Universidade do Minho
Escola de Engenharia

Ana Cecília Sousa da Rocha Coimbra

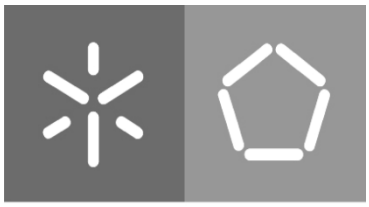
Improving Clinical Problem List with Evidence Based Medicine, Patient Oriented Medical Records and Intelligence

April 2021

Improving Clinical Problem List with Evidence Based Medicine, Patient Oriented Medical Record and Intelligence

Ana Coimbra

UMinho | 2021



Universidade do Minho
Escola de Engenharia

Ana Cecília Sousa da Rocha Coimbra

Improving Clinical Problem List with Evidence Based Medicine, Patient Oriented Medical Records and Intelligence

PhD Thesis

Doctoral Program in Biomedical Engineering

Supervised by

Professor José Manuel Ferreira Machado and

Professor António Carlos da Silva Abelha

April 2021

ACKNOWLEDGEMENTS

I would like to thank all the people who, in different ways, supported me during this journey.

First of all, I would like to thank my two supervisors, Professor Doutor José Machado and Professor Doutor António Abelha, for all the availability, patient, support and attention, without their time and advices this work would not be possible.

To the professionals of the Information Systems Service of Hospital de Santo António, for the reception and support shown. Special thanks to Dr. Isabel Santos who guided me on the "e-consentimento" project, the opinions, criticisms, the motivation and help were very important for the development of the project and to the entire Centro Hospitalar Universitario do Porto coding department for all their patience and availability.

To my friends, Adriana, Tiago, Inês and Flávio for all the shared moments, I thank you for always presenting me with words of motivation, patience, laughter and emotional support. Sometimes we need to relieve our head to get back to work more strongly.

To my colleague Filipe Miranda for all the good advice and ideas that contributed to the work presented, for all the help and exchange of opinions, no doubt that this project has become richer with your help.

To my family who always supported me and believed in me, thank you very much.

Finally, a very special thanks to my husband João, who was always by my side, and with whom I shared my joys and frustrations, thank you for your patience, strength and support along this route and for never letting me give up. Without your support this journey would have been much more difficult.

DECLARAÇÃO DE INTEGRIDADE

Declaro ter atuado com integridade na elaboração do presente trabalho académico e confirmo que não recorri à prática de plágio nem a qualquer forma de utilização indevida ou falsificação de informações ou resultados em nenhuma das etapas conducente à sua elaboração.
Mais declaro que conheço e que respeitei o Código de Conduta Ética da Universidade do Minho.

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 University of Minho.

RESUMO

As listas de problemas clínicos são muito importantes na prestação de cuidados de saúde, principalmente em termos de precisão. A presente tese tem como base um conjunto de estudos realizados no Centro Hospitalar Universitário do Porto, onde o principal objetivo é o de melhorar as listas de problemas clínicos.

O primeiro estudo foca-se nos passos iniciais do desenvolvimento de um sistema de registo clínico inovador que utiliza openEHR e terminologia SNOMED CT. Este sistema irá permitir a criação de registos estruturados através da utilização de arquétipos, terá também definidos protocolos baseados nas guidelines HL7 versão 3.

O segundo e terceiro estudo centram-se na codificação dos relatórios de alta. A codificação dos relatórios de alta permite um melhor agrupamento de episódios nos Grupo de Diagnóstico Homogéneos, daí a importância de tornar este processo o mais eficiente possível e com o mínimo de erros. Deste modo foi desenvolvida uma plataforma para que os médicos possam facilmente codificar os referidos episódios, tendo em *background* processos de gestão para auxiliar o *workflow* de todo o processo de codificação.

O quarto e último estudo refere-se ao desenvolvimento de uma plataforma capaz de disponibilizar consentimentos informados personalizados, onde os médicos podem adaptar os consentimentos aos diferentes tipos de casos que encontram.

A metodologia adotada é a Design Science Research (DSR) suportada por uma filosofia pragmática. Ao longo do desenvolvimento do projeto um conjunto de grupos de foco irão contribuir para a continua avaliação do sistema.

Palavras-Chave: Arquétipos; Consentimento Informado; ICD-10-CM/PCS; Lista de Problemas Clínicos; Machine Learning

ABSTRACT

The clinical problems list is very important in the provision of health care, mainly in terms of accuracy. This thesis is based on a set of studies carried out at the Centro Hospitalar Universitário do Porto where the main objective is improving the lists of clinical problems.

