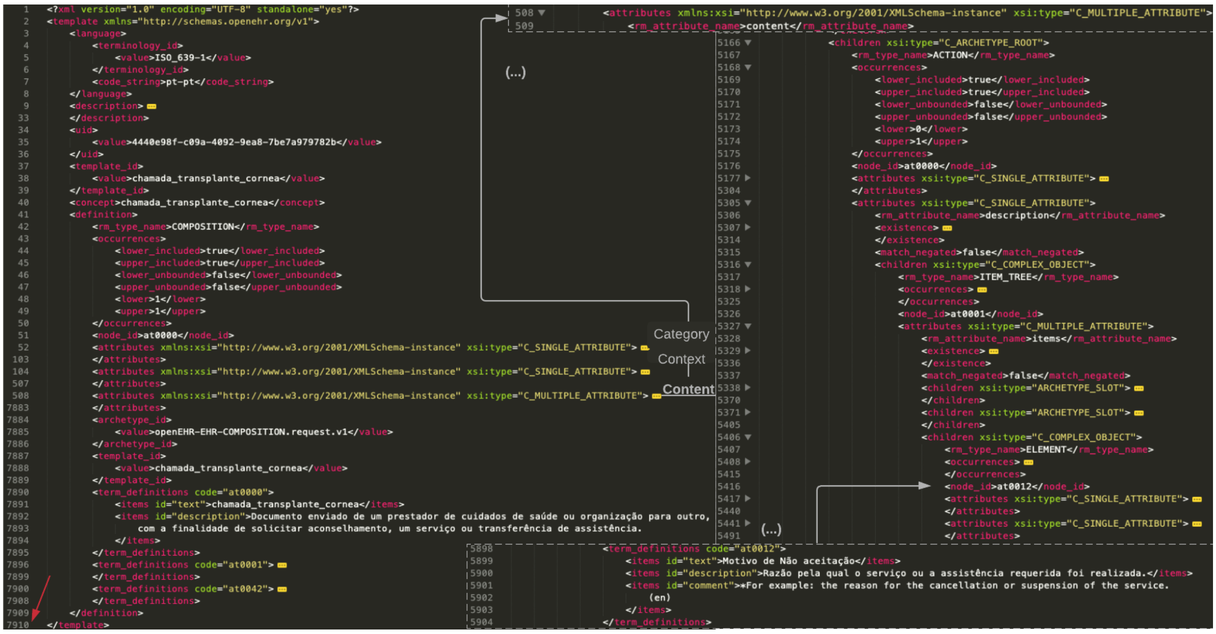

Figure 28: OPT structure for the patient's call for cornea transplant.

Figure 28 shows the template structure for the patient's call for cornea transplantation. The templates have a fixed structure, consisting of language, description, uid, template_id, concept, and definition components. Inside the *definition* component, the base archetype of class *COMPOSITION* is declared, which will originate the remaining child elements. This class is subdivided into three main attributes - Category, Context, and Content - the latter being the most complex in terms of component proliferation. As it can be seen in schema, the attribute *Content* has a child attribute of type *C_ARCHEYPE_ROOT*, which means that the attribute is an archetype of type *ACTION*. In turn, the class *ACTION* inherits the class *CARE_ENTRY* and has the attributes time, ism_transition, instruction_details, and description. On the right side of the figure, it is possible to see the expanded description attribute and its constituents in the *ITEM_TREE* structure. This structure is composed of multiple items that can be of various types, including the type *ELEMENT* whose description is obtained by binding its node_id (at0012) to the term_definitions (*"Motivo de Não aceitação"*). As it can be seen, interpreting a OPT structure is kind of difficult. As a result, the necessity emerged to create an algorithm that convert OPT into smaller, easier-to-manipulate structures, based on recursive methods designed to produce a simpler structure in JSON format, named by JDT. The following Figure 29 demonstrates the result of the OPT to JDT conversion algorithm for calling patients for corneal transplantation.


```json
{
    "themeColor": "default",
    "templateName": "Chamada para Transplante de Córnea",
    "templateId": "4440e98f-c09a-4092-9ea8-7be7a979782b",
    "rm_type_name": "COMPOSITION",
    "occurrences": { …
    },
    "archetypeId": "openEHR-EHR-COMPOSITION.request.v1",
    "path": "[openEHR-EHR-COMPOSITION.request.v1]",
    "node": {
        "code": "at0000",
        "text": "Chamada para Transplante de Córnea",
        "description": "Documento enviado de um prestador de cuidados de saúde ou organização para outro"
    },
    "items": [
            { …
            },
            { …
            },
            {
                "occurrences": {
                    "lowerOccurrences": 0,
                    "upperOccurrences": 1
                },
                "xsi_type": "C_ARCHETYPE_ROOT",
                "archetype_id": "openEHR-EHR-ACTION.service.v0",
                "path": "[openEHR-EHR-COMPOSITION.request.v1]/[openEHR-EHR-ACTION.service.v0]",
                "node": {
                    "code": "at0000",
                    "text": "Identificação",
                    "description": "Actividade clínica geral direccionada para o utente"
                },
                "items": [
                    { …
                    },
                    {
                        "path": "[openEHR-EHR-COMPOSITION.request.v1]/[openEHR-EHR-ACTION.service.v0]/[at0012]",
                        "node": {
                            "code": "at0012",
                            "text": "Motivo de Não aceitação",
                            "description": "Razão pela qual o serviço ou a assistência requerida foi realizada."
                        },
                        "dataType": "DV_TEXT",
                        "occurrences": {
                            "lowerOccurrences": 0,
                            "upperOccurrences": 1
                        },
                        "value": null
                    },
                    { …
                    },
                    { …
                    }
                ],
                "data_type": "Title"
            },
            { …
            }
        ]
}
```


Figure 29: JDT structure for the patient's call for cornea transplant.

It is observable that the structure has decreased considerably (7910 to 541 rows) as well as its complexity. The JDT consists of the key information from the template and the maximum possible syntax is maintained. The main node is the *openEHR-EHR-COMPOSITION.request.v1* archetype and is described as *"Chamada para Transplante de Córnea"*. Following this, its components are contained in the array called *items*, as are all child components at whatever level. As in figure 28, the *ACTION* archetype is expanded and you can see the simplification of its content without loss of relevant information.

Algorithm 1 ProcessOptNodes algorithm

function processNode($rm\_type, path, terms, node$)
    if $node.children \neq undefined$ **then**
        if $length(node.children) == 1$ **then**
            return processChildren(rm_type,path,terms,node.children)
        else if length($nodeNode.children$) > 1 **then**
            for $c \leftarrow 0, length(node.children) - 1$ **do**
                return processChildren(rm_type,path,terms,node.children[c])
            end for
        end if
    else if $node.attributes \neq undefined$ **then**
        return processNode($rm\_type, path, terms, node.attributes$)
    end if
end function

Algorithm 1 describes the recursive function that processes each node of the OPT file. Its input parameters correspond to the RM type of node, its path inside the OPT structure, the terms for binding between the codes and descriptions, and the node to be processed. Throughout the OPT processing, the RM classes of each archetype were saved, as can be seen in figure 30. Also, this RM was processed through its JSON specification schema using the most recent 1.1.0 release [54].

```
{
  "": {
    "path": "openEHR-EHR-COMPOSITION.request.v1", //Main archetype
    "rm_type": "COMPOSITION"
  },
  "items.0.2": {
    "path": "[openEHR-EHR-COMPOSITION.request.v1]/[openEHR-EHR-ACTION.service.v0]",//Archetype Identificacao
    "rm_type": "ACTION"
  },
  "items.0.3": {
    "path": "[openEHR-EHR-COMPOSITION.request.v1]/[openEHR-EHR-ACTION.service.v0]", //Archetype Aceitação
    "rm_type": "ACTION"
  },
  "items.0.1.items.0": {
    "path": "[openEHR-EHR-COMPOSITION.request.v1]/[openEHR-EHR-SECTION.adhoc.v1]/[openEHR-EHR-ACTION.service.v0]", //Archetype "1º Contacto Telefónico"
    "rm_type": "ACTION"
  },
  "items.0.1.items.1": {
    "path": "[openEHR-EHR-COMPOSITION.request.v1]/[openEHR-EHR-SECTION.adhoc.v1]/[openEHR-EHR-ACTION.service.v0]", //Archetype "2º Contacto Telefónico"
    "rm_type": "ACTION"
  },
  "items.0.1.items.2": {
    "path": "[openEHR-EHR-COMPOSITION.request.v1]/[openEHR-EHR-SECTION.adhoc.v1]/[openEHR-EHR-ACTION.service.v0]", //Archetype "3º Contacto Telefónico"
    "rm_type": "ACTION"
  },
  "items.0.1.items.3": {
    "path": "[openEHR-EHR-COMPOSITION.request.v1]/[openEHR-EHR-SECTION.adhoc.v1]/[openEHR-EHR-ACTION.service.v0]", //Archetype "4º Contacto Telefónico"
    "rm_type": "ACTION"
  }
}
```

Figure 30: RM types of each archetypable item for the patient's call for cornea transplant.

Parallel to the development of the new JDT structure, its associated items were saved with their relevant data for future use, and an auxiliary JSON structure called ***InitialComposition*** was developed and illustrated in figure 31 to complement JSON in some aspects.

```
[
  {
    "itemName": "chamada_transplante_cornea",
    "itemPath": "",
    "TaskVisible": [
      "All"
    ],
    "TaskEditable": [
      "All"
    ],
    "TaskInReport": [
      "All"
    ],
    "Functions": [],
    "InternalFunctions": [],
    "ruleMandatory": [],
    "ruleVisibility": []
  },
```

```
  {
    "itemName": "Motivo de Não aceitação",
    "itemPath": "items.0.2.items.1",
    "TaskVisible": [
      "All"
    ],
    "TaskEditable": [
      "All"
    ],
    "TaskInReport": [
      "All"
    ],
    "Functions": [],
    "InternalFunctions": [],
    "ruleMandatory": [],
    "ruleVisibility": []
  },
```

Figure 31: *initialComposition* auxiliar structure for the patient's call for cornea transplant.

This structure contains all JDT components, including archetypes, clusters, and elements, and its major purpose is to help in the control process of the dynamic GUI built to generate openEHR forms.

In short, after submitting a new OPT to the governance platform, several methods are invoked and new structures are generated and instantiated. For a better understanding, the following UML sequence diagram was developed.

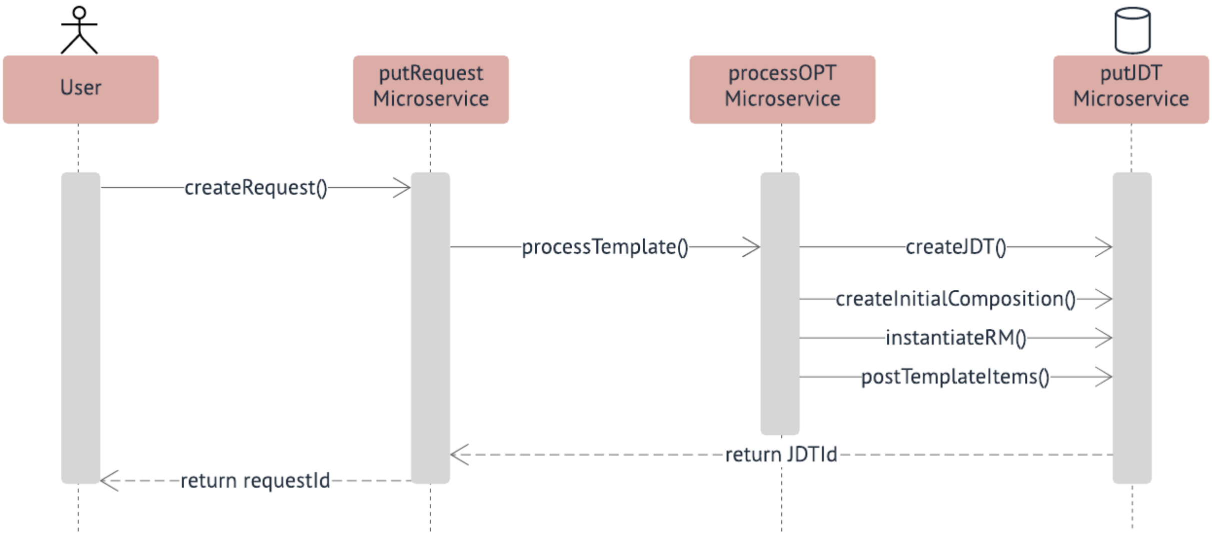

Figure 32: UML sequence diagram of the OPT process.

4.2.2 Web Application Frontend - *AIDAEHR*

With the main goal of governing openEHR frameworks and everything that underlies them, a web platform subdivided into several modules was developed.

The authentication module is "Sign In"and "Sign Up"and all subsequent operations are performed in the backend. The user type "Administrator"can change the permissions of each user on the platform, having to primarily approve each new user enrolled. The user can create his own account, and if he forgets his password, he receives in his e-mail an activation code to create a new password. Note that all user passwords are encrypted when stored in the database.

One of the independent modules is the "Requests"module, where the user can submit the modulated OPT structure so that it can then be processed by the backend process and consequently submitted to the openEHR governance system. This module is accessible by all types of users, although with different permissions. Only the user "Modulator"has the permission to evaluate the submitted request for its importance and modulation and to reject the request or even remove it from the list of requests.

After the "Requests"module evaluates each request, the template is stored in the system, and the modulator can remove or replace it. As it can be seen in figure 33 , the "Templates"module is dependent on the "Requests"module.

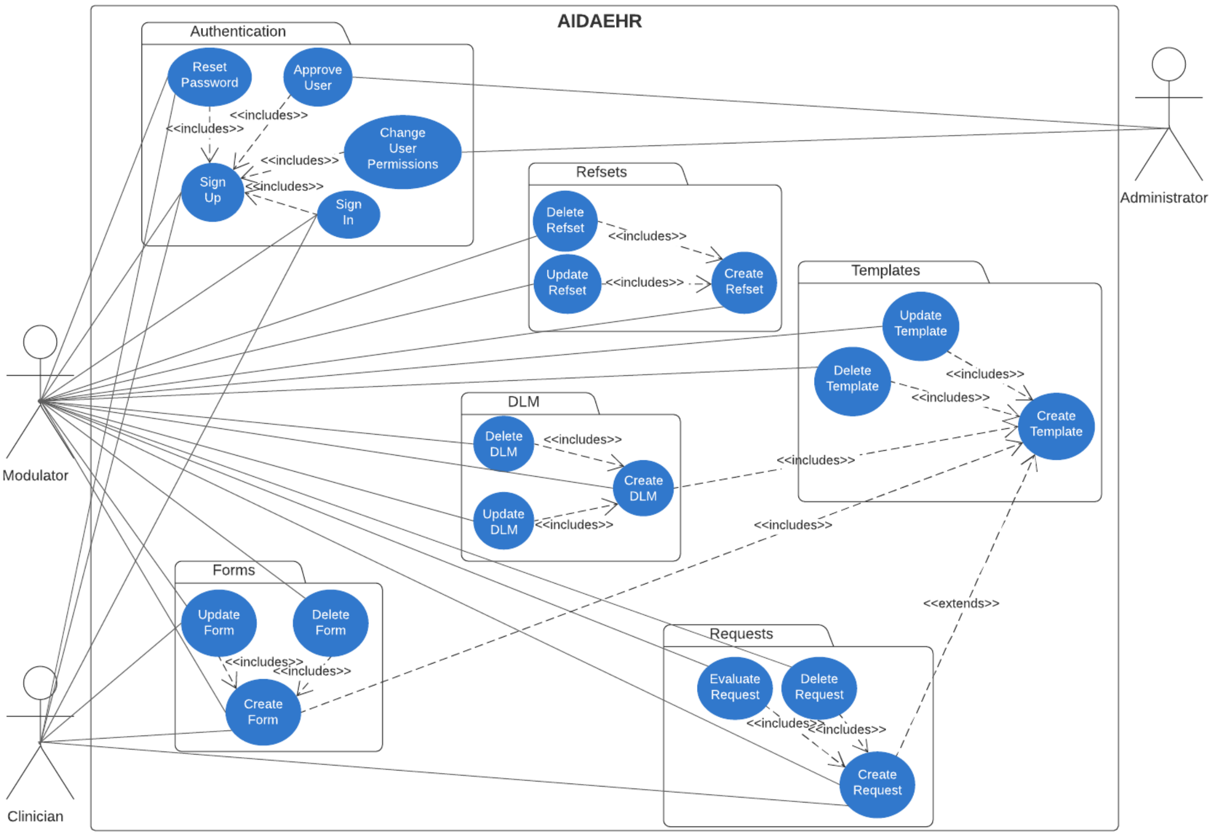

Figure 33: UML use case diagram of the governance web platform.

The other independent module is called ”RefsetBuilder”and allows you to create data sets to answer the *defining_code* attribute of the *DV_CODED_TEXT* datatype. This datatype can have internal concepts, where they are created when modulating the template, or it can have external concepts created ”a posteriori”. Refsets can be classified as static or dynamic, and the static ones, after being created, can only be manually changed. This type of refset can be created based on several data sources:

- **SNOMED CT**: through a developed internal server that processes its ”Expression Contraint Queries” and returns the set of concepts as a response;
- **ICD**: through the diagnoses registered by each institution;
- **MCDT**: through the procedures, exams, and clinical analyses prescribed by each institution.

The dynamic refsets are built by creating *ad hoc* queries to the different databases available at each institution. Each dynamic refset stores the developed query and the respective database and when the refset is called, the query is executed and its result is returned. This type of refset is very useful for value sets that are constantly changing.

The ”FormBuilder”module allows the management of forms generated through the existing templates in the ”Templates”module. This module is very useful for developing several forms from the same template, for different purposes. Through a package that transforms the JDT structure into dynamic GUIs, it is possible to edit each form and associate the developed refsets with certain fields, or even hide a field or section of the form. After making the various changes available in the package, it is possible to save the changes to the edited form or create a new form with the new changes. Similar to other modules, the ”Clinician”user only has permissions to create new forms, while the ”Modulator”profile can perform all available operations.

The above-mentioned package, in addition to transforming the *JDT* structure into a dynamic *GUI*, is equipped with devices to provide intelligence to the forms. For this, it was necessary to develop a ”DLMBuilder”module that aims to build small logical decision modules through the creation of rules and conditions. The result of the combination of these rules and conditions is expressed through a structure interpretable by the GUI package. The ”DLMBuilder”area also allows the management of the DLM blocks already created. It should be noted that the engine which interprets each DLM was not developed in the scope of this doctoral thesis, although it is mentioned for its symbiosis with the platform and described here.

In terms of work management and task governance, the ”Backlog”module was developed in which users can create tasks, associate them with other users and deadlines. All tasks are presented in the form of a Kanban board for easy viewing and understanding of the status of developments.

4.2.3 Demographic and Clinical Information System

Developing the architecture of a new clinical IM is a complex process consisting of several adjacent areas. For this purpose, two core modules were developed to make the implementation of the standard in a healthcare institution possible: the DIM and the CIM, following the openEHR development guidelines.

4.2.3.1 Demographic System Developed

For the whole ecosystem to be in symbiosis, openEHR structures were modeled to represent each *ORGANISATION*, *GROUP*, *PERSON* and *ROLE*, as main classes of the DIM system. The DIM RM can be observed in figure 34, an excerpt taken from figure 25. Each of these classes corresponds to the healthcare institutions, teams, healthcare and administrative professionals, patients, and roles played by each defined *ACTOR*.

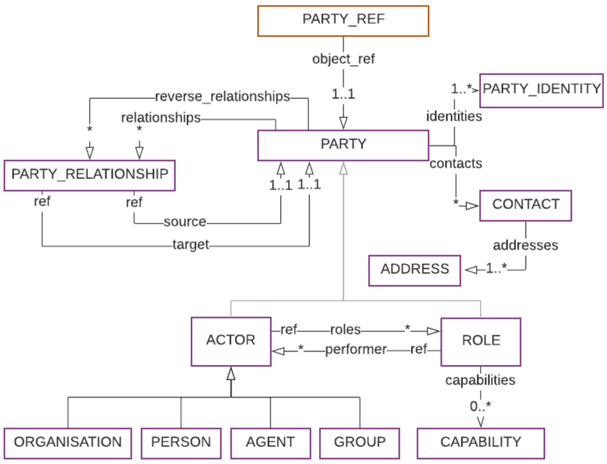

Figure 34: OpenEHR RM specification of DIM.

In order to obtain the maximum advantage of its RM, some templates were developed based on demographic archetypes. Because the *LOCATABLE* class is inherited by the majority of the AM package classes, the knowledge modeled in it is as valuable as the RM that supports it. This class is used to enter data into the RM, allowing clinical or administrative data to be combined with registration and audit data in the system.

The majority of the templates were created using Better's Archetype Designer platform [25], although on some levels, this tool becomes unable to respond to a need for new demographic structures, such as from the *ROLE* demographic class, and only permits the creation of demographic templates from the *PERSON* class, which does not contain the same RM data. As a result, the missing openEHR structures were modelled using Ocean Informatics' Template Designer Desktop application [113].

Some of the archetypes provided in various worldwide Clinical Knowledge Manager (CKM) instances were reused and adapted, while others were created from new [23]. Table 4 lists all of the archetypes and templates that were used and created to build the DIM.

Table 4: Templates modeled in OpenEHR and the archetypes reused.

Template	**Archetypes**
Organisation.opt	openEHR-DEMOGRAPHIC-ADDRESS.address.v0 openEHR-DEMOGRAPHIC-ADDRESS.electronic_communication-provider.v0 openEHR-DEMOGRAPHIC-ADDRESS.electronic_communication.v0 openEHR-DEMOGRAPHIC-PARTY_IDENTITY.organisation_name-CHUP.v0 openEHR-DEMOGRAPHIC-PARTY_IDENTITY.organisation_name.v0 openEHR-DEMOGRAPHIC-CLUSTER.provider_identifier_chup.v0 openEHR-DEMOGRAPHIC-PERSON.organisation_demographic.v0 openEHR-EHR-CLUSTER.telecom_details.v0
Care_team.opt	openEHR-DEMOGRAPHIC-ADDRESS.electronic_communication.v0 openEHR-DEMOGRAPHIC-PARTY_IDENTITY.team_name.v0 openEHR-DEMOGRAPHIC-PERSON.group.v1
Person.opt	openEHR-DEMOGRAPHIC-ADDRESS.address.v0 openEHR-DEMOGRAPHIC-ADDRESS.electronic_communication.v0 openEHR-DEMOGRAPHIC-PERSON.person.v0 openEHR-DEMOGRAPHIC-ITEM_TREE.person_details.v0 openEHR-DEMOGRAPHIC-PARTY_IDENTITY.person_name.v0 openEHR-EHR-CLUSTER.telecom_details.v0
Role-team_member.opt	openEHR-DEMOGRAPHIC-ROLE.team_member.v0 openEHR-DEMOGRAPHIC-CAPABILITY.individual_team-functions.v1
Role-organisation_member.opt	openEHR-DEMOGRAPHIC-ROLE.employee.v0 openEHR-DEMOGRAPHIC-CAPABILITY.individual_credentials.v0
Role-organisation_patient.opt	openEHR-DEMOGRAPHIC-CLUSTER.person_identifier.v0 openEHR-DEMOGRAPHIC-ROLE.patient.v0

The underlined archetypes were the base structures from which each template was built. Note that they were all modeled with classes from the demographic model, with the exception of using classes *CLUSTER* and *ITEM_TREE* to complete the record with additional information such as address or email details.

Therefore, the first template, called *Organisation.opt* aims to represent a health institution and other external institutions with which it may collaborate. Characteristics such as address, telephone and electronic contacts, identification, among others, define a given healthcare organization.

Like many companies and organizations, openEHR suggests that health care be provided by members of certain local or cross-departmental teams. For this, it was necessary to develop the *Care_team.opt* template, which is based on the *GROUP* class of the RM and also has their identification and some other relevant details. Also, the *Person.opt* template was developed to define all the actors in an institution, whether they are healthcare or administrative professionals, patients or their family members, with all their most pertinent information.

For the system to recognize the role of a particular person in an institution, these classes described above have to be related through the *ROLE* class, and the last three templates of the table were developed. It should also be noted that, except for the *Role-organisation_patient.opt* template, the other two templates use the *CAPABILITY* class to define the roles that a given member has in an institution and a team.

After the more technical explanation of the specification, the following figure 35 shows an easy-to-understand example of the implemented system.

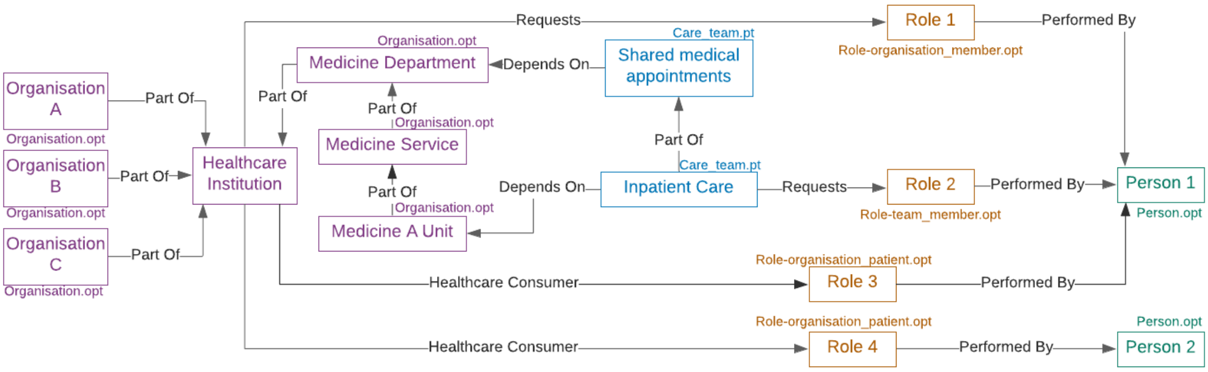

Figure 35: A DIM system scenario illustration.

Each entity - organisation, care team, role, person - is supported by its own openEHR structure, and they link with another via the RM *PARTY_REF* class with different meanings. It was also decided to completely separate the entities in order to have more control over each one, taking advantage of the potential of the RM class of *PARTY_RELATIONSHIP*. Besides that, the existence of a hierarchical structure from the principal organisation to its underlying and dependent organisations can be detected using this architecture.

The main organisation is also divided into departments, each of which is affiliated with distinct clinical services, and each of these services is associated to different care teams in different contexts. However, one of the most critical decisions was how to model the patient, the employees of the institution, and the members of each care team. Only one person structure was built, which provides identification to the subject independent of their relationship with the institution. Consequently, these relationships are clearly specified independently and linked to their source and target via roles. Also, these roles have certain associated capabilities that intrinsically define the credentials of a specific role.

Figure 35 demonstrates that the Person 1 is linked to two roles: Role 1 defines the actor as an employee of the healthcare organization, and Role 2 defines the actor as a member of the Medicine A inpatient care team. Furthermore, the same Person 1 is associated with Role 3, which defines him as a patient at the same healthcare facility. Person 2, on the other hand, only has Role 4 associated with him, indicating that he is a patient of the institution. In the situation presented, organizations A, B, and C are part of the main healthcare institution, and other *ORGANISATIONS*, which correspond to departments in this example, can also be part of the main organization. In this scenario, the Medicine department has the responsibility of the Medicine clinical service, which in turn is in control of the Medicine A care unit, which in turn is in responsibility of the inpatient care team. This team's members are also members of the Shared medical appointments team.

Listing 2: Medicine A unit structure

```
{
 "LOCATABLE": {
  "UID": "87a58847-555f-43cd-8fb2-2f6aea37106b",
  "TEMPLATE_ID": "c7752924-fab5-44f8-a3f4-c7011d5b7109
      ↪ ",
  "RM_VERSION": "1.1.0",
  "ORIGINATING_SYSTEM_AUDIT": {...}},
 "IDENTITIES": [{
  "PARTY_IDENTITY": {
   "Ativa": "True",
   "Tipo": {
    "code": 4,
    "text": "Unidade",
    "terminology": "NIVEIS HIERARQUICOS"
   },
   "Nome": "Medicina A",
   "Identificador": 8008005
  },
  "PURPOSE": "Legal"
 }],
 "CONTACTS": [...],
 "RELATIONSHIPS": [{
  "SOURCE": {
   "ID": "87a58847-555f-43cd-8fb2-2f6aea37106b",
   "NAMESPACE": "Medicina A",
   "TYPE": "ORGANISATION"
  },
  "TARGET": {
   "ID": "0bbd910e-7132-48e7-a56c-046f6c0f36d0",
   "NAMESPACE": "Serviço de Medicina",
   "TYPE": "ORGANISATION"
  },
  "TIME_VALIDITY": {
   "start": "2000-01-01",
   "end": "2900-01-01"
  }
 }],
 "TYPE": "Organisation",
 "LANGUAGES": "Portuguese"
}
```

Listing 3: Medicine A Inpatient team structure

```
{
 "LOCATABLE": {
  "UID": "321b2278-13dc-4c52-a0c1-5005c20a2827",
  "TEMPLATE_ID": "721f2b0b-7b39-4db6-ad26-707a80f16f6d
      ↪ ",
  "RM_VERSION": "1.1.0",
  "ORIGINATING_SYSTEM_AUDIT": {...},
 "IDENTITIES": [{
  "PARTY_IDENTITY": {
   "Identificador": "31403",
   "Nome": "INT MEDICINA A",
   "Representação em uso": "Internamento Medicina A"
  },
  "PURPOSE": "Legal"
 }],
 "DETAILS": {...},
 "CONTACTS": [...],
 "RELATIONSHIPS": [{
  "SOURCE": {
   "ID": "321b2278-13dc-4c52-a0c1-5005c20a2827",
   "NAMESPACE": "INT MEDICINA A /HSA",
   "TYPE": "GROUP"
  },
  "TARGET": {
   "ID": "87a58847-555f-43cd-8fb2-2f6aea37106b",
   "NAMESPACE": "Medicina A",
   "TYPE": "ORGANISATION"
  },
  "TIME_VALIDITY": {
   "start": "2000-01-01",
   "end": "2900-01-01"
  }}],
 "TYPE": "GROUP",
 "LANGUAGES": "Portuguese",
 "ROLES": [{
  "id": "dc0bcdba-7e70-4f23-bae6-2dffb1cc4885",
  "namespace": "Nome Fictício",
  "type": "ROLE"
 },{...}]
}
```

The JSON structures listed above are the result of DIM materialization based on openEHR standard specifications for an *ORGANISATION* (Medicine A) and for a *GROUP* (Medicine A Inpatient).

The *TARGET* attribute holds the dependency with another structure, and each *GROUP* (team) contains the *ROLES* attribute, which contains all the *PARTY_REF's* of all the roles associated with that team. Each of these responsibilities is then associated with a *PARTY_REF* for the individual who will play that part.

4.2.3.2 Clinical Information System developed

The EHR top package of all RM is represented briefly in figure 36. Each patient entering a healthcare facility that has the openEHR standard implemented has his or her EHR associated with it. This structure contains all the *PARTY_REFS* of its status, its access permissions, its interactions with the institution, along with the log register always updated. This log register is called *CONTRIBUTIONS*, in which each contribution is constituted by the set of direct and indirect changes to the logs of a given EHR on a given day.

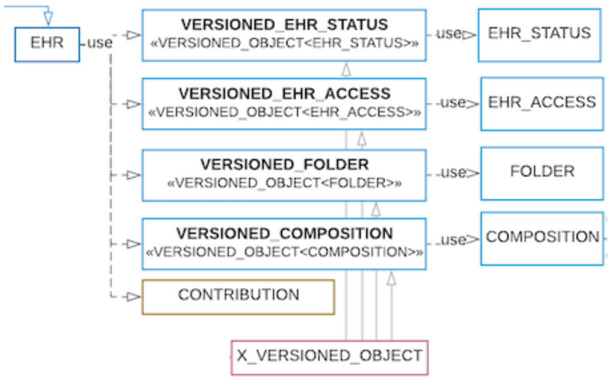

Figure 36: OpenEHR RM specification of EHR IM.

The use case in figure 37 is intended to represent some interactions of a patient with a particular healthcare facility, whose IS's are based on the openEHR standard.

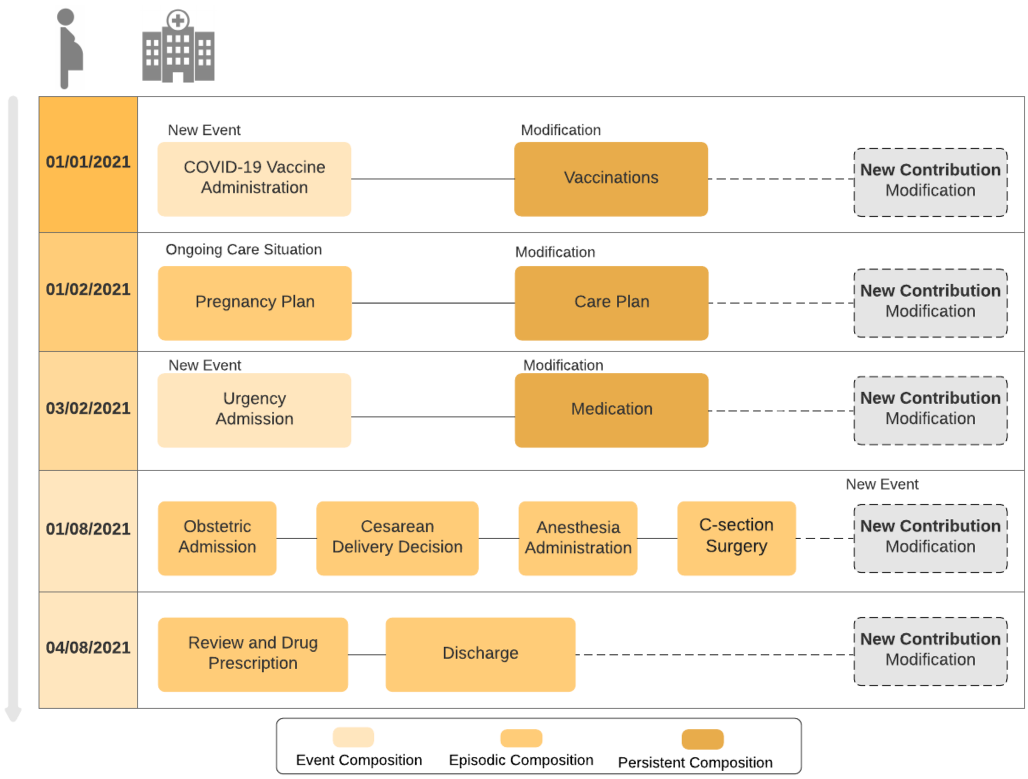

Figure 37: OpenEHR RM specification of EHR IM.

Firstly, the use case represents the COVID-19 vaccine administration to a female patient. By itself, this clinical act is a new event that is recorded in an openEHR system as an ***Event Composition***. An event composition records clinical occurrences during certain acts performed by the health care organization with and/or for the patient, such as observations, or actions during patient contact or even clinical acts during surgery, or even pathology tests where the patient is not present. These types of compositions are important to record also the context information of the event, i.e. who, when, where and why the event happened.

With the occurrence of new events, if the submitted event composition is associated with a given ***Persistent Composition***, this in turn is also updated with the new information. In this use case, the patient's EHR contains a *Persistent Composition* of Vaccinations, which was updated when the COVID-19 vaccine was taken.

Following the timeline of the use case presented, on the first day of February 2021, the patient makes a pregnancy planning appointment, the plan for which will last 5 to 6 months until delivery takes place. In this sense, we are facing an episodic context with a finite duration, whose information is stored in ***Episodic Compositions***. When planning the pregnancy, the system received a new episodic composition and automatically there was a modification of the persistent composition called "Care Plan".

The use case patient was admitted to the hospital's emergency department and prescribed medication in the days that followed. Her Medication persistent composition was modified, and a new element was introduced. As a result, a new contribution was generated for that EHR_ID on that particular day. Also, on the first of August 2021, the same patient was admitted to obstetrics, but this time it was a clinical act scheduled by the pregnancy plan specified. That is why it is referred to as episodic or episodic compositions.

MAS versioning and organization developed

To increase the flexibility of the EHR structure, OpenEHR is provided the option of adopting a set of folders and their respective root directories. This feature allows for more flexibility in clinical record structure as well as the linking of profiles with the physician who is consulting or modifying a specific EHR at the time. A system of dynamic **Folders** was implemented in this project, which means that the previously described governance platform will allow those in charge of each department, service, or team to configure not only how they want to see the forms (UI generated by JDT structure) presented, but also how they want to see the information of a specific patient in a specific context.

It was required to create a consistent versioning and organizing system in order to create a long-lasting EHR that is easy to access and manage. The versioning system created was based on some openEHR principles, and it was organized using a **Folders** and **Directories** system. The number of folders produced, as well as their content, can be volatile or persistent over time, giving the openEHR a high degree of flexibility. Persistent folders are built up of persistent compositions that contain information that is transversal throughout the patient's life and can readily summarize his health state. Persistent folders, can be divided into categories such as problem lists, vaccines, and allergies, among other clinical topics. Otherwise, folders based on episodic or event data, with their corresponding types of compositions as content, can be also generated.

To enable the creation of new folders, a new template was designed in Better's Archetype Designer tool and is presented in figure 38. This template records relevant data such as the templates linked with the new folder, the date it was created, and the person responsible of its development. Furthermore, the *SLOT* of type *CLUSTER* was added to allow subfolders to be associated with the newly constructed folder.

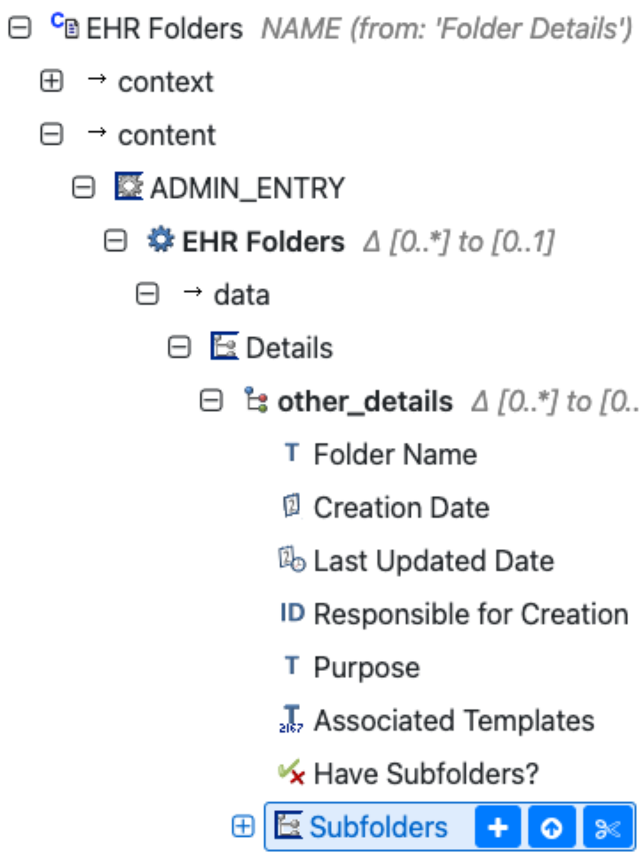

Figure 38: Modeled template for the creation of folders.