The first study focuses on the initial steps of developing an innovative clinical record system that uses openEHR and SNOMED CT terminology. This system will allow the creation of structured records through the use of archetypes, it will also have defined protocols based on the guidelines HL7 version 3.

The second and third studies focus on the codification of discharge reports. The codification of discharge reports allows for a better grouping of episodes in the Homogeneous Diagnostic Groups, hence the importance of making this process as efficient as possible and with the minimum of errors. In this way, a platform was developed so that doctors can easily code these episodes, with management processes in the background to assist the workflow of the entire coding process.

The fourth and final study refers to the development of a platform capable of providing personalized informed consent where doctors can adapt the consent to the different types of cases they encounter.

The methodology adopted is Design Science Research (DSR) supported by a pragmatic philosophy. Throughout the development of the project, a set of ´focus groups will contribute to the continuous evaluation of the system.

Keywords: Archetypes; Clinical Problem List; ICD-10-CM/PCS; Informed consent; Machine Learning

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

A

Administração Central de Sistema de Saúde
ACSS 29, 60, 64
Archetype Definition Language
ADL 34, 37

C

Centro Hospitalar Universitário do Porto
CHUP 17, 23, 58, 71, 87
Clinical Decision Support Systems
CDSS 23

D

Data Mining
DM 30
Department Information System
DIS 24, 46
Design Science Research
DSR 47, 48, 49, 60, 61, 69, 70
Diagnosis-Related Groups
DRG 17, 28, 29, 42, 43, 44, 59, 86, 87
Digital Imaging and Communications in Medicine
DICOM 23

E

Electronic Clinical Process
PCE 23, 24, 25, 42, 46, 47, 88
Electronic Health Records
EHR 15, 16, 18, 20, 23, 25, 26, 30, 31, 32, 33, 34, 37, 47, 85, 94
Extensible Markup Language
XML 23, 26, 35, 36, 37

G

Graphical User Interface
GUI 25, 26, 33, 36

H

Health Information Systems
HIS 21, 22, 23, 38
Health Level 7
HL7 ..... 22, 31, 32, 33, 34, 36, 37, 38, 86
Hospital Patient Management System
SONHO 23, 46

I

J

L

M

N

O

P

R

S

U

1. INTRODUCTION

The present dissertation focuses on the presentation of five case studies that aim to improve the lists of clinical problems.

The project is part of the doctoral dissertation of the PhD in Biomedical Engineering at the University of Minho.

This chapter presents a contextualization and framing of the theme, as well as the motivation for its elaboration. Some questions are also raised, and objectives set out to answer these questions.

Finally, the structure of the dissertation is presented.

1.1. Contextualization and framing

A clinical problem list is a list of current and past problems relevant to the current care of the patient. The use of problem lists is quite advantageous not only in terms of providing direct care, helping physicians to more easily identify the most important factors for each patient, but also in terms of studies. It is easier to identify specific populations of diseases, to improve care delivery through case studies, thus perceiving the populations of each health center and also to facilitate the identification of patients for potential research studies [1], [2].

However, maintaining the problem list current and accurate it is not an easy task since the most Electronic Health Records (EHR) systems use problem lists populated by physicians, which implies a routinely updating by them, which unfortunately becomes too laborious leading to inaccurate, incomplete and duplicative lists. The use of machine learning techniques leads to reduce these inaccurate, incomplete and duplicative lists [2], [3].

The use of terminologies associated with the list of clinical problems is a solution to reduce the existing redundancy, and at the same time to facilitate information retrieval [2]. An example of a terminology is SNOMED CT. SNOMED CT contains around 350 000 active concepts, and is open-ended, never being complete, since in the medical world it is in constant discovery what lead to constant updates in the terminology, thus becoming an appropriate candidate to integrate with EHR [4], [5], [6].

1.2. Motivation

As mentioned in the previous section the medicine field is continually in evolution which leads to the need of continuous updating of the resources used to support clinical practice, However the conventional EHR systems are restricted and rigid leading to a time-consuming and costly process when it comes to updating existing systems. The use of OpenEHR archetypes can be a solution to solve this problem.

On the other hand, the quality of health care provided can be impaired if there is no correct hospital financing. To avoid this, it is necessary to quantify hospital products, grouping hospitalization episodes into groups of similar resource expenditures. This grouping is facilitated through the codification of hospitalization reports by the coding physicians.