An area called *EHR Folder Builder* was created to allow users of the built governance platform, to control this level of structuring and hierarchization of EHR records. This section enables platform users to control the folders created through *delete*, *update* and *creation* operations. These last two processes will be performed at the template level, as described above, via the GUI layer. On the one hand, the *update* events take place through the retrieval of an already generated composition and the subsequent development of a new version of it. The *creation* operation, on the other hand, is only executed through the instantiation of a new template and its subsequent submission as a versioned composition.

Figure 39 shows the workflow of the MAS developed to ensure the organization and versioning of the CIM system.

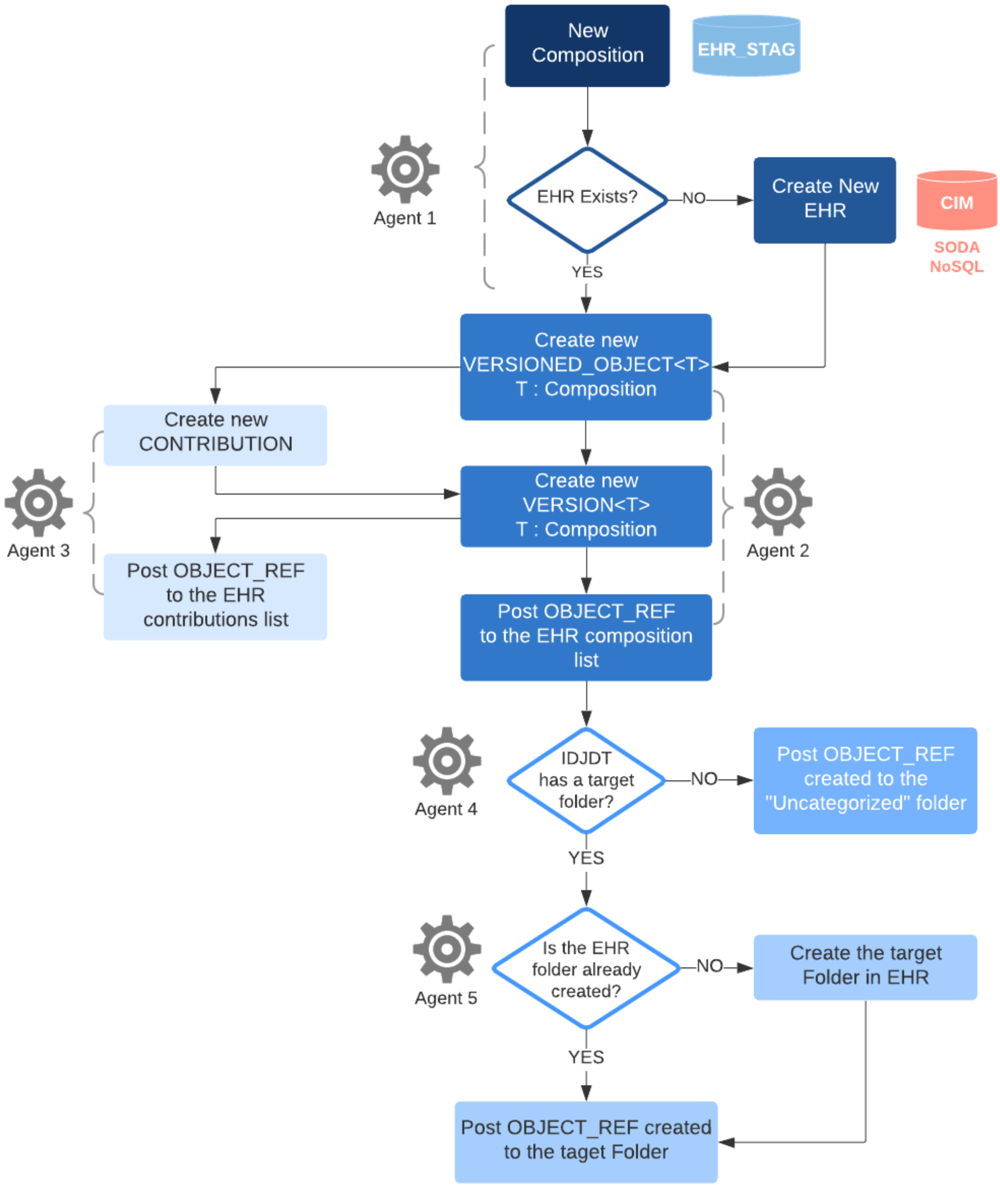

Figure 39: CIM versioning and organization system.

New records and records that are not yet ”completed”are kept in a staging area. Agent 1 is constantly looking for new complete records in the staging database for the system to proceed. This agent takes care of querying the staging database to discover new ”compositions”to integrate into a given EHR. Once it finds a new record, it will check if the *UID* of the patient associated to the record has a specific EHR associated to it. If it does, it sends a message to agent 2 telling it that it can process the new record for that EHR. If agent 1 doesn't find an EHR associated with the patient, it will create one and then send the message to agent 2.

The occurrence of new data leads to the creation of a new *VERSIONED_COMPOSITION* for the versioning of the information to be possible. This task is in charge of agent 2, which only initializes it after the message received by agent 1. After generating the *UID* of the new *VERSIONED_COMPOSITION*, agent 2 sends a message to agent 3 with the respective *OBJECT_REF* of the new versioning so that it can perform its tasks. Thus, agent 3 can create a new *CONTRIBUTION* and save its reference in the list of contributions of the respective EHR. Agent 2 finishes its task by saving the versioning reference of the new record in the list of compositions of the same EHR.

The termination of the tasks of the versioning agents happens when agent 3 sends a message to agent 4 with its respective *OBJECT_REF* created and the record's source *IDJDT*. With that information, agent 4 can check if there is any target folder for that template and if not, the received reference will be stored in a default folder. On the other hand, if there is a folder associated with the *IDJDT*, agent 4 sends a message to agent 5 so that it can save the reference to the new information in the target folder, within the EHR.

The following BPMN diagram presented in figure 40 was designed to represent the operations performed by the built MAS.

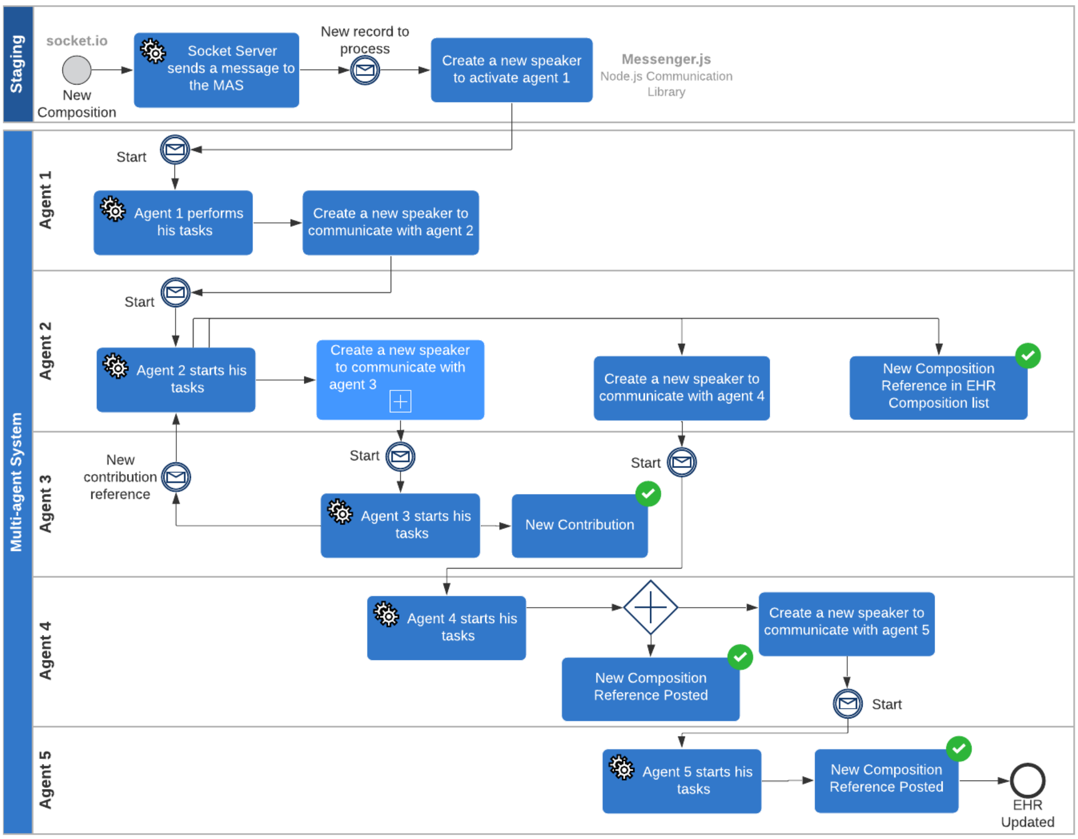

Figure 40: MAS BPMN diagram.

The staging area contains a **socket.io** server, which generates an event whenever a new composition is submitted. To capture this event, MAS connects to the socket server as its client and subscribes the *trigger MAS event* to activate its first agent.

A Node.js communication library named **Messenger.js** was used to create message exchange amongst the agents. A **listener** was created for each agent using this package, so that each of them is available on a specific port and ready to be triggered upon receipt of each of the messages sent by a certain **speaker**.

4.3 Conclusion

This chapter was written throughout the second stage of this research project. During the "Prototype Design"phase and after the research methodology was chosen, a main chapter was written describing all the work developed, including the adopted technologies at different levels - Databases, Backend and Frontend technologies - in order to answer to RQ5. A main system architecture was proposed and, after this, sub-processes were described too, responding to RQ6.

On the one hand, one of the major difficulties in the third stage, Artefact Development, was the high complexity of the openEHR specification and how to unfold it to achieve a module-based system architecture. Another significant problem was ensuring data syncronization between the LS and the novel system. These barriers were overcome during the development of the artefact in order to answer the RQ7.

In the fourth stage, in Artefact Evaluation, one of the biggest challenges was the development of a test and production deployment environment. Infrastructures were then created at various levels, including data schemas and machine configurations to place the backends and frontend developed. When deployed in a testing environment, several testing plans were developed in order to perform a survey of initial inconsistencies of the developed prototype, at different levels. These difficulties experienced when implementing and testing the artefact were taken as an answer to question RQ8.

Chapter 5

Case Studies

5.1 Data Migration from Legacy Systems

Converting and mapping data from a healthcare institution's current systems is one of the most critical components of building an HIS based on the openEHR standard. These LS, on the one hand, could be in use and constantly updated, while others may no longer be in clinical usage but nevertheless include useful clinical information gleaned from years of data collecting. Online and offline LS are distinguished in this project due to the differences in the ways in which they are transformed and used in the newly built OpenEHR system.

A mapping system for offline LS conversion was created and afterwards included in the new openEHR structure. Furthermore, the mapping system that was developed consists of the development of intelligent agents that retrieve relevant information from the already existing offline LS database and then bind that information to the new openEHR structure. This new structure is always saved in the form of a *"Composition"* that corresponds to a determined template that has been specifically designed for the purpose of storing it. At the same time as filling out the composition, the RM of the template that will be used in each context is filled out and saved as well. Because the data from the offline systems is already immutable, the conversion to the new structure may be done in bulk as a large data batch process, while the data's meaning and value are preserved.

5.1.1 Use Case 1: Clinical data migration to openEHR in an emergency context

The presented use case proposes making use of data from an offline emergency system that has immutable data for more than 20 years. Its main purpose is to convert emergency data to openEHR format and identify the significant difficulties faced throughout the conversion process.

In addition to information on a patient's admission, triage, and discharge, the emergency system data includes the patient's complaints and problem history. Furthermore, in order to identify diagnoses detected in an emergency context, it is generally necessary to collect information through the measurement of vital signs, analysis of procedures, tests, or analyses required by health professionals. The system under study also consulted the patient's problem list.

The data under analysis came from the CHUP institution's emergency department, where more than 1,300,000 emergency room admissions were collected for analysis between 2004 and 2018, making an average of 265 admissions per day. Also, some indicators of the data under study can be seen in figure 41.

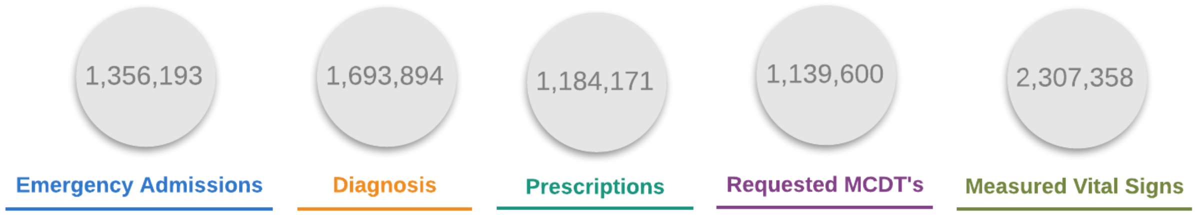

Figure 41: Numbers of studied records of the main clinical acts in emergency department.

This case study was essential to determine the most viable approach of establishing interoperability between existing systems and the newly developed openEHR system. Because of the security of working with already immutable records, this process could be cyclical and evolutionary.

5.1.1.1 Methods

To make the retrieval process as realistic as possible, the clinical experts collaborated with the IT experts to analyze the relevant information for the case study. This symbiotic relationship is essential for information retrieval as well as the continuous evolution of openEHR systems. This analysis resulted in the planning of openEHR structure modulation, as indicated in table 5 below.

It should be mentioned that the most difficult aspect of this initial phase was the existence of unstructured data with no standardization. These information can be found, for example, in the patient's disease history or in the patient's complaints during the emergency episode. The clinicians in these situations recorded their observations without any consistency in terms of data structuring. These same observations were inconsistent, as the same term was frequently recorded in different ways in different records. In these circumstances, modulation was restricted to *DV_TEXT* elements.

Table 5: Templates modeled in OpenEHR and the archetypes reused.

Template	Archetypes
urg_admission.opt	openEHR-EHR-COMPOSITION.encounter openEHR-EHR-SECTION.adhoc openEHR-EHR-OBSERVATION.story openEHR-EHR-CLUSTER.symptom_sign openEHR-EHR-ADMIN_ENTRY.admission openEHR-EHR-EVALUATION.reason_for_encounter openEHR-EHR-ADMIN_ENTRY.triage openEHR-EHR-CLUSTER.demographic
urg_discharge.opt	openEHR-EHR-COMPOSITION.administrative openEHR-EHR-ADMIN_ENTRY.episode_institution
diagnosis.opt	openEHR-EHR-COMPOSITION.encounter openEHR-EHR-EVALUATION.problem_diagnosis
procedures.opt	openEHR-EHR-COMPOSITION.encounter openEHR-EHR-ACTION.service
medication.opt	openEHR-EHR-COMPOSITION.encounter.v1 openEHR-EHR-ACTION.medication.v1 openEHR-EHR-CLUSTER.dosage.v1
disease_history.opt	openEHR-EHR-COMPOSITION.progress_note openEHR-EHR-OBSERVATION.progress_note
vital_signs.opt	openEHR-EHR-COMPOSITION.encounter openEHR-EHR-OBSERVATION.laboratory_test_result openEHR-EHR-CLUSTER.laboratory_test_analyte openEHR-EHR-CLUSTER.demografico_sonho openEHR-EHR-OBSERVATION.blood_pressure openEHR-EHR-OBSERVATION.pulse openEHR-EHR-OBSERVATION.glasgow_coma_scale openEHR-EHR-OBSERVATION.respiration openEHR-EHR-OBSERVATION.pulse_oximetry openEHR-EHR-OBSERVATION.body_temperature
requests.opt	openEHR-EHR-COMPOSITION.request openEHR-EHR-SECTION.adhoc openEHR-EHR-INSTRUCTION.service_request

5.1.1.2 Results

There were multiple data sources that supplied the offline system under study, requiring connections to Oracle databases of various versions and, as a result, the installation of different *oracle clients*.

To handle the different versions of the existing databases and the multiple JDTs generated by modulation (after the submission of the developed templates to the AIDAEHR platform), a control table was created to ensure that all of the essential requirements were connected. A *retrievalData* algorithm had to be created in order for the data conversion to actually occur. The main function of this method is to process each row of the control table before proceeding with the data conversion. Each control line includes the SQL query created by IT experts, its respective JDT id and the binding structure of its items, as well as the source database to be queried. The algorithm runs in the background for a set amount of time, depending on the number of records to be processed, until the last record is retrieved. The records are then transformed into their target composition, coupled with their respective RM, and stored in the staging area. The aforementioned MAS is triggered after each insertion of a new record into this staging repository. All the processes described are shown in the schema presented in figure 42 below.

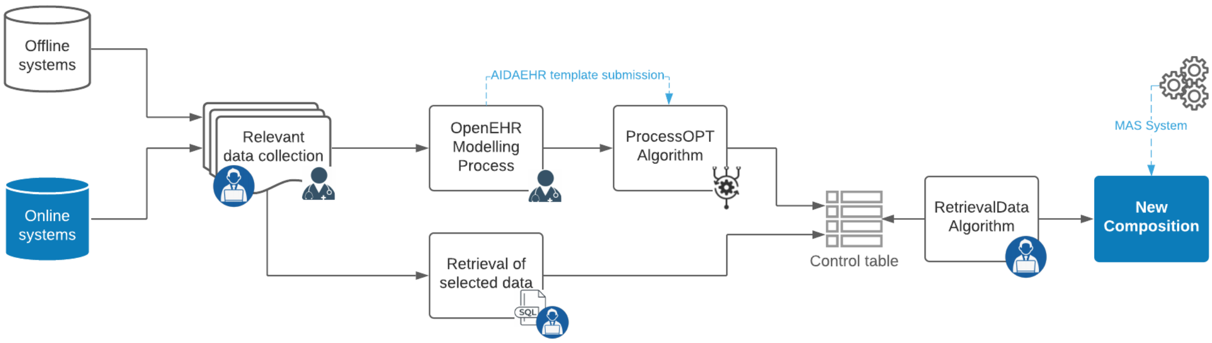

Figure 42: Data retrieval process to the new openEHR system.

5.1.1.3 SWOT Analysis

An internal and external SWOT analysis was performed to assess the developed openEHR migration system's internal strengths and weaknesses, as well as the external opportunities and threats.

Table 6: SWOT analysis of developed data migration system.

Internal Environment	**External Environment**
Strenghts	Opportunities
Retrieval of important clinical data	Opportunity to create a structured and open clinical repository
Modelling of new openEHR structures in emergency context	sharing openEHR structures developed for an emergency department
Interoperation between LS and the new system	Discovery of new knowledge from a big amount of transformed data
Weaknesses	Threats
Connections to old version databaseses	More efficient externally migration systems
Existence of non-standard free text after migration	Continued existence of free-text fields in the clinical act record

5.1.1.4 Conclusions

The created backend system is flexible to whatever data source and openEHR target structures are used. The existence of a high number of free-text records made the conversion from an outdated offline system challenging. The main conclusion extracted from this case study was the future necessity for the development of a NLP module using Named Entity Recognition (NER) methods, which would result in improved clinical modulation and knowledge discovery. This NLP module would overcome both the existence of non-standard text, and the improvement of collection of reliable and shareable clinical information in a future central open repository.

5.2 Fast Developed System Implementations

Throughout the four years of development of the system presented in this doctoral thesis, several tests were performed and use cases developed. The initial use cases were in applications that were already developed and modules were added to interact with the openEHR system. When a minimally stable artefact was achieved, some of the system modules were incorporated in several works in progress, such as the module DIM to be used in clinical work plans and in the authentication systems of some platforms. In terms of decision making, the feeding of the DLM (still under development) was done through structures from the items generated by the OPT conversion algorithm and the development of the "Rules and Conditions"area in the *AIDAEHR* governance platform.

One of the case studies that had the most impact due to the easy adaptation and use of the developed openEHR system was the implementation of the COVID-19 patient workflow in the *CHUP* healthcare institution, described below. The second case study presented is mainly focused with the knowledge discovery layer of the openEHR data from case study 1.

5.2.1 Use Case 2: The use of OpenEHR modeling in the fight against the COVID-19 pandemic

The COVID-19 pandemic provoked the collapse of healthcare systems in several countries, forcing a restructuring of institutions at essentially all levels. This case study was created at the CHUP, where the first COVID-19 patient was diagnosed in early March 2020.

Faced with an unknown crisis, the institution chose to build a supplement to the already implemented HIS, based on the openEHR specification, to support the workflow associated with suspected and confirmed cases, in order to respond quickly to the local pandemic growth.

In order to create a feasible solution, a multidisciplinary team of researchers and healthcare professionals used a front-line mining technique preceded by openEHR modelling.

This case study tries to provide answers to multiple research questions that arose during the implementationof new approaches in the HIS of healthcare institutions. These RQs are as follows:

- RQ1: Why is interoperability important in healthcare institutions?
- RQ2: What is the significance of adopting global clinical standards?
- RQ3: What is the ideal strategy for designing and developing a new clinical workflow in a pandemic situation?
- RQ4: How will the newly created workflow be integrated into the institutions' information systems?

5.2.1.1 Methods

It was established and adopted a methodology for developing well-founded templates that can support the healthcare process and save all relevant patient information in order to create well-founded templates that can support the healthcare process and store all relevant patient information figure 43. As the medical community became more informed of COVID-19, a requirements elicitation process was done to identify which concepts should be saved and how they should be represented. It was also necessary to analyze the administrative and clinical concerns of the institution in order to collect domain concepts for modelling. In terms of modelling, searches on the openEHR CKM returned valuable structures through the additional COVID-19 specific archetypes and templates, generated and shared by the openEHR community. The Archetype Designer by *Better* [25] company was also used in order to modulate valuable structures applied to the CHUP institution's necessities.

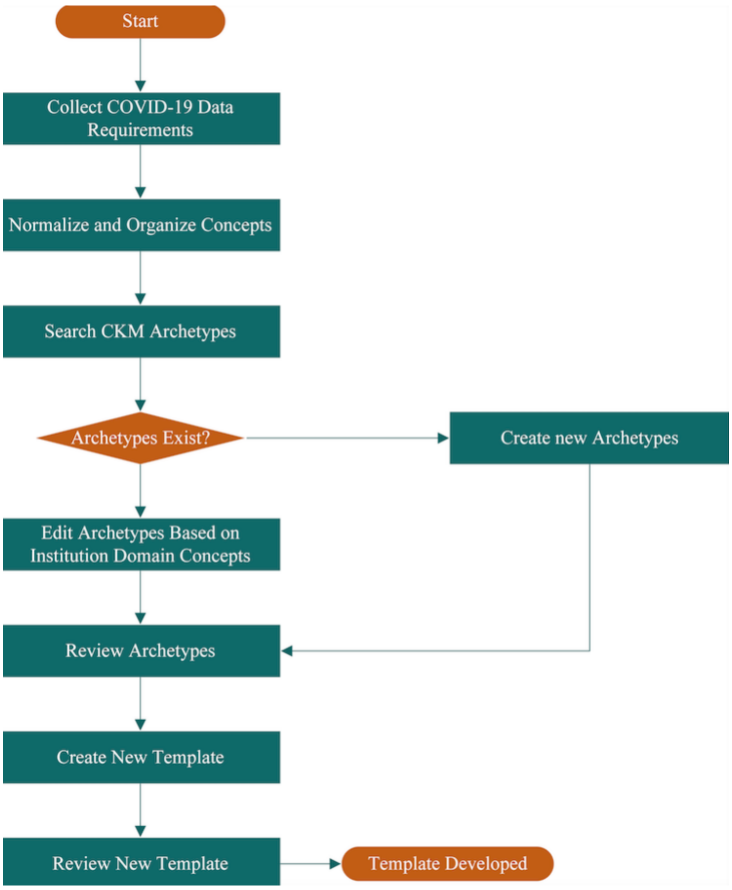

Figure 43: Modelling approach for openEHR structure development.

Throughout the process, communication between CHUP healthcare experts and patients was continuous and direct, via text messages and phone calls. Thus, during the pandemic, the workflow of health professionals to treat COVID-19 patients was determined and depicted in figure 44.

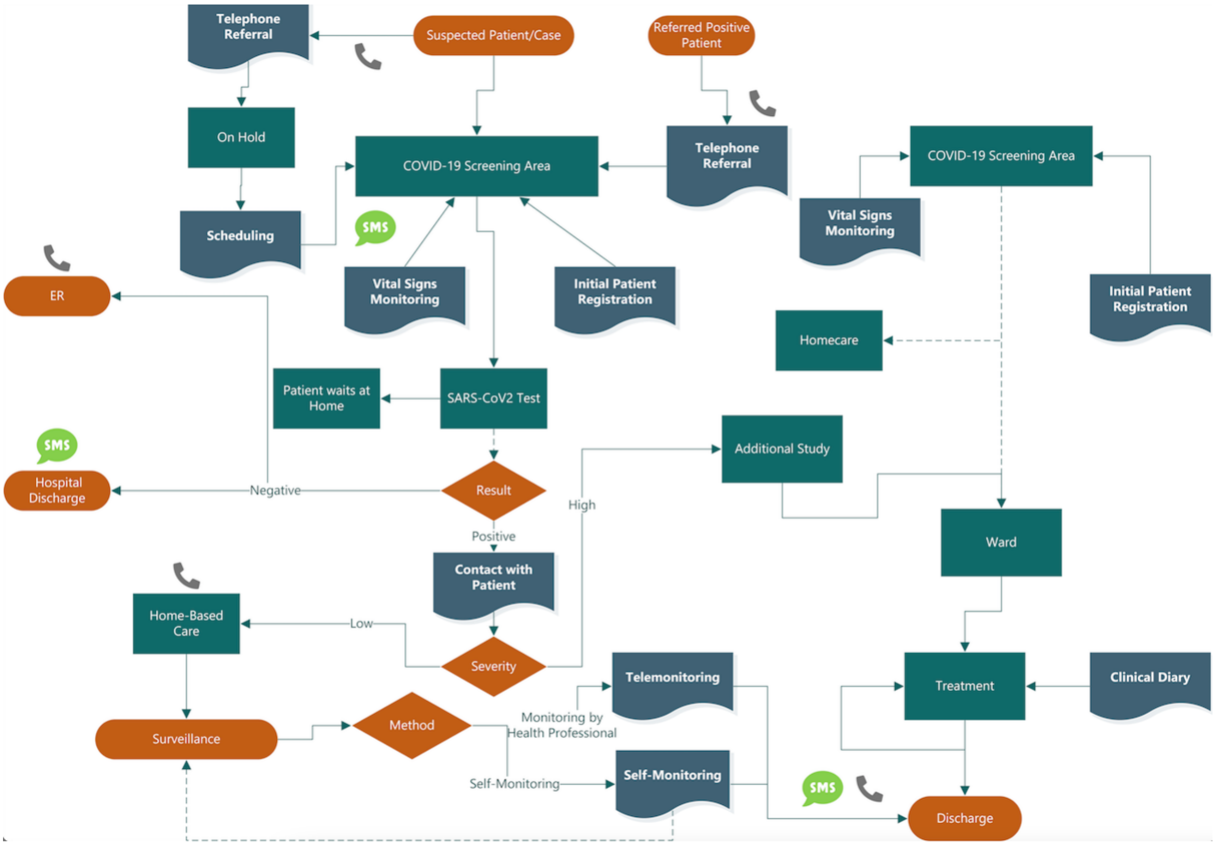

Figure 44: CHUP Patient Treatment Flow Covid-19.

A patient suspected of carrying the SARS-CoV2 virus will be referred by telephone or seen directly in the COVID-19 testing area, depending on his/her health status. A SARS-CoV2 test is scheduled if their health has not worsened. Both the "Telephone Referral"and "Scheduling"events make use of openEHR forms designed for this purpose. A patient who has already tested positive for the SARS-CoV2 virus is referred for follow-up by telephone.

Except for the most severe cases of infection, in which the patient has to be hospitalized, patients wait at home for their COVID-19 test results. If the result is positive, and depending on the severity of the patient's symptomatology, the patient may be followed at home, in home care, or an additional study may be necessary to choose more intense treatments. In the case of home internment with less severe symptoms, the patient monitors himself, registering every day his health status, his body temperature, the medication he has taken, among other relevant parameters. This self-monitoring by the patient is recorded using a specifically developed application that uses the openEHR "Self-Monitoring"form. Depending on each patient's records, there may be a need to call the patient to take stock of their health status ("Telemonitoring"form).

5.2.1.2 Results

The implementation of a new system based on the openEHR specification has allowed health professionals to optimize and automate their medical procedures, particularly during the COVID-19 epidemic. Also, the use of many archetypes stored in CKM demonstrates the relevance of sharing clinical models.

A platform for patient monitoring and referral was also developed, which included web applications and a microservice-based architecture. Thousands of compositions have been submitted and processed since the proposed system and its COVID-19 methodology were implemented. Several measures of the platform's usage were gathered and evaluated using visual analytics software, namely Tableau Desktop, to show the project's Key Indicators of Performance (KPI). The given results are separated into two categories: modulation and clinical knowledge extraction.

On figure 45 a representation of which archetypes were used in each created template is presented.

Figure 45: Archetypes used in the templates modelling process.

From March to June 2020, the amount of openEHR records submitted in this process could be tracked, as shown in figure 46. There was a peak in the first half of April 2020, accompanied by the pandemic scenario in Portugal. As the crisis unfolded, by April 10, 2020, Portugal has registered more than 435 deaths and 1516 daily cases counted [92]. This indicator was also evaluated between February and September of 2020.

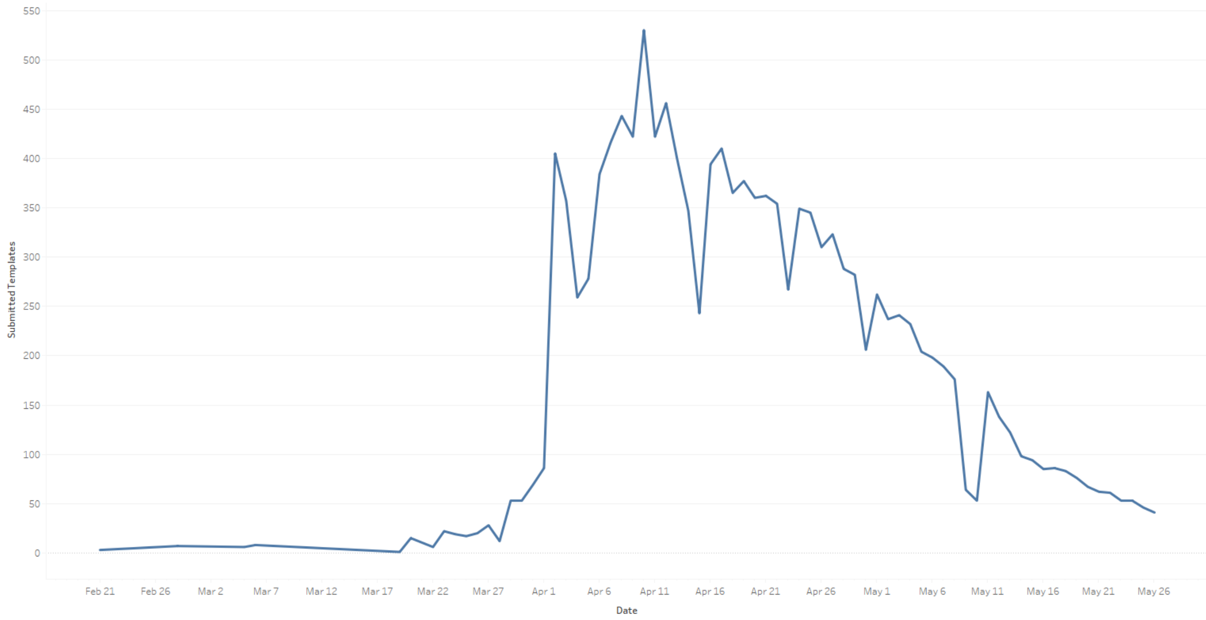

Figure 46: Records submitted according on openEHR structures developed.

Then, from February through May, figure 47 shows the number of submitted templates per template. The largest number of records submitted were of the "Self-Monitoring"type, which is understandable since each patient recorded his or her health condition more than once a day.

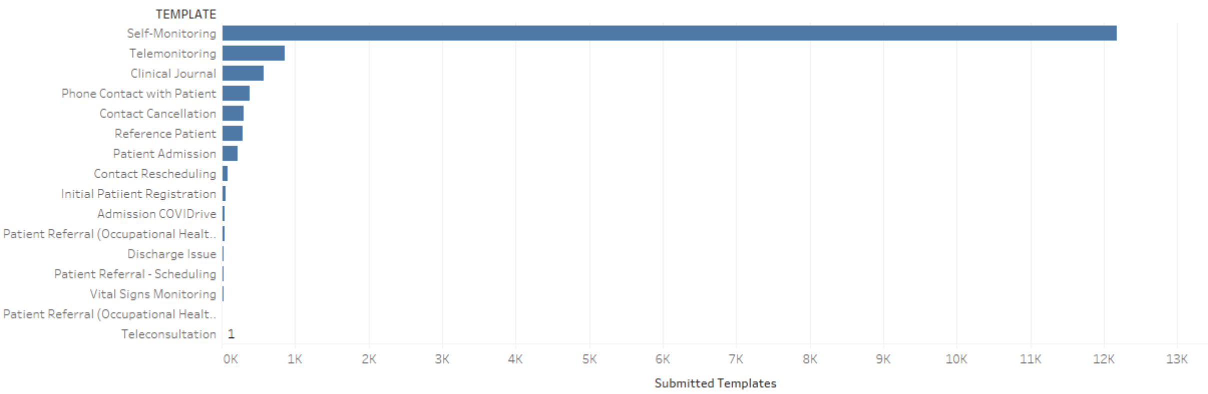

Figure 47: Number of records submitted per Template.

It was possible to evaluate how much time a patient had to wait for his/her COVID-19 test results by using the date range between the "Initial Patient Registration"and the "Telephone Referral"templates. The average COVID-19 test waiting time from the day the test was completed, is depicted in figure 48. The time window to evaluate this indicator was from February to September 2020.

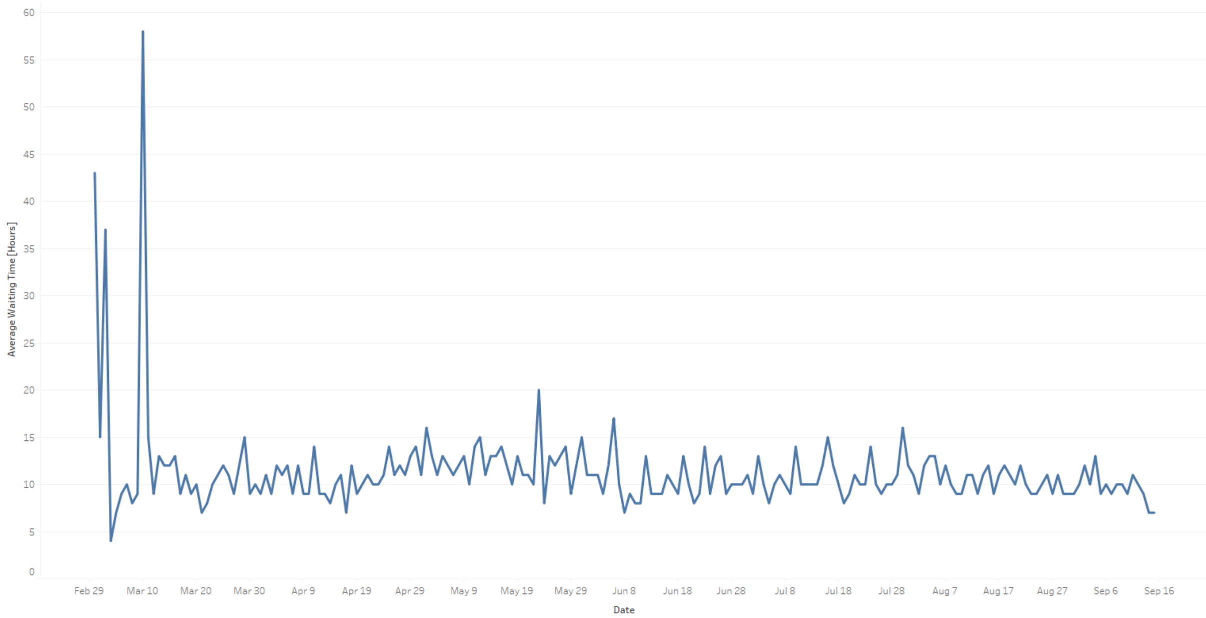

Figure 48: Average COVID-19 Test Waiting Time.

5.2.1.3 Discussion

With the dual architecture in place, it was possible to quickly obtain performance to reuse worldwide data models developed in the OpenEHR community and adapt to local needs with minimal effort. This approach enables the reuse of archetypes in different templates throughout the COVID-19 patient care workflow. Furthermore, these templates have been used in a variety of scenarios, indicating a comprehensive and extensive modulation. The most gratifying aspect of using openEHR is its high adaptability in the face of continually changing knowledge circumstances. When changing clinical knowledge, it was only necessary to change the template, preserving the consistency of the software used in the openEHR system. The modelling process must be accompanied by the essential clinical guidelines to maintain data integrity in a particular clinical environment, and it must always be designed as a continuous process so that its update is quick and effective.

From the results obtained in this study, it was possible to answer the research questions that were initially formulated.

Why is interoperability important in healthcare institutions?

It is crucial to be able to use technology to promote the sharing and exchange of data between hospitals, laboratories, and pharmacies. A high level of interoperability in healthcare institutions improves the patient experience, decreases waiting times, improves and simplifies clinical records, reduces expenses associated with old-fashioned techniques as well as potential errors, and protects patient privacy.

What is the significance of adopting global clinical standards?

The use of global standards ensures the symbiosis of reliable and efficient software with other devices and components. These standards establish healthcare interoperability, allowing communication with external applications. Using openEHR allows reuse clinical data model defined at an international level, with compromising local needs.

What is the ideal strategy for designing and developing a new clinical workflow in a pandemic situation?

The effective approach for designing workflows according to specific clinical guidelines is to adopt a dual architecture that separates RM from the clinical knowledge model. As a result, a development process is distributed by clinical and IT professionals who complement one another. Furthermore, it converts into a continuous improvement plan, with dynamic and ever-changing modulation structures. As an independent process, these changes are not reflected in software errors. In a pandemic scenario, professional guidelines and clinical knowledge are constantly changing, making it necessary to use this type of architecture.