However, this coding does not have a defined structure, which leads to the possibility of errors occurring, such as not filling in all the necessary parameters or typing errors.

In terms of carrying out procedures, informed consent is extremely important to ensure a good doctor-patient relationship and for the patient to really understand what they are proposing.

1.3. Objectives and research questions

Regarding the first study, the following questions are asked:

- **Question 1:** How is it possible to implement an openEHR based Electronic Health Record (EHR) in a Portuguese major healthcare unit?
- **Question 2:** What is the importance of implementing this system?

To answer these questions, the following objectives were outlined:

- Planning and definition of requirements for the development of a new openEHR based EHR system:
    - Define the information workflow;
    - Define the requirements for data generation;
    - Define how exchange of information will work.
- Performing a SWOT analysis.

Regarding the second and third studies the following questions are asked:

- **Question 3:** What does ICD-10-CM/PCS represent in health institutions?
- **Question 4:** What are the advantages of ICD-10-CM/PCs over ICD-9-CM?

- **Question 5:** What is the role of ICD-10-CM / PCS in Diagnosis Related Groups (DRG)?
- **Question 6:** What is the importance of DRG in health institutions?
- **Question 7:** What gaps are found in the codification process of the inpatient report?
- **Question 8:** How can you guarantee compliance with the data protection regime?
- **Question 9:** How can the platform interoperability with SIMH be guaranteed?

The following objectives have been outlined to answer these questions:

- Change from ICD-9-CM to ICD-10-CM/PCS on the existing platform:
    - Change the codification area for only two tables: diagnosis and procedures;
    - Add dynamic help in the code insertion.
    - Development of a platform to manage the episode coding process:
    - Definition of episode status workflow;
    - Creation of types of users with different types of access to patient information.
- Performing a SWOT analysis.

Regarding the fourth study the following questions are asked:

- **Question 10:** How are informed consent available at Centro Hospitalar Universitário do Porto (CHUP)?
- **Question 11:** What is the importance of personalized informed consent in health institutions?
- **Question 12:** Does "*e-consentimento*" have the capacity to arrest the user through an intuitive interface and easy learning to adapt it?
- **Question 13:** Is the immediate implementation of "*e-consentimento*" viable?

To answer this question the following objectives have been outlined:

- Development of a platform for providing informed consent:
    - Structuring the interface in order to facilitate its use;
    - Provision of informed consent divided by procedures divided by specialties;
    - Restriction of access to informed consent belonging to a specialty other than that of the doctor himself.
- Performing a SWOT analysis.

The following questions can be raised in relation to the fourth study:

- **Question 14:** Will using the entire ICD-10-CM code set or just the first three digits be more efficient?

- **Question 15:** Is it relevant how many diagnoses are used as parameters in the data mining process?
- **Question 16:** It is possible to apply data mining algorithms using just the *"e-codificação"* platform's parameters?

The following objectives were established to address these questions:

- Structuring the dataset:
    - Create three distinct diagnostic datasets, two with three major diagnoses each and one with five major diagnoses each.
- Use the WEKA tool to create predicting models using four different data mining algorithms and compare performance between them.

Finally, it is possible to raise the main research question:

Question 17: How the use of evidence-based medicine, patient oriented medical records and intelligence, can improve the clinical problem lists?

1.4. Thesis Organization

This document is divided into five chapters. The first one being the present chapter where the introduction is presented, a brief contextualization and a framework, the motivation and the objectives.

The remaining chapters:

Chapter 2: the main concepts for carrying out this project are described, such as hospital information systems, interoperability, and the theoretical introduction to proofs of concepts carried out throughout this dissertation.

Chapter 3: the papers resulting of the five case of studies are presented. the first case study concerns the initial steps for the implementation of an openEHR based EHR in a Portuguese hospital, where the architecture of the platform to be developed is described. The second and third case studies concern the "*e-codificação*" platform, a platform used for the codification of discharge reports in ICD-10-CM / PCS terminology. These studies aim to explain the process of upgrading the platform from ICD-9-CM to ICD-10-CM / PCS, as well as the management system of the coding process and how it is interoperable with the SIMH platform. The fourth case study presents a platform developed in order to personalize the informed consent used by doctors, describing which tools are used and how it works.

Chapter 4: A list of papers publish or accepted for publication.