How will the newly created workflow be integrated into the institution's IS?

The designed templates must be converted to dynamic GUIs so that they can be used in all target IT systems later on. Also, mechanisms for entering and querying data to feed the new GUI must be developed.

5.2.1.4 SWOT Analysis

An internal and external SWOT analysis was performed to assess the internal strengths and weaknesses of the proposed approach, as well as the external opportunities and threats.

Table 7: SWOT analysis of use case performed.

Internal Environment	**External Environment**
Strenghts	Opportunities
Structured clinical data	Competitive advantages to collaborate with international groups
Fast and simple adaptation in unexpected scenarios	Sharing the developed internal structures with the openEHR community
Use of a globally recognized standard	Creation a worldwide centralized openEHR repository
Reuse of structures for different use cases	Continuous active research in the scientific community
Weaknesses	Threats
Small intern team of openEHR modulators	More efficient externally developed systems
Constantly changing of intern clinical requirements	Low openEHR professionals number slowing its global expansion
	Constantly changing of openEHR specifications

5.2.1.5 Conclusions

This case study is part of a pilot project at the CHUP institution to implement openEHR. Following an analysis of the institution requirements, it was determined to implement the openEHR clinical approach, which has the support of a large worldwide community and is constantly evolving around the world. This study prototype was implemented in Portugal at the start of the COVID-19 pandemic, supporting the institution and its health professionals in organizing themselves in times of crisis and high pressure. The methodology also allowed the institution to define its clinical guidelines in a structured way, as well as deal with the constant changes of the institution. Based on the findings and user feedback, the system built on the openEHR standard has proven to be simple to use and highly valued. The system translates into the dynamic adaptability associated with the present epidemic, as well as the ability to readily interoperate with other HISs. It was feasible to quickly get the most diverse indicators in terms of their workflow as well as the evolution of the pandemic in the healthcare institution using standardized and structured data records. To round out this conclusion, I would want to thank my colleagues and healthcare professionals who helped make this case study a reality.

5.2.2 Use Case 3: Using OpenEHR artefacts to Predict COVID-19 Diagnosis

As mentioned earlier, the ongoing pilot project focusing on the OpenEHR specification was used by the CHUP institution to refine the COVID-19 patient treatment workflow. As a result, a hybrid solution was developed that interacted with existing LSs to ensure that users could communicate fast and effectively with minimal effort. This implementation generated 13,000 records based on patients who fulfilled the ”Self-Monitoring”template. In the face of such an unknown and unpredictable disease, these structured records with clinically important information were used to develop a system for tracking symptoms and health problems of suspected or confirmed SARS-CoV-2 infected patients. All data on patients' evolutionary status in home care, as well as the results of their COVID-19 test, were used to train different ML models with the goal of creating a predictive model capable of identifying COVID-19 infections based on the severity of symptoms reported by patients.

The major purpose for this use case is to explore the collected data in order to develop and train machine learning models capable of predicting COVID-19 infections, thereby assisting health professionals at healthcare facilities. In addition, the Cross Industry Standard Process for Data Mining (CRISP-DM) methodology, which is used to construct DM procedures, was implemented to validate that the generated models are valid and repeatable.

5.2.2.1 Methods

The extraction of useful patterns and knowledge from the inquired patient data was carried out in the hope of developing a predictive model capable of distinguishing between healthy people and people with COVID-19, thereby assisting healthcare professionals in the early detection of infected patients, allowing them to isolate as quickly as possible and, as a result, decreasing the virus's spread.

CRISP-DM is a common framework for designing DM processes all over the world. The European Community funded this methodology, which enables for project replication as well as project planning and management. **CRISP-DM** is a cyclical process with six steps, as illustrated in figure 49: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment.

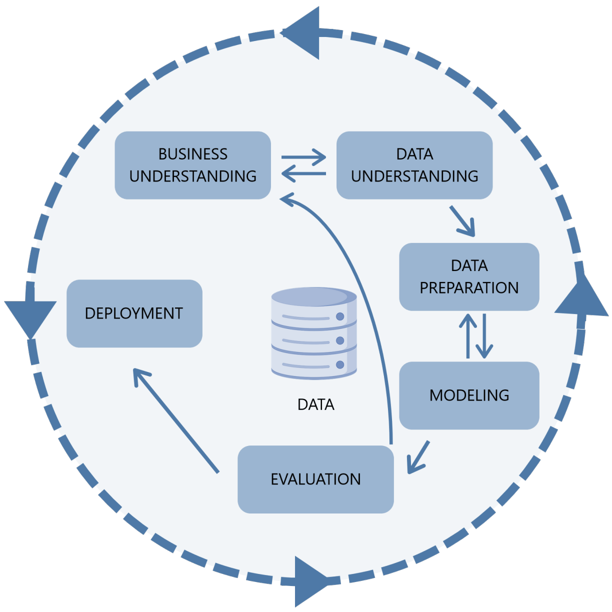

Figure 49: Cycle of **CRISP-DM** methodology.

With the exception of the final phase, model deployment, the following topics provide a full discussion of each of CRISP-DM phases.

Business Understanding

COVID-19 is associated with a wide range of symptoms, including fever, headache, respiratory tract symptoms such as cough and dyspnea, and loss of taste or smell, among others. Because this is a new disease that manifests different ways in different people, diagnosing it is difficult because most of the symptoms presented are either difficult to identify because they are usually less severe or because they are common day-to-day conditions for some people, such as people who frequently suffer from headaches, or because they are symptoms of common diseases, such as the flu. As a result, the COVID-19 diagnosis is based on a series of laboratory tests. However, as the pandemic spreads and grows at an exponential rate, the number of tests available becomes limited.

As a result, a model capable of detecting the presence of COVID-19 based on survey responses about a patient's clinical symptomatology should significantly improve the health-care system's ability to respond rapidly and efficiently to newly identified cases. Thus, the business goal of the work presented in this research is to determine which characteristics have a higher impact on COVID-19 diagnosis, as well as to build an efficient prediction model for early disease detection based on clinical symptomatology data from patients who have been inquired. For this, different machine learning algorithms were used to create models capable of extracting useful information from the data. As a result, the end purpose is always to provide better healthcare for patients and to deliver them the appropriate therapy as soon as possible.

Data Understanding

This stage was crucial to interpreting the data and determining the relationships between the attributes. As aforementioned, the dataset used in this study contains a wide range of information concerning COVID-19 patients and the symptoms they encountered. The extracted data included 13,434 instances and 14 attributes. Also, each of these instances refers to a patient who has been inquired and has his or her own medical data. The description of the attributes can be found in the next table 8 .

Table 8: Description of the dataset's attributes.

Attribute	**Description**	**Type**
Patient_id	Patient's Identifier	Integer
Age	Patient's age	Integer
Gender	Patient's gender [1]	Integer
Temperature	Patient's body temperature [2]	Integer
Headache	Patient's headache evaluation [2]	Polynomial
Muscle_pain	Patient's muscle pain evaluation [2]	Polynomial
Cough	Patient's cough evaluation [2]	Polynomial
Diarrhea	Patient's diarrhea evaluation [2]	Polynomial
Thoracalgia	Patient's thoracalgia evaluation [2]	Polynomial
Shortness_of_breath	Patient's shortness of breath evaluation [2]	Polynomial
Shortness_of_smell_taste	Patient's shortness of smell and taste evaluation [2]	Polynomial
Medication_last_24h	Medications taken in the previous 24 hours [2]	Polynomial
Global_evaluation	Patient's health status [2]	Polynomial
Result	COVID-19 test result [3]	Binomial

[1] {Female, Male} [2] {No,I have now, Keeps, Improved, Worsened} [3] {Negative, Positive}

The figures below illustrate briefly the type of information collected from the dataset as well as some of its trends. The graphs depict the distribution of patient data based on age, gender, and temperature.

Figure 50 illustrates the age distribution from newborns to the elderly, with the age range of 25 to 60 years holding the most records. Because the COVID-19 screening questionnaires were available online, it is probable that the larger concentration of information in this age group reflects a greater familiarity with current technologies.

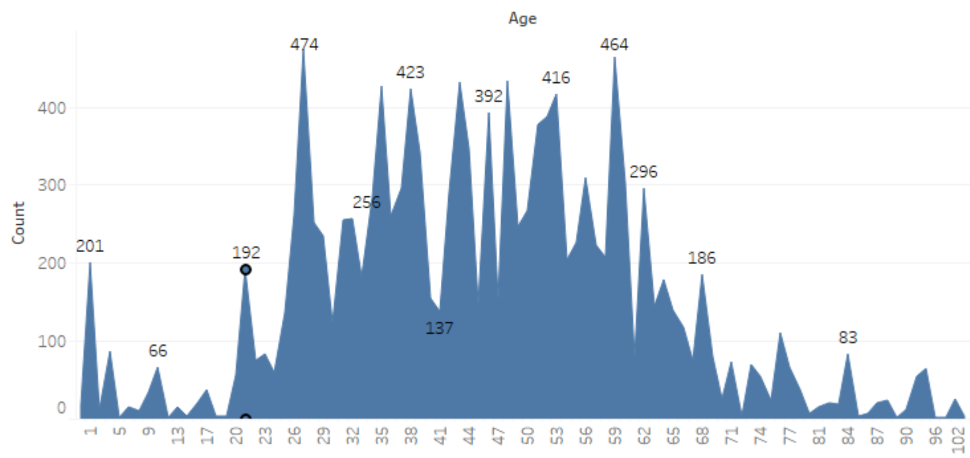

Figure 50: Patient distribution by age.

In terms of patient gender, figure 51 reveals that the female gender was more abundant in the research dataset than the male gender.

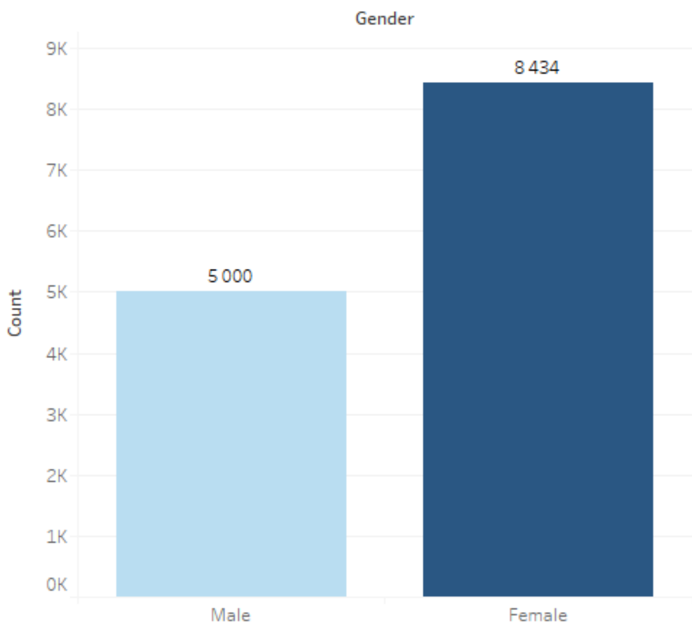

Figure 51: Patient distribution by gender.

The patient's body temperature, which was read by the patient himself/herself, is one of the most important established metrics. The temperature distribution of the patients, which ranges from 35 to 42 degrees Celsius, is shown in figure 52. The analysis of this graphic reveals that the majority of patients had a temperature of 36 degrees Celsius, and a nearly symmetrical distribution of temperatures to either side may be shown by plotting a perpendicular axis at that point. Surprisingly, the majority of patients have normal body temperatures, ranging between 35 and 37 degrees Celsius. This is remarkable given that the majority of people included in this study are COVID-19 positive cases, as seen below.

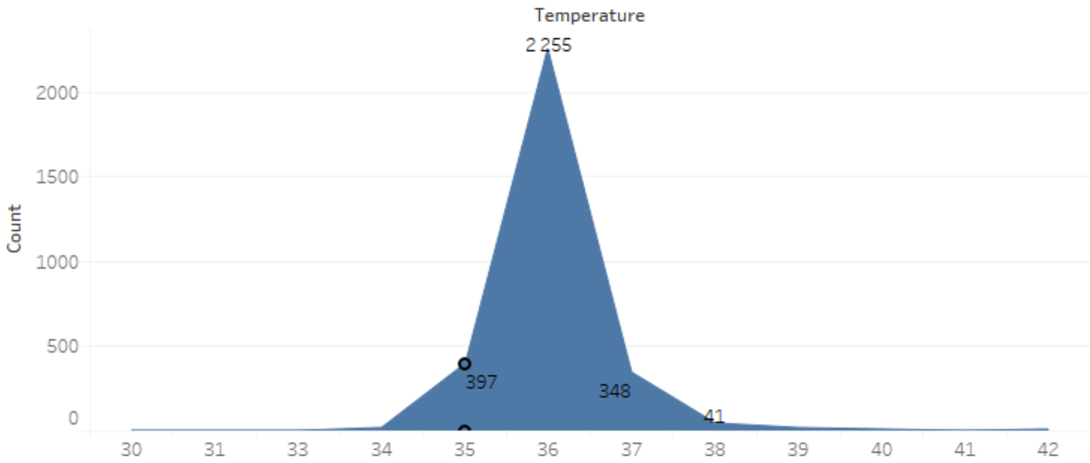

Figure 52: Patient distribution by temperature.

Finally, figure 53 illustrates the distribution of COVID-19 results among the patients in this case study (the target attribute). The target class distribution is severely imbalanced, with just 7.24% of occurrences generating a negative result and the remaining 92.76% providing COVID-19 positive results, as shown in this figure.

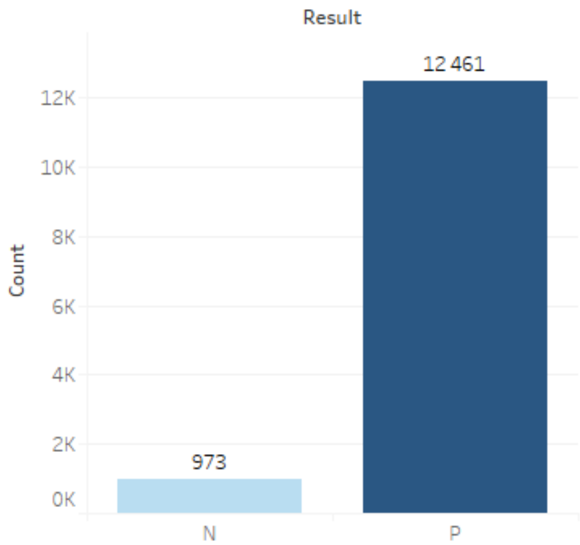

Figure 53: Patient distribution by COVID-19 result test.

These statistics show that the majority of users who answered the screening questionnaire provided in the built web application only wanted to be remotely monitored if they tested positive for COVID-19, resulting in an unbalanced dataset.

Data Preparation

Data preparation is the most time-consuming part of the CRISP-DM process, involving the integration, cleaning, transformation, and sampling of data. The data was initially integrated, and the data cleaning process was used to identify duplicate data, missing values, outliers, and inconsistencies.

There were no duplicates or outliers detected throughout the data cleaning procedure. However, some inconsistencies were discovered and needed to be rectified. The *Temperature* attribute has the most inconsistencies. Because this is a numerical attribute, there may be some variation, such as some patients filling in the values with commas and others rounding them, as well as some patients putting the unit of measurement and others don't, and in the case of putting the unit, the formatting may differ, such as "°C", "degrees", and "degrees Celsius". As a result, all units were removed, and all temperature values were rounded to convert into Integer type. In addition to this attribute, the attribute *Medication last 24h* required a specific transformation process since, because it is a free text field, patients used multiple designations to designate the same drug. As a result, a time-consuming and labor-intensive transformation process was implemented to verify that the drug names were consistent.

In addition to data inconsistencies, missing values were detected and treated by replacing them with the mean for numeric attributes and the mode for nominal attributes. As a result, the missing *Temperature* values were replaced by the attributes average value. In turn, the missing values of nominal attributes corresponding to the patient's symptoms were substituted by the most common value, "No".

In this phase, the under and over types of sampling methods were evaluated for the definition of different data approaches in order to determine which form of sampling is better for COVID-19 case classification. During the Modeling stage, many scenarios were constructed by selecting specific attributes in order to evaluate their impact on the final prediction.

Modeling

This phase consisted in the preparation of different DMM using the *RapidMiner* software with the dataset resulting from the Data Preparation stage. Each DMM can be described as belonging to an Approach (A), being composed by a Scenario (S), a Missing Values Approach (MVA), a DMT, a Sampling Method (SM), a Data Approach (DA) and a Target (T), as expressed in equation 1.

$$DMM = \{A, S, MVA, DMT, SM, DA, T\} \tag{1}$$

There was only one target (T), which was the *result* variable. Since Classification was the chosen Approach (A), six different classifiers were selected to be used as DMT, namely Decision Tree (DT), Random Forest (RF), Random Tree (RT), Naive Bayes (NB), Naive Bayes - Kernel (NB-K) and Deep Learning (DL). The DL algorithm is an implementation of the Rapidminer operator which is based on a multi-layer feed-forward artificial neural network that is trained with stochastic gradient descent using back-propagation on a node of H2O cluster.

For each DMT, three Sampling Methods (SM) were tested:

- *Split Validation*, with 80% of the data used for training and the remaining amount for testing.
- *Split Validation*, with 70% of the data used for training and the remaining amount for testing.
- *Cross Validation*, using 10 folds and where all data is used for testing.

Because the target variable's class distribution was considerably unbalanced, two Data Approaches (DA) were investigated: undersampling and oversampling, with the SMOTE upsampling methodology being used.

Regarding the scenarios, the first scenario (S1) includes all attributes. In the second scenario (S2), it was decided to remove the *Thoracalgia* attribute. On the other hand, the third scenario (S3) includes all attributes except the *Shortness_of_smell_taste* attribute since the loss of smell does not always imply the loss of taste, and vice versa. Therefore, it was decided that it was important to investigate the influence of this attribute on the prediction process.

As a result, in this study, the DMM are defined as follows:

- A = {Classification}
- S = {S1, S2, S3}
- MVA = {N/A, Replace (Average and Replenishment)}
- DMT = {DT, RF, RT, NB, NB-K, DL}
- SM = {Split Validation (80%), Split Validation (70%), Cross Validation (10 folds)}
- DA = {Undersampling, Oversampling (SMOTE upsampling)}
- T = {result}

In total, 216 models were induced according to equation 2.

$$DMM = 1(A) \times 3(S) \times 2(MVA) \times 6(DMT) \times 3(SM) \times 2(DA) \times 1(T) \tag{2}$$

Evaluation

As a classification approach, each model produced a confusion matrix for its evaluation, which represents the number of False Positive (FP), False Negative (FN), True Positive (TP), and True Negative (TN). Several evaluation metrics can be calculated from these values, this study in particular used the accuracy (3), precision (4), sensitivity (5), and specificity (6) to support the evaluation and conclusion of this research case. The description and calculation of each of these metrics is presented below:

– **Accuracy:** This indicator calculates the ratio between the instances correctly classified by the forcast model and all classified instances for the correctly TP classified instances, that responds to the question:

How many patients were accurately classified out of the total?

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP} \tag{3}$$

– **Precision:** This parameter measures the proportion of positive occurrences properly classified by the model to the total number of positive instances, that responds to the question:

How many of patients who were classified with COVID-19 actually had the disease?

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

– **Sensitivity:** This metric is considered as an integrator indicator and measures the ratio of positive instances correctly classified by the model to the total positive instances, that responds to the question:

How many COVID-19 patients were successfully predicted out of all of them?

$$Sensitivity = \frac{TP}{TP + FN} \tag{5}$$

– **Specificity:** Reveals the correctly TN classified instances through the calculation of the proportion of negative occurrences correctly classified by the model to the total negative instances, that responds to the question:

How many healthy patients were correctly predicted?

$$Specificity = \frac{TN}{TN + FP} \tag{6}$$

5.2.2.2 Results

The outcomes analysis is separated into sections based on the many metrics to be examined, which were previously specified in the CRISP-DM cycle evaluation stage. After six predictive algorithms were evaluated, the best outcome for each parameter for every DMT was tested.

Table 9 shows that S1 had the best results in terms of Accuracy, i.e., the assertiveness of patient labeling, when combined with DT, RF, NB-K, and NB *data mining* techniques, achieving results above 85%. It is worth noting that the S3 scenario, which removed the *Shortness_of_smell_taste* attribute, combined with the DL algorithm produced a good result with 89,07% accuracy. Finally, it is important to note that the best results are not related with an MVA.

Table 9: Highest accuracy DMM for each DMT.

DMM	DMT	S	SM	MVA	DA	Accuracy (%)
3	DT	S1	Split Validation (80%)	N/A	SMOTE	96,25
7	RF	S1	Split Validation (70%)	N/A	SMOTE	91,36
105	DL	S3	Split Validation (80%)	N/A	SMOTE	89,07
27	NB-K	S1	Split Validation (80%)	N/A	SMOTE	87,45
19	NB	S1	Split Validation (70%)	N/A	SMOTE	86,32
13	RT	S1	Split Validation (70%)	N/A	SMOTE	68,26

The precision metric obtained excellent results, as can be seen in table 10, with all models achieving values greater than 90%, indicating that there is a high degree of certainty that the individuals classified with COVID-19 had the condition. It should be noted that the best precision result was 99.91% while employing the DMM3, which is defined by the DT algorithm, the S1 scenario, the Split Validation (80%) method, and the SMOTE Upsampling technique. The worst model, DMM83, based on the RF algorithm along with the S3 scenario, the SMOTE data approach, and Cross Validation SM, achieved an accuracy of 91.97%, which still remains an very high value.

Table 10: Highest precision DMM for each DMT.

DMM	DMT	S	SM	MVA	DA	Precision (%)
3	DT	S1	Split Validation (80%)	N/A	SMOTE	99,91
13	RT	S1	Split Validation (70%)	N/A	SMOTE	98,99
21	NB	S1	Split Validation (80%)	N/A	SMOTE	98,80
63	NB-K	S2	Split Validation (80%)	N/A	SMOTE	98,36
103	DL	S3	Split Validation (70%)	N/A	SMOTE	97,57
83	RF	S3	Cross Validation	N/A	SMOTE	91,97

In terms of sensitivity, table 11 shows that the predictive models with the best performance were RF, DT, DL, and NB-K. This means that these models guarantee that at least 82% of patients infected with COVID-19 were correctly predicted. Sensitivity is the most relevant metric for evaluating models in the context of this research, because it is risky to predict that a patient infected with SARS-CoV-2 is healthy. It is important to mention that this is the only metric in which DMM3 did not produce the greatest results, but its performance was still adequate, as it was the second DMM with the highest sensitivity value - 92,58%.

Table 11: Highest sensitivity DMM for each DMT.

DMM	DMT	S	SM	MVA	DA	Sensitivity (%)
9	RF	S1	Split Validation (80%)	N/A	SMOTE	93,42
3	DT	S1	Split Validation (80%)	N/A	SMOTE	92,58
105	DL	S3	Split Validation (80%)	N/A	SMOTE	89,37
28	NB-K	S1	Split Validation (80%)	Replace	SMOTE	82,57
22	NB	S1	Split Validation (80%)	Replace	SMOTE	79,09
18	RT	S1	Cross Validation	Replace	SMOTE	50,80

Specificity, on the other hand, indicates how many healthy patients were correctly predicted. The DMM3, which is defined by the DT algorithm, the S1 scenario, Split Validation with 80% of the data used for training, and the SMOTE Upsampling technique, reveals the best combination to achieve the best specificity result - 99,92%.

Table 12: Highest specificity DMM for each DMT.

DMM	DMT	S	SM	MVA	DA	Specificity (%)
3	DT	S1	Split Validation (80%)	N/A	SMOTE	99,92
13	RT	S1	Split Validation (70%)	N/A	SMOTE	99,63
21	NB	S1	Split Validation (80%)	N/A	SMOTE	99,12
63	NB-K	S2	Split Validation (80%)	N/A	SMOTE	98,76
103	DL	S3	Split Validation (70%)	N/A	SMOTE	98,18
83	RF	S3	Cross Validation	N/A	SMOTE	92,35

According to the majority of results, Split Validation was the most successful SM in this study. Split Validation is useful for large datasets and complex preprocessing procedures since it allows for some uncertainty regarding the robustness of the model. Cross Validation, on the other hand, requires more fully validated models, which slows down the process computational efficiency when the data is huge.

In terms of scenarios, S1, which included all of the attributes, generated the best results, followed by S3, which made it to six of the best performing models, and S2, which made it to two of the best performing models. This indicates that all of the attributes used in this case study should be included when evaluating the diagnosis of COVID-19 cases.

In terms of the MVA, it can be shown that the models performed better in general when the missing values were not replaced by their mean or mode value, depending on whether they are numerical or nominal. This is not surprising given that the dataset contains a significant number of missing values, with the majority of the attributes referring to the evolution of the symptoms, which are extremely subjective.

Oversampling, using the SMOTE Upsampling technique, produced the best results. Although the Undersampling method was tried, besides it did not produce the optimal results. This was most likely owing to the minority class being far smaller than the majority class, resulting in severe data loss.

5.2.2.3 SWOT Analysis

An internal and external SWOT analysis was performed to assess the developed predictive model's internal strengths and weaknesses, as well as the external opportunities and threats.

Table 13: SWOT analysis of use case performed.

Internal Environment	**External Environment**
Strenghts	Opportunities
Large original dataset for analysis	Sharing the final dataset for other research studies
Promotion of interoperability between home-based care and professionals	Sharing of openEHR COVID-19 structures built with the international community
High compliance of patients and healthcare professionals in this process	Scientific contribution to fight a pandemic situation
Weaknesses	Threats
Inconsistencies in the filling of some data	New predictive models with best results
An unbalanced dataset in terms of results	Virus variations that can change the disease's symptoms anytime.
Large number of missing values	

5.2.2.4 Conclusions

IT systems are changing the healthcare industry in ways never thought before, from the discovery of cures for diseases and the development of new treatment techniques to the improvement of patients' diagnoses and their enhancement. As a consequence, the benefits of using IT approaches in clinical procedures, such as improving the patients' quality of care and optimising the health institution's resources, have become widely recognized, from health centres to large-scale hospitals around the world.

Accordingly, one of the most promising outcomes of this project is the discovery of how quickly and efficiently globally recognized methodologies and standards such as openEHR can be implemented, as well as how they can interoperate with LSs already in place at each health institution.

Through the methodologies and research strategies chosen, it was possible to delineate a valid strategy starting from topics and key ideas that became more solid and justified with the revision of the literature. Additionally, this study demonstrated that the data generated by this new system can be used to train predictive machine learning models with acceptable performance. In this context, the openEHR standard was quickly adapted and implemented in a COVID-19 patient circuit, and latter the data from inquired patients was used to feed forecasting models based on the symptoms and current health state of the patients.

In terms of results, practically all models achieved accuracy rates over 80%, which is very impressive. DMM3 had the best results with 96.25% accuracy, 99.91% precision, 92.58% sensitivity, and 99.92% specificity, combining the dataset with all symptoms, the Split Validation (80%) method, and the Decision Tree algorithm without replacing the missing values. Hence, after collecting more data and subjecting the models to additional testing and rigorous evaluations, the predictive model could be latter implemented in a CDSS to assist healthcare professionals.

For future work, it is suggested to change some occurrences in the items of the openEHR templates developed, thus implying the need to fill them in order to reduce the number of missing values in the dataset, and therefore improve the trustworthiness of the produced results.

To end this conclusion, I would like to thank my colleagues, specially Diana Ferreira, and the health professionals involved for their collaboration in the development of this case study.

5.3 Conclusion

The fifth phase of the project started after the development of the test and production environments and the subsequent implementation of the developed artefact. Called Proof of Concept, it was in this phase that the different types of users performed their usability tests. The biggest difficulty experienced at this stage was the lack of knowledge of the openEHR specifications by health professionals, leaving the main burden of modelling to the same elements. Even so, the case studies and their respective proofs of concept were successfully developed, and the intention is to extend the proposed system to other clinical contexts. Therefore, the RQ9 and RQ10 imposed in stage IV are thus answered.

Chapter 6

Discussion and Conclusions

This thesis concludes with a detailed chapter titled "Discussion and Conclusions". Thus, in the first section (Section 6.1), the main goals accomplished through the completion of this doctorate thesis are discussed, including answering to the primary Research Question given in Chapter 1's Section 1.3. Following that, thoughts on future study on this research topic are shared (Section 6.2).

6.1 Main Achievements

This research work arose to address needs for structuring, exchange, and interpretation of data in a hospital setting. To this end, the following research question was initially imposed and will be answered below.

How to define the structuring, data exchange and knowledge discovery processes in a healthcare context, creating an interoperable Electronic Health Record (EHR)?

Several globally used standards were studied, and the standard that stood out the most was openEHR, demonstrating the most potential not only for its vast specification, but also for covering clinical workflows through its Process model (PROC) and Clinical Decision Model (CDM). Its broad scope justified its choice for this doctoral thesis in the biomedical engineering doctoral program. This work focused on developing an open and accessible artefact for healthcare institutions to create their own interoperable network between their HIS. The autonomy of each institution is as important as the sharing and understanding of information between each other. For this, the artefact was developed in several modules, ensuring that they work separately or together.

The demographic system developed translates into a simplistic way for healthcare institutions to organize themselves as work teams, with the creation of roles and capabilities associated with each member, and to represent their patients anonymously and securely. Note that the demographic specifications are also contained in the reference model and are archetypable. This module can be implemented on its own, for demographic and administrative management in a hospital context.

The clinical module was developed with the main purpose of creating each patient's EHR in a chronological, secure and versioned way. Each EHR tends to have a duration similar to the patient's lifetime, which can translate into thousands of records of various purposes. It is up to the clinical EHR system to organize this information in backoffice so that it can be easily and quickly consulted afterwards.

The openEHR governance module was translated into the development of a web platform capable of meeting the needs of interoperability. Through the platform, the operational templates are loaded and the conversion algorithm presented in section 4.2.1.1 is called in order to create an interpretable structure and its respective reference model to be instantiated when it is used. Through this process, syntactic interoperability is guaranteed, always starting from a stable version of the reference model, besides the guarantee of structural interoperability, treating archetypes as meaningful entities. Semantic interoperability is established by guaranteeing the meaning of concepts, avoiding their replication (managing archetypes and templates) and using the module for managing refsets through terminologies.

The speed and efficiency of developing clinical systems in different contexts and patient-oriented is one of the major advantages of the new system. The increased quality of data, greater flexibility and dynamics given to the most diverse platforms of health institutions, through the new system developed based on openEHR, highlights a greater symbiosis between health professionals and systems, leading to increased quality of care. These highlights of the new system can respond to RQ11 of the sixth stage. As an answer to RQ12, compared to other existing systems in the same field and performance, this study distinguishes itself by its complete and modular architecture, and its integration as a governance platform.

In response to RQ13, this doctoral thesis may inspire future work in the interoperation between institutions or even countries, by joining efforts both at the research level and at the governmental level. For data sharing to be possible, it is not enough just to develop this type of system based on international standards, but to make sure that, duly authorized, there is sharing of relevant information for research purposes or even health care in different geographic areas.

The development of this work took four years due to the enormity of the specifications studied and the projection of the most appropriate architectures and technologies to make this artefact completely useable in the future. Furthermore, it was really gratifying to realize how the work developed was able to assist health professionals in such a short period of time during an unprecedented pandemic situation that I will never forget. This final paragraph is dedicated to all health professionals and support employees who are assisting in the fight against the coronavirus. A heartfelt thank you to all.

6.2 Future Work

As this is a work in continuous development, some areas are still unexplored. At the national level, during these 4 years, a lack in the Portuguese openEHR community was noted, and one of the steps to follow would be to make the developed *AIDAEHR* web platform available online, so that all interested institutions could start contributing to the development of openEHR structures, thus obtaining a central national repository ready to be used. Also, the OPT conversion algorithm will have to be improved in the future to be able to process *work plan* templates and then allow the follow-up of clinical processes.

At the level of data mapping automation, another of the next goals to be achieved will be the development of a NER system capable of recognizing free text data patterns and suggesting openEHR structures for those data, as already mentioned in the conclusions of section 5.1.1.

At the level of information organization within each EHR, we intend to use clustering algorithms to group the *COMPOSITIONS* into their respective folders, through unsupervised or semi-supervised learning, through already predefined anchors.

Bibliography

[1] A. Abelha, C. Analide, J. Machado, J. Neves, M. Santos, and P. Novais. "Ambient intelligence and simulation in health care virtual scenarios." In: *Working Conference on Virtual Enterprises*. Springer. 2007, pp. 461–468.

[2] A. Abelha, E. Pereira, A. Brandão, F. Portela, M. F. Santos, J. Machado, and J. Braga. "Improving quality of services in maternity care triage system." In: *International Journal of E-Health and Medical Communications (IJEHMC)* 6.2 (2015), pp. 10–26.

[3] *About Health Level Seven International | HL7 International*. (Accessed on 08/03/2021). url: http://www.hl7.org/about/index.cfm?ref=nav.

[4] *About Node.js*. (Accessed on 08/21/2021). url: https://nodejs.org/en/about/.

[5] M. Alexandre, P. da Silva, V. Cesário, and C. Araújo. "A Decoupled Health Software Architecture using Microservices and OpenEHR Archetypes." In: *International Journal of Computer Applications* 975 (2020), p. 8887.

[6] *Ant Design - The world's second most popular React UI framework*. (Accessed on 08/24/2021). url: https://ant.design/.

[7] *Architecture Overview*. (Accessed on 08/20/2021). url: https://specifications.openehr.org/releases/BASE/latest/architecture_overview.html.

[8] *Architecture Overview*. https://specifications.openehr.org/releases/BASE/latest/architecture_overview.html. (Accessed on 08/14/2021).

[9] *Axios Docs*. (Accessed on 08/24/2021). url: https://axios-http.com/docs/intro.

[10] *Background & Overview of ICD-11 Before Implementation in 2022*. (Accessed on 08/03/2021). url: https://yes-himconsulting.com/icd-11-overview/.

[11] I. R. Bardhan and M. F. Thouin. "Health information technology and its impact on the quality and cost of healthcare delivery." In: *Decision Support Systems* 55.2 (2013), pp. 438–449.

[12] I. Benbasat, D. K. Goldstein, and M. Mead. "The case research strategy in studies of information systems." In: *MIS quarterly* (1987), pp. 369–386.

[13] Y. Bendavid, S. F. Wamba, and L. A. Lefebvre. “Proof of concept of an RFID-enabled supply chain in a B2B e-commerce environment.” In: (2006), pp. 564–568.

[14] A. Bento. “Como fazer uma revisão da literatura: Considerações teóricas e práticas.” In: *Revista JA (Associação Académica da Universidade da Madeira)* 7.65 (2012), pp. 42–44.

[15] E. V. Bernstam, W. R. Hersh, I. Sim, D. Eichmann, J. C. Silverstein, J. W. Smith, and M. J. Becich. “Unintended consequences of health information technology: a need for biomedical informatics.” In: *Journal of biomedical informatics* 43.5 (2010), p. 828.

[16] O. Bodenreider, R. Cornet, and D. J. Vreeman. “Recent developments in clinical terminologies—SNOMED CT, LOINC, and RxNorm.” In: *Yearbook of medical informatics* 27.01 (2018), pp. 129–139.

[17] T. Cardoso, I. Alarcão, and J. A. Celorico. *Revisão da literatura e sistematização do conhecimento*. Porto Editora, 2010.

[18] S. C. O. Castro. “Normas, nomenclaturas e uniformização do registo clínico.” Doctoral dissertation. 2013.

[19] V. Chooprayoon and C. C. Fung. “TECTAM: an approach to study Technology Acceptance Model (TAM) in gaining knowledge on the adoption and use of e-commerce/e-business technology among small and medium enterprises in Thailand.” In: *InTech* (2010), pp. 31–38.

[20] J. J. Cimino and E. H. Shortliffe. *Biomedical Informatics: Computer Applications in Health Care and Biomedicine (Health Informatics)*. Springer-Verlag, 2006.

[21] *Classification of Diseases (ICD)*. (Accessed on 08/03/2021). url: https://www.who.int/classifications/classification-of-diseases.

[22] W. S. Cleveland. “Data science: an action plan for expanding the technical areas of the field of statistics.” In: *International statistical review* 69.1 (2001), pp. 21–26.

[23] *Clinical Knowledge Manager*. (Accessed on 09/02/2021). url: https://ckm.openehr.org/ckm/.

[24] *Commission Recommendation of 2 July 2008 on cross-border interoperability of electronic health record systems*. (Accessed on 08/01/2021). url: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32008H0594&from=PT.

[25] B. company. *Archetype Designer*. (Accessed on 09/02/2021). url: https://tools.openehr.org/designer/#/.

[26] C. M. Costa, M. Menárguez-Tortosa, and J. T. Fernández-Breis. “Clinical data interoperability based on archetype transformation.” In: *Journal of biomedical informatics* 44.5 (2011), pp. 869–880.

[27] *Current and future standardization issues in the e-Health domain: Achieving interoperability.* (Accessed on 07/31/2021). url: https://share.ansi.org/shared%20documents/Standards%20Activities/Healthcare%20Informatics%20Technology%20Standards%20Panel/Standardization%20Committees/International%20Landscape/eHealthStandardizationReport-Part2-finalversion%20-%202005-03-011.pdf.

[28] F. D. Davis. "Perceived usefulness, perceived ease of use, and user acceptance of information technology." In: *MIS quarterly* (1989), pp. 319–340.

[29] F. D. Davis, R. P. Bagozzi, and P. R. Warshaw. "User acceptance of computer technology: A comparison of two theoretical models." In: *Management science* 35.8 (1989), pp. 982–1003.

[30] R. S. Dick, E. B. Steen, D. E. Detmer, et al. *The computer-based patient record: an essential technology for health care*. National Academies Press, 1997.

[31] J. Duarte, M. Salazar, C. Quintas, M. Santos, J. Neves, A. Abelha, and J. Machado. "Data quality evaluation of electronic health records in the hospital admission process." In: *2010 IEEE/ACIS 9th International Conference on Computer and Information Science*. IEEE. 2010, pp. 201–206.

[32] G. Duftschmid, T. Wrba, and C. Rinner. "Extraction of standardized archetyped data from Electronic Health Record systems based on the Entity-Attribute-Value Model." In: *International journal of medical informatics* 79.8 (2010), pp. 585–597.

[33] E. H. Durfee, V. R. Lesser, and D. D. Corkill. "Trends in cooperative distributed problem solving." In: *IEEE Transactions on knowledge and data Engineering* (1989).

[34] *eHealth in Latin America and the Caribbean: interoperability standards review*. (Accessed on 08/22/2021). url: https://iris.paho.org/handle/10665.2/28189.

[35] *Express - Node.js web application framework*. (Accessed on 08/24/2021). url: https://expressjs.com/.

[36] G. Eysenbach. "What is e-health?" In: *Journal of medical Internet research* 3.2 (2001), e20.

[37] U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy. "Advances in knowledge discovery and data mining." In: (1996).

[38] D. Ferreira, S. Silva, A. Abelha, and J. Machado. "Recommendation system using autoencoders." In: *Applied Sciences* 10.16 (2020), p. 5510.

[39] *Fetch - APIs*. (Accessed on 08/24/2021). url: https://developer.mozilla.org/pt-BR/docs/Web/API/Fetch_API/Using_Fetch.

[40] M. Fishbein and I. Ajzen. "Belief, attitude, intention, and behavior: An introduction to theory and research." In: *Philosophy and Rhetoric* 10.2 (1977).

[41] F. M. Gabor-Harosa, O. P. Stan, L. Daina, and F. Mocean. “Proposed model for a Romanian register of chronic diseases in children.” In: *Computer methods and programs in biomedicine* 130 (2016), pp. 198–204.

[42] S. Garde, R. Chen, H. Leslie, T. Beale, I. McNICOLL, and S. Heard. “Archetype-based knowledge management for semantic interoperability of electronic health records.” In: *Medical Informatics in a United and Healthy Europe*. IOS Press, 2009, pp. 1007–1011.

[43] S. Garde, P. Knaup, E. J. Hovenga, and S. Heard. “Towards semantic interoperability for electronic health records.” In: *Methods of information in medicine* 46.03 (2007), pp. 332–343.

[44] S. Ghazinoory, M. Abdi, and M. Azadegan-Mehr. “SWOT methodology: a state-of-the-art review for the past, a framework for the future.” In: *Journal of business economics and management* 12.1 (2011), pp. 24–48.

[45] B. Gillham. *Case study research methods*. Bloomsbury Publishing, 2000.

[46] R. L. Glass, I. Vessey, and V. Ramesh. “Research in software engineering: an analysis of the literature.” In: *Information and Software technology* 44.8 (2002), pp. 491–506.

[47] L. Grandia. “Healthcare information systems: a look at the past, present, and future.” In: *Health Catalyst* (2017), pp. 1–4.

[48] C. Health. *The HL7 Evolution: Comparing HL7 Version 2 to Version 3, Including a History of Version 2*. 2009.

[49] G. Hernandez-Ibarburu, D. Perez-Rey, E. Alonso-Oset, R. Alonso-Calvo, D. Voets, C. Mueller, B. Claerhout, and N. Custodix. “ICD-10-PCS extension with ICD-9 procedure codes to support integrated access to clinical legacy data.” In: *International Journal of Medical Informatics* 122 (2019), pp. 70–79. issn: 1386-5056. doi: https://doi.org/10.1016/j.ijmedinf.2018.11.002. url: https://www.sciencedirect.com/science/article/pii/S1386505618302181.

[50] R. Hillestad, J. Bigelow, A. Bower, F. Girosi, R. Meili, R. Scoville, and R. Taylor. “Can electronic medical record systems transform health care? Potential health benefits, savings, and costs.” In: *Health affairs* 24.5 (2005), pp. 1103–1117.

[51] *HIMSS Proposes a New Interoperability Definition in Healthcare*. (Accessed on 07/30/2021). url: https://www.imprivata.com/blog/himss-proposes-a-new-interoperability-definition-in-healthcare-what-it-means-for-patients-and-professionals.

[52] *https://www.openehr.org/programs/specification/*. (Accessed on 08/20/2021). url: https://www.openehr.org/programs/specification/.

[53] S. M. Huff, R. A. Rocha, B. E. Bray, H. R. Warner, and P. J. Haug. "An event model of medical information representation." In: *Journal of the American Medical Informatics Association* 2.2 (1995), pp. 116–134.

[54] S. Iancu. *specifications-ITS-JSON/components/RM/Release-1.1.0*. (Accessed on 09/01/2021). May 2021. url: https://github.com/openEHR/specifications-ITS-JSON/tree/master/components/RM/Release-1.1.0.

[55] *ICD-10 Version:2019*. (Accessed on 08/03/2021). url: https://icd.who.int/browse10/2019/en#/I.

[56] "IEEE Standard Dictionary of Electrical and Electronics Terms." In: *IEEE Transactions on Power Apparatus and Systems* PAS-99.6 (1980), 37a–37a. doi: 10.1109/TPAS.1980.319816.

[57] H. L. S. Internationa. *HL7 Governance and Operations Manual*. (Accessed on 08/03/2021). 2021. url: http://www.hl7.org/documentcenter/public/membership/HL7_Governance_and_Operations_Manual.pdf.

[58] *Interoperability in Healthcare | HIMSS*. (Accessed on 07/29/2021). url: https://www.himss.org/resources/interoperability-healthcare.

[59] *ISO - ISO 18308:2011 - Health informatics — Requirements for an electronic health record architecture*. (Accessed on 07/29/2021). url: https://www.iso.org/standard/52823.html.

[60] *ISO 13606 Standard - EHR Interoperability*. (Accessed on 08/04/2021). url: http://www.en13606.org/information.html.

[61] *ISO/IEC TR 10000-1:1998(en), Information technology — Framework and taxonomy of International Standardized Profiles — Part 1: General principles and documentation framework*. (Accessed on 07/31/2021). url: https://www.iso.org/obp/ui/.

[62] D. G. Katehakis, A. Kouroubali, and I. Karatzanis. "Health Level Seven." In: *Interoperability in eHealth - A Course* 4 (2020).

[63] *Knowledge Base – LOINC*. (Accessed on 08/03/2021). url: https://loinc.org/kb/loinc-release-notes/.

[64] T. A. Koleck, C. Dreisbach, P. E. Bourne, and S. Bakken. "Natural language processing of symptoms documented in free-text narratives of electronic health records: a systematic review." In: *Journal of the American Medical Informatics Association* 26.4 (2019), pp. 364–379.

[65] D. Kumawat and A. Pavate. "Correlation of NOSQL & SQL Database." In: *IOSR J. Comput. Eng.(IOSR-JCE)* 18 (2016), pp. 70–74.

[66] E. Lamine, W. Guédria, A. Rius Soler, J. Ayza Graells, F. Fontanili, L. Janer-García, and H. Pingaud. "An Inventory of Interoperability in Healthcare Ecosystems: Characterization and Challenges." In: *Enterprise Interoperability: INTEROP-PGSO Vision* 1 (2017), pp. 167–198.

[67] K. C. Laudon, J. P. Laudon, et al. "Essentials of management information systems." In: (2011).

[68] P. J. Lavrakas. *Encyclopedia of survey research methods*. Sage publications, 2008.

[69] Y. Li and S. Manoharan. "A performance comparison of SQL and NoSQL databases." In: (2013), pp. 15–19.

[70] Z. H. Liu, B. Hammerschmidt, D. McMahon, H. Chang, Y. Lu, J. Spiegel, A. C. Sosa, S. Suresh, G. Arora, and V. Arora. "Native JSON datatype support: maturing SQL and NoSQL convergence in Oracle database." In: *Proceedings of the VLDB Endowment* 13.12 (2020), pp. 3059–3071.

[71] *LOINC Term Basics – LOINC*. (Accessed on 08/03/2021). url: https://loinc.org/get-started/loinc-term-basics/.

[72] J. Machado, A. Abelha, J. Neves, and M. Santos. "Ambient intelligence in medicine." In: *2006 IEEE Biomedical Circuits and Systems Conference*. IEEE. 2006, pp. 94–97.

[73] J. M. Machado, V. Alves, A. Abelha, and J. Neves. "Ambient intelligence via multiagent systems in the medical arena." In: (2007).

[74] N. Marangunić and A. Granić. "Technology acceptance model: a literature review from 1986 to 2013." In: *Universal access in the information society* 14.1 (2015), pp. 81–95.

[75] S. T. March and V. C. Storey. "Design science in the information systems discipline: an introduction to the special issue on design science research." In: *MIS quarterly* (2008), pp. 725–730.

[76] A. Mardan, Mardan, and Corrigan. *Practical Node.js*. Springer, 2018.

[77] C. Martínez-Costa, M. Menárguez-Tortosa, and J. Fernandez-Breis. "An approach for the semantic interoperability of ISO EN 13606 and OpenEHR archetypes." In: *Journal of biomedical informatics* 43 (Oct. 2010), pp. 736–46. doi: 10.1016/j.jbi.2010.05.013.

[78] S. D. McArthur, E. M. Davidson, V. M. Catterson, A. L. Dimeas, N. D. Hatziargyriou, F. Ponci, and T. Funabashi. "Multi-agent systems for power engineering applications—Part I: Concepts, approaches, and technical challenges." In: *IEEE Transactions on Power systems* 22.4 (2007), pp. 1743–1752.

[79] C. J. McDonald. "The barriers to electronic medical record systems and how to overcome them." In: *Journal of the American Medical Informatics Association* 4.3 (1997), pp. 213–221.

[80] M. Miranda, A. Abelha, M. Santos, J. Machado, and J. Neves. "A group decision support system for staging of cancer." In: *International Conference on Electronic Healthcare*. Springer. 2008, pp. 114–121.

[81] M. Miranda, J. Duarte, A. Abelha, J. M. Machado, and J. Neves. "Interoperability and healthcare." In: (2009).

[82] M. Miranda, M. Salazar, F. Portela, M. Santos, A. Abelha, J. Neves, and J. Machado. "Multi-agent systems for hl7 interoperability services." In: *Procedia Technology* 5 (2012), pp. 725–733.

[83] J. L. C. de Moraes, W. L. de Souza, L. F. Pires, and A. F. do Prado. "A methodology based on openEHR archetypes and software agents for developing e-health applications reusing legacy systems." In: *Computer methods and programs in biomedicine* 134 (2016), pp. 267–287.

[84] A. Moreno-Conde, D. Moner, W. D. d. Cruz, M. R. Santos, J. A. Maldonado, M. Robles, and D. Kalra. "Clinical information modeling processes for semantic interoperability of electronic health records: systematic review and inductive analysis." In: *Journal of the American Medical Informatics Association* 22.4 (2015), pp. 925–934.

[85] M. J. Mortenson and R. Vidgen. "A computational literature review of the technology acceptance model." In: *International Journal of Information Management* 36.6 (2016), pp. 1248–1259.

[86] P. Muñoz, J. D. Trigo, I. Martínez, A. Muñoz, J. Escayola, and J. García. "The ISO/EN 13606 standard for the interoperable exchange of electronic health records." In: *Journal of Healthcare Engineering* 2.1 (2011), pp. 1–24.

[87] T. B. Murdoch and A. S. Detsky. "The inevitable application of big data to health care." In: *Jama* 309.13 (2013), pp. 1351–1352.

[88] J. Neves, M. R. Martins, J. Vilhena, J. Neves, S. Gomes, A. Abelha, J. Machado, and H. Vicente. "A soft computing approach to kidney diseases evaluation." In: *Journal of medical systems* 39.10 (2015), pp. 1–9.

[89] *openEHR - Reference Model (RM) Component*. (Accessed on 02/09/2021). url: https://specifications.openehr.org/releases/RM/latest.

[90] *openEHR CDS, Guidelines and Planning Examples*. (Accessed on 08/20/2021). url: https://specifications.openehr.org/releases/PROC/latest/process_examples.html.

[91] *openEHR Platform Service Model*. (Accessed on 08/20/2021). url: https://specifications.openehr.org/releases/SM/latest/openehr_platform.html.

[92] *Os números da covid-19 em Portugal | Coronavírus | PÚBLICO*. (Accessed on 09/08/2021). url: https://www.publico.pt/2020/03/26/infografia/situacao-registada-portugal-480.

[93] N. Pahl and A. Richter. *SWOT analysis: Idea, methodology and a practical approach*. Grin Verlag Munich, 2007.

[94] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee. “A design science research methodology for information systems research.” In: *Journal of management information systems* 24.3 (2007), pp. 45–77.

[95] A. Pereira, F. Marins, B. Rodrigues, F. Portela, M. F. Santos, J. Machado, F. Rua, Á. Silva, and A. Abelha. “Improving quality of medical service with mobile health software.” In: *Procedia Computer Science* 63 (2015), pp. 292–299.

[96] R. Pereira, J. Duarte, M. Salazar, M. Santos, A. Abelha, and J. Machado. “Usability of an electronic health record.” In: *2012 IEEE International Conference on Industrial Engineering and Engineering Management*. IEEE. 2012, pp. 1568–1572.

[97] *Population Health Management: A Roadmap for Provider-Based Automation in a New Era of Healthcare*. (Accessed on 07/29/2021). url: http://c4fd63cb482ce6861463-bc6183f1c18e748a49b87a25911a0555.r93.cf2.rackcdn.com/iHT2_BigData_2013.pdf.

[98] *React Router: Declarative Routing for React.js*. (Accessed on 08/24/2021). url: https://reactrouter.com/.

[99] *React – A JavaScript library for building user interfaces*. (Accessed on 08/24/2021). url: https://reactjs.org/.

[100] M. Saunders, P. Lewis, and A. Thornhill. “Research methods forbusiness students.” In: *Essex: Prentice Hall: Financial Times* (2003).

[101] B. Schmidt. “Proof of principle studies.” In: *Epilepsy research* 68.1 (2006), pp. 48–52.

[102] *Sequelize ORM*. (Accessed on 08/24/2021). url: https://sequelize.org/.

[103] A. B. Sergey, D. B. Alexandr, and A. T. Sergey. “Proof of concept center—a promising tool for innovative development at entrepreneurial universities.” In: *Procedia-Social and Behavioral Sciences* 166 (2015), pp. 240–245.

[104] A. A. Shaikh and H. Karjaluoto. “Making the most of information technology & systems usage: A literature review, framework and future research agenda.” In: *Computers in Human Behavior* 49 (2015), pp. 541–566.

[105] *Simple Oracle Document Access (SODA)*. (Accessed on 08/02/2021). url: https://docs.oracle.com/en/database/oracle/simple-oracle-document-access/index.html.

[106] *SNOMED - 5-Step Briefing*. (Accessed on 08/04/2021). url: https://www.snomed.org/snomed-ct/five-step-briefing.

[107] *SNOMED - Why SNOMED CT?* (Accessed on 08/04/2021). url: https://www.snomed.org/snomed-ct/why-snomed-ct.

[108] *SNOMED CT | NRCeS*. (Accessed on 08/04/2021). url: https://www.nrces.in/standards/snomed-ct.

[109] *Stack Overflow Developer Survey 2019*. https://insights.stackoverflow.com/survey/2019. (Accessed on 08/24/2021).

[110] *Stack Overflow Developer Survey 2020*. (Accessed on 08/21/2021). url: https://insights.stackoverflow.com/survey2020#technology.

[111] *Stack Overflow Developer Survey 2021*. (Accessed on 08/24/2021). url: https://insights.stackoverflow.com/survey/2021#technology.

[112] *Tailwind CSS - Rapidly build modern websites without ever leaving your HTML.* (Accessed on 08/24/2021). url: https://tailwindcss.com/.

[113] *Template Designer by Ocean Health Systems*. (Accessed on 09/02/2021). url: http://downloads.oceaninformatics.com/downloads/TemplateDesigner/.

[114] *The Difference between EMR and EHR | Ksatria Medical System*. (Accessed on 07/29/2021). url: https://ksatria.com/blog/emr-vs-ehr/.

[115] *Unified Medical Language System (UMLS)*. (Accessed on 09/15/2021). url: https://www.nlm.nih.gov/research/umls/index.html.

[116] V. Venkatesh and H. Bala. "Technology acceptance model 3 and a research agenda on interventions." In: *Decision sciences* 39.2 (2008), pp. 273–315.

[117] V. Venkatesh and F. D. Davis. "A theoretical extension of the technology acceptance model: Four longitudinal field studies." In: *Management science* 46.2 (2000), pp. 186–204.

[118] M. Virtanen, B. Ustun, J. M. Rodrigues, G. Surjan, A. Rector, V. N. Stroetmann, K. A. Stroetmann, P. Lewalle, P. E. Zanstra, and D. Kalra. "Semantic interoperability for better health and safer healthcare: deployment and research roadmap for Europe." In: *European Comission* (2013). doi: 10.2759/38514.

[119] C. N. Vorisek, S. A. I. Klopfenstein, J. Sass, M. Lehne, C. O. Schmidt, and S. Thun. "Evaluating Suitability of SNOMED CT in Structured Searches for COVID-19 Studies." In: *Public Health and Informatics*. IOS Press, 2021, pp. 88–92.

[120] W. L. Wallace. *The logic of science in sociology sound recording*. Transaction Publishers, 1971.

[121] *What Is a Database | Oracle*. (Accessed on 08/02/2021). url: https://www.oracle.com/database/what-is-database/.

[122] J. L. Whitten and L. D. Bentley. *Systems Analysis and Design Methods*. 2005.

[123] C. Williams. "Research methods." In: *Journal of Business & Economics Research (JBER)* 5.3 (2007).

[124] A. Wulff, C. Baier, S. Ballout, E. Tute, K. K. Sommer, M. Kaase, A. Sargeant, C. Drenkhahn, D. Schlüter, M. Marschollek, et al. "Transformation of microbiology data into a standardised data representation using OpenEHR." In: *Scientific reports* 11.1 (2021), pp. 1–12.

[125] A. Wulff, M. Mast, M. Hassler, S. Montag, M. Marschollek, and T. Jack. "Designing an openEHR-based pipeline for extracting and standardizing unstructured clinical data using natural language processing." In: *Methods of information in medicine* 59.S 02 (2020), e64–e78.

[126] R. K. Yin. *Case study research: Design and methods*. Vol. 5. sage, 2009.

[127] S. Yu, D. Berry, and J. Bisbal. "Clinical coverage of an archetype repository over SNOMED-CT." In: *Journal of biomedical informatics* 45.3 (2012), pp. 408–418.

Appendix A

Interoperability in Healthcare Information Systems

This questionnaire is a component of a study conducted as part of a PhD thesis in Biomedical Engineering at the University of Minho. The results will be used solely for academic purposes, and it is emphasized that respondents' responses reflect only their own personal view. By anonymizing your responses, your data will be protected.

1. **Which Portuguese institution are you affiliated with?**

   ____

2. **How many databases do you manage on a daily basis as part of your job?**
   ☐ 1-2 ☐ 2-4 ☐ 4-6 ☐ 6-8 ☐ 8-10

3. **How many distinct database versions do you handle on a daily basis as part of your job?**
   ☐ 1-2 ☐ 2-4 ☐ 4-6 ☐ 6-8 ☐ 8-10

4. **Are standards implemented in your institution to ensure the interoperability of information systems?**
   ☐ Yes ☐ No

5. **If you replied "Yes"to the above question, please identify the standards that your institution adopts.**

   ____

   ____

6. **How would you rate the interoperability of your institution's information systems on an intra-hospital level?**

   ☐ Inexistent ☐ Substancial ☐ Satisfatory ☐ Efficient

7. **How would you rate the interoperability of your institution's information systems on an inter-hospital level?**

   ☐ Inexistent ☐ Substancial ☐ Satisfatory ☐ Efficient

8. **Is your institution connected to other Extra-Hospital Information Systems (EHIS)?**

   ☐ Yes ☐ No

9. **Please indicate any area(s) in which you wish to increase your institution's interoperability.**

   ____

   ____

   ____

Thank you for your cooperation.

Appendix B

System's Technology Acceptance Assessment

This appendix contains the TAM questionnaire designed to assess the proposed artifact's acceptance by target users. All survey questions are scored on a seven-point Likert scale, described bellow.

1 - strong disagree **2** - moderate disagree **3** - some disagree **4** - neutral
5 - some agree **6** - moderate agree **7** - strong agree

Table 14: Survey to evaluate the new system's acceptance based on the TAM3 on a 7-point Likert scale.

	1	**2**	**3**	**4**	**5**	**6**	**7**
Perceived Ease of Use							
I believe the system is simple to use.	○	○	○	○	○	○	○
My interaction with the system is simple and understandable.	○	○	○	○	○	○	○
With the system, I have easy access to information.	○	○	○	○	○	○	○
With the system, I am able to easily organize my tasks.	○	○	○	○	○	○	○
Subjective Norm							
Healthcare and IT professionals believe we should use the system.	○	○	○	○	○	○	○
In general, the institution has supported the system's use.	○	○	○	○	○	○	○
Image							
Health professionals who use the system are more recognizable than those who do not.	○	○	○	○	○	○	○
In such a competitive industry, having the system is a							

Continued on next page

Table 14 – *Continued from previous page*

	1	**2**	**3**	**4**	**5**	**6**	**7**
tremendous asset for healthcare institutions.	○	○	○	○	○	○	○
Job Relevance							
Using the system is crucial and necessary for me to execute my job-related tasks.	○	○	○	○	○	○	○
Output Quality							
I believe that the system will enable me to achieve high-quality results in general.	○	○	○	○	○	○	○
Result Demonstrability							
It is simple for me to describe to others the outcomes that can be reached using the system.	○	○	○	○	○	○	○
Perceived Usefulness							
Using the system can help me perform better and be more efficient in my job-related roles.	○	○	○	○	○	○	○
I believe the system is useful and an excellent tool for managing patients and their clinical information.	○	○	○	○	○	○	○
Computer Self-efficacy							
I am able to operate the system even when there is no one nearby to assist me.	○	○	○	○	○	○	○
I would be able to operate the system only if someone demonstrated how to do so beforehand.	○	○	○	○	○	○	○
Perceptions of External Control							
Given the necessary resources and knowledge, I find it simple to use the system.	○	○	○	○	○	○	○
The healthcare institutions supply the essential resources for system use.	○	○	○	○	○	○	○
Computer Anxiety							
I believe myself to be at ease and at comfortable when using a computer.	○	○	○	○	○	○	○

Continued on next page

Table 14 – *Continued from previous page*

	1	**2**	**3**	**4**	**5**	**6**	**7**
I am fearful and apprehensive when it comes to using a computer.	○	○	○	○	○	○	○
Perceived Enjoyment							
I believe using the system to be pleasurable and enjoyable.	○	○	○	○	○	○	○
Intention to Use the Tecnhology							
I plan to take advantage of the system.	○	○	○	○	○	○	○
I intend to make frequent and consistent use of the system.	○	○	○	○	○	○	○

Thank you for your cooperation.

Appendix C

Publications

This appendix contains the scientific papers published during this doctoral thesis, as well as the manuscripts that have already been submitted and are awaiting publication.

C.1 Prediction of COVID-19 Diagnosis based on OpenEHR Artefacts

Authors: Daniela Oliveira, Rui Miranda, Diana Ferreira, António Abelha, and José Machado

Title: Prediction of COVID-19 Diagnosis based on OpenEHR Artefacts

Year of Submission: 2021

Abstract: The complexity and momentum of monitoring COVID-19 patients calls for the usage of agile and scalable data structure methodologies. A system for tracking symptoms and health conditions of suspected or confirmed SARS-CoV-2 infected patients was developed based on the openEHR architecture. All data on the evolutionary status of patients in home care as well as the results of their COVID-19 test were used to train different ML algorithms, with the aim of developing a predictive model capable of identifying COVID-19 infections according to the severity of symptoms identified by patients. The results obtained were promising, with the best model achieving an accuracy of 96.25%, a precision of 99.91%, a sensitivity of 92.58%, and a specificity of 99.92%, using the Decision Tree algorithm and the Split Validation method.

C.2 OpenEHR modeling: improving clinical records during the COVID-19 pandemic

Authors: Daniela Oliveira. Rui Miranda. Pedro Leuschner. Nuno Abreu. Manuel Filipe Santos. Antonio Abelha and José Machado

Title: OpenEHR modeling: improving clinical records during the COVID-19 pandemic

Journal: Health and Technology, Springer Nature

Year of Publication: 2021

Abstract: The COVID-19 pandemic had put pressure on various national healthcare systems, due to the lack of health professionals and exhaustion of those avaliable, as well as lack of interoperability and inability to restructure their IT systems. Therefore, the restructuring of institutions at all levels is essential, especially at the level of their information systems. Furthermore, the COVID-19 pandemic had arrived in Portugal at March 2020, with a breakout on the northern region. In order to quickly respond to the pandemic, the CHUP healthcare institution, known as a research center, has embraced the challenge of developing and integrating a new approach based on the openEHR standard to interoperate with the institution's existing information and its systems. An openEHR clinical modelling methodology was outlined and adopted, followed by a survey of daily clinical and technical requirements. With the arrival of the virus in Portugal, the CHUP institution has undergone through constant changes in their working methodologies as well as their openEHR modelling. As a result, an openEHR patient care workfow for COVID-19 was developed.

Keywords: COVID-19, OpenEHR, Archetypes, Templates, Interoperability, Healthcare information systems, Electronic health record

State of Publication: Published

DOI: 10.1007/s12553-021-00556-4

C.3 Steps towards an Healthcare Information Model based on openEHR

Authors: Daniela Oliveira, Rui Miranda, Francini Hak, Nuno Abreu, Pedro Leuschner, António Abelha abd José Machado

Title: Steps towards an Healthcare Information Model based on openEHR

Conference Proceedings: The 4th International Conference on Emerging Data and Industry 4.0 (EDI40), Vol. 184, pp. 893-898. Procedia Computer Science, Elsevier

Year of Publication: 2021

Abstract: During COVID-19 pandemic crisis, healthcare institutions globally were experiencing a VUCA - Volatile, Uncertain, Complex, and Ambiguous - environment. Effcient clinical and administrative management had never been so emergent. To achieve this goal, different components of the Healthcare Information System (HIS) must cooperate and interoperate flawlessly. Data standardization is a necessary step towards normalization and interoperability between existing Legacy Systems (LSs), and provides for longitudinal, highly reliable and persistent Electronic Health Records (EHRs). The openEHR standard was chosen for its overall dual domain architecture, where the more dynamic clinical information model may evolve independently from the relatively stable Reference Model (RM). Its Information Model (IM) comprises demographic, administrative and clinical systems. Critical clinical terms have been aligned to the FHIR HL7 standard, as to further support interoperability.

Keywords: Electronic Health Record, Healthcare Information Systems, OpenEHR, Clinical Information model, Reference Model

State of Publication: Published

DOI: 10.1016/j.procs.2021.04.015

C.4 Management of a Pandemic Based on an openEHR approach

Authors: Daniela Oliveira, Rui Miranda, Nuno Abreu, Pedro Leuschner, António Abelha, Manuel Santos and José Machado

Title: Management of a Pandemic Based on an openEHR approach

Conference Proceedings: The 10th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH 2020), Vol. 177, pp. 522-527. Procedia Computer Science, Elsevier

Year of Publication: 2020

Abstract: The COVID-19 pandemic has collapsed several national health systems, due to the lack of healthcare professionals and exhaustion of those employed, as well as the lack of interoperability and capacity to restructure their informatic systems. Therefore, the restructuring of institutions at all levels is essential, mainly at the level of their Information Systems. When the COVID-19 pandemic had spread to Portugal in March 2020, with a breakout on the northern region, the Centro Hospitalar Universitário do Porto (CHUP) healthcare institution had felt the need to develop and integrate a new approach based on the openEHR standard to interoperate with the institution's existing information systems, with the aim of responding quickly to the pandemic's evolution.

Keywords: COVID-19, openEHR, Healthcare Information Systems, Electronic Health Record, Interoperability

State of Publication: Published

DOI: 10.1016/j.procs.2020.10.072

C.5 An OpenEHR Adoption in a Portuguese Healthcare Facility

Authors: Francini Hak, Daniela Oliveira, Nuno Abreu, Pedro Leuschner, António Abelha and Manuel Santos

Title: An OpenEHR Adoption in a Portuguese Healthcare Facility

Conference Proceedings: International Workshop on Hospital 4.0 (Hospital), Vol. 170, pp. 1047-1052. Procedia Computer Science, Elsevier

Year of Publication: 2020

Abstract: The quality and safety of clinical decisions depend to a large extent on the knowledge acquired by the records of health professionals. However, a traditional Electronic Health Record (EHR) has become insufficient in terms of knowledge acquisition and clinical decision support. The development of these aspects may bring marked improvements in the quality and safety of health care. The usage of open models promotes interoperability between systems, minimizing the impact of information systems on the efficient production of knowledge useful for clinical decisions. In this sense, this article describes an implementation project of a system that support the production and use of knowledge in clinical environments, based on OpenEHR two levels modelling open data approach, in a healthcare facility on the north of Portugal.

Keywords: openEHR, Healthcare Information Systems, Electronic Health Record, Interoperability

State of Publication: Published

DOI: 10.1016/j.procs.2020.03.075

C.6 Review of Trends in Automatic Human Activity Recognition based in Synthetic Data

Authors: Ana Coimbra, Cristiana Neto, Diana Ferreira, Júlio Duarte, Daniela Oliveira, Francini Hak, Filipe Gonçalves, Joaquim Fonseca, Nicolas Lori, António Abelha and José Machado

Title: Review of Trends in Automatic Human Activity Recognition based in Synthetic Data

Conference Proceedings: Intelligent Data Engineering and Automated Learning – IDEAL 2020. Lecture Notes in Computer Science, Vol. 12490, pp. 368-376, Springer

Year of Publication: 2020

Abstract: Driverless vehicles are more and more becoming a reality. However, people still have some concerns in using them, the main concern is fear, hence the importance of creating a surveillance system inside those vehicles. For the detection and classification of human movements to be possible it is necessary to train the system with data representative enough for all kinds of possibilities. Although the production of large quantities of data becomes an expensive process and adds the problem of data protection, the use of synthetic data once they are artificially generated allows lower costs and eliminates the problem of data protection. A bibliographic study was carried out in this paper with articles from 2017 or later on the use of synthetic data. In these studies, it is noted that synthetic data is widely used with good results. As far as image capture is concerned, they show that 3D cameras have better results, but they are more expensive, so 2D cameras are more often used with later conversion to 3D images. The stitched puppet (SP) model is capable of adapting to the most difficult poses having obtained good results in its application in the FAUST dataset.

Keywords: Autonomous car, Multisensory fusion, Audio-visual synthetic data

State of Publication: Published

DOI: 10.1007/978-3-030-62365-4_35

C.7 Step Towards Interoperability in Nursing Practice

Authors: Daniela Oliveira, Júlio Duarte, António Abelha and José Machado

Title: Step Towards Interoperability in Nursing Practice

Conference Proceedings: Data Analytics in Medicine: Concepts, Methodologies, Tools, and Applications. IGI Global, pp. 865-878, 2020

Year of Publication: 2020

Abstract: Hospital inpatient care compromises one of the most demanding services in health institutions for providing a careful and continuous healthcare assistance. Such demands require a constant update of the patients' health record allied with support systems responsible for monitoring their clinical information. In this context, the problem in this study becomes a process of continuous improvement. To define the case study, it was necessary to use research tools such as questionnaires and interviews. With these techniques, it was possible to delineate the state and dimension of the problem. Subsequently, the approach and solution was established and a new web platform for the daily monitoring of patients was proposed focused on nurses. The tool incorporates a real-time data visualization, and a patient record during an inpatient care episode. Moreover, this article also highlights the required adaptability of this platform for each health unit according to needs. With this solution, it is expected to correct many of the problems detected through quantitative results.

Keywords: Healthcare, Hospital Information System, Interoperability, Multi-Agent System, Nursing Practice

State of Publication: Published

DOI: 10.4018/978-1-7998-1204-3.ch046

C.8 Intelligent Support System for the Provision of Inpatient Care

Authors: Sónia Faria. Daniela Oliveira. António Abelha. and José Machado

Title: Intelligent Support System for the Provision of Inpatient Care

Conference Proceedings: Advances in Intelligent Systems and Computing, Vol. 1137, pp.364-374.

Year of Publication: 2020

Abstract: Inpatient care is seen as a rigorous healthcare environment, as several daily tasks are performed to provide adequate treatment to inpatients and a minor flaw in these tasks may result in irreversible damage to patients. It is therefore required that the information related to the patient is always updated and available to all health professionals. Thus, comes up the motivation of the project described in this paper, which presents an intelligent system to support the practice of inpatient healthcare through a Web platform that allows the monitoring of patients admitted to a health facility. Thus, the developed system culminates in an application where all relevant information is gathered to monitor the different hospitalization episodes, presenting this information in a simplistic and intuitive way and alerting the professionals to the occurrence of events related to medical exams and analysis, surgical procedures, among others. This paper presents the architecture, the requirements and a SWOT analysis of the solution proposed, the main conclusions and a proposed future work.

Keywords: Inpatient support system, Multi-Agent System, Clinical Decision Support, System Healthcare

State of Publication: Published

DOI: 10.1007/978-3-030-40690-5_36

C.9 Mobile Computing in Patient Relationship Management – A Case Study

Authors: Pedro Moreira, Daniela Oliveira, Filipe Miranda, António Abelha and José Machado

Title: Mobile Computing in Patient Relationship Management – A Case Study

Conference Proceedings: 16th International Industrial Simulation Conference, EUROSIS.

Year of Publication: 2018

Abstract: Around fourteen years ago, the first hospital in Portugal started to use a patient notification system, to alert patients of their medical events at the hospital, such as appointments, surgeries, exams, treatments through text messages. This notification system is used nowadays, but it faces a big problem: a huge amount of money spent for the telecommunication companies involved. Although each message cost a fraction of cents, it can easily reproduce its value in more than 50,000 euros per year per institution. Since technology and the use of smartphones has been evolving in such a quick way, it is estimated that in no more than 10 years, almost all the Portuguese population will use smartphones or have access to them. For those reasons, the main purpose of the present work, is to design and develop a mobile application in order to substitute the previous notification systems, through push up notifications on the app and by email, that can be saved on the smartphone calendar, translating in no costs associated with the notifications sent by the hospitals. The main motivation is, therefore, suppressing these costs for the hospital, bring the patients closer integrating other systems on the app and make the notification alert more efficient. Thereby, the mobile app will be able, not only to manage each notification and notify the patients, but also to check its medical event history and to schedule medical appointments.

Keywords: Information Technologies, PRM, push notifications, mobile application, e-Health

State of Publication: Published

C.10 Data Mining for Prediction of Length of Stay of Cardiovascular Accident Inpatients

Authors: Cristiana Silva, Daniela Oliveira, Hugo Peixoto, José Machado and António Abelha

Title: Data Mining for Prediction of Length of Stay of Cardiovascular Accident Inpatients

Conference Proceedings: Digital Transformation and Global Society, Vol. 858, pp. 516-527, Springer.

Year of Publication: 2018

Abstract: The healthcare sector generates large amounts of data on a daily basis. This data holds valuable knowledge that, beyond supporting a wide range of medical and healthcare functions such as clinical decision support, can be used for improving profits and cutting down on wasted overhead. The evaluation and analysis of stored clinical data may lead to the discovery of trends and patterns that can significantly enhance overall understanding of disease progression and clinical management. Data mining techniques aim precisely at the extraction of useful knowledge from raw data. This work describes an implementation of a data mining project approach to predict the hospitalization period of cardiovascular accident patients. This provides an effective tool for the hospital cost containment and management efficiency. The data used for this project contains information about patients hospitalized in Cardiovascular Accident's unit in 2016 for having suffered a stroke. The Weka software was used as the machine learning toolkit.

Keywords: Data mining, Weka, Prediction, Cardiovascular accident

State of Publication: Published

DOI: 10.1007/978-3-030-02843-5_43

C.11 Step Towards Interoperability in Nursing Practice

Authors: Cristiana Silva, Daniela Oliveira, Hugo Peixoto, José Machado and António Abelha

Title: Step Towards Interoperability in Nursing Practice

Journal: International Journal of Public Health Management and Ethics (IJPHME), Vol. 3, IGI Global.

Year of Publication: 2018

Abstract: Hospital inpatient care compromises one of the most demanding services in health institutions for providing a careful and continuous healthcare assistance. Such demands require a constant update of the patients' health record allied with support systems responsible for monitoring their clinical information. In this context, the problem in this study becomes a process of continuous improvement. To define the case study, it was necessary to use research tools such as questionnaires and interviews. With these techniques, it was possible to delineate the state and dimension of the problem. Subsequently, the approach and solution was established and a new web platform for the daily monitoring of patients was proposed focused on nurses. The tool incorporates a real-time data visualization, and a patient record during an inpatient care episode. Moreover, this article also highlights the required adaptability of this platform for each health unit according to needs. With this solution, it is expected to correct many of the problems detected through quantitative results.

Keywords: Healthcare, Hospital Information System, Interoperability, Multi-Agent System, Nursing Practice

State of Publication: Published

DOI: 10.4018/IJPHME.2018010103

C.12 New Approach to an openEHR Introduction in a Portuguese Healthcare Facility

Authors: Daniela Oliveira, Ana Coimbra, Filipe Miranda, Nuno Abreu, Pedro Leuschner, José Machado and António Abelha

Title: New Approach to an openEHR Introduction in a Portuguese Healthcare Facility

Conference Proceedings: World Conference on Information Systems and Technologies. in Advances in Intelligent Systems and Computing, Vol. 747, pp. 205-211, Springer

Year of Publication: 2018

Abstract: Implementing a new EHR data system is not easy, as the systems already in place and user mentality are very difficult to change. The openEHR architecture introduces a new way of organizing clinical information using archetypes and templates. The present paper focuses on the initial steps of the implementation of an openEHR based EHR in a Portuguese major HealthCare provider. The system comprises operational templates creation through the creation of a validation mechanism and after that storage, a platform for data generation dynamically constructed from templates and an interoperability mechanism through the implementation of an HL7 V3/CDA message system.

Keywords: EHR, openEHR, Archetypes, Operational templates, HL7 V3/CDA, SNOMED CT, Interoperability

State of Publication: Published

DOI: 10.1007/978-3-319-77700-9_21

C.13 Improving Nursing Practice through Interoperability and Intelligence

Authors: Daniela Oliveira, Júlio Duarte, António Abelha and José Machado

Title: Improving Nursing Practice through Interoperability and Intelligence

Conference Proceedings: 2017 5th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW), pp. 194-199, IEEE

Year of Publication: 2017

Abstract: Hospital inpatient care compromises one of the most demanding services in health institutions when providing a careful and continuous healthcare assistance. Such demands require constant update of the patients' electronic health record allied with support systems responsible for monitoring their clinical information. In this context, this paper presents a new web platform for daily monitoring of patients, designed to be used by health professionals, especially nurses. The application is based on React, an open-source JavaScript library for building user interfaces. The developed tool incorporates two main features: the real-time visualization of the data, and the storage of the patient's historic during an inpatient care episode. The storage capability allows keeping the data updated among hospital shifts. Moreover, this work also highlights the required adaptability of this platform for each health units inside a hospital center according with its needs.

Keywords: Interoperability, Hospital Information System, Healthcare, Nursing Practice

State of Publication: Published

DOI: 10.1109/FiCloudW.2017.92