Chapter 5: The main conclusions and answers to the questions raised at the beginning of this dissertation, and is also highlighted is some future work that can be done

2. STATE-OF-ART

In this chapter it is presented a review of the literature and the state of the art of the concepts associated with the development of this thesis proposal.

2.1 Clinical Problem List

As already mentioned in the previous section, clinical problem lists are lists with all problems relevant to the patient, whether past or current. The use of these lists it is an asset and quite used by the physicians, hence their importance.

Matney et al. [7] have as main goal "To create an interoperable set of nursing diagnoses for use in the patient problem list in the EHR to support interoperability", for this all nursing diagnoses across four nursing terminologies were retrieved from queries executed against the Unified Medical Language System (UMLS) Metathesaurus. From these queries were retrieved 1320 concepts, but after a data set cleaning the subset has as result 369 SNOMED CT concepts in the nursing problem list. The authors highlight that "The problem list is a key component of the patient care and has been acknowledged as critical by the EHR Meaningful use criteria".

Galanter, Hier, Jao, & Sarne [8] presented a system based on alerts on inpatient medication. In cases of patients with one of the predefined diagnoses the validity of the alert was 96 ±1% and some alerts leads to addition of new problems to the list, in that case, the accuracy of problem list addition was 95 ± 1%. This leads the authors to conclude that the clinical decision support system integrated into the process of medication order is an asset, in terms of accuracy, to the addition of problems. They also defend that the patient safety and quality of care is influenced by the accurate problem lists.

Hartung, Hunt, Siemienczuk, Miller, & Touchette [9] has as research question "What is the impact of problem list documentation of heart failure on the likelihood that evidence-based pharmacotherapy that has been prescribed?". This cross-sectional study has as sample active patients with left ventricular ejection fraction of 40% or less. In conclusion the likelihood of being prescribed medication with known clinical benefit increases when the patient has a systolic dysfunction on the problem list.

All these studies highlight how important is to have an accurate and updated clinical problem lists and at the same time prove that this is feasible and advantageous in relation to what is normally used.

2.1 Health Information Systems

Health Information Systems (HIS) emerged around the 60s, 70s, and were called Departmental Information Systems, as they consisted of functionally limited applications, existing in certain departments of the hospitals, such as laboratory, radiology, among others. Later, they were extended to the entire hospital, thus having a broader view and therefore adopting the name HIS. However, despite this update, they were directed only to support doctors, and only later that support was extended to all health professionals [10].

In order for a decision on diagnosis, therapy or other procedures to be made with quality, adequate access to relevant data is necessary. When this access does not occur, it can lead to adverse events. Adverse events are understood not as those that result from the disease that led the patient to seek medical help, but rather as those that derive from medical conduct, with insufficient communication and lack of information being some of the factors that most contribute in this regard [10], [11].

Therefore, it is clear that using HIS as a support tool will help reduce the occurrence of adverse events, reduce clinical errors, support healthcare practitioners, and improve the efficiency and quality of healthcare. These positive effects are demonstrated in several studies. This decrease results from the timely availability of selected and targeted information, since HIS should not automate medical decision-making but rather reduce the cognitive burden of health professionals and improve the basis for their decisions [11], [12].

It must be taken into account that within a hospital, there are two main processes: the organizational process and the medical treatment process, the first being related to the organization of records such as results reports, patient admission, among others and the second related to the diagnoses and therapeutic procedures to be performed on the patient. Thus, HIS must assist in both types of processes, the first providing support in the coordination of health professionals and organizational units, and the second increasing the quality of medical care [11].

However, associated with HIS, there are dangers, such as the cost of modern Information Systems (IS) and the possibility of their failure, the latter of which can lead to negative effects on patients [12].

Health care is evolving towards isolated treatment of patients for continuous treatment where multiple professionals and institutions are involved. Thus, there is a need to create a wide network capable of connecting multiple hospitals, as well as insurance companies and government organizations. However, this will raise a number of security and confidentiality issues. On the other hand, this type of systems will result in an increase in opportunities both in the field of Medical Informatics and in the field of statistics, enabling a greater variety of studies [10], [11].

2.2 Interoperability

The concept of interoperability arises from the need for continuous communication and exchange of information between systems, thus always making information accessible. However, most information systems do not respect this concept, which leads to an accumulation of systems of different languages, thus making it difficult for physicians to consult different systems to access the different types of information intended [11], [13], [14].

Therefore, it is extremely important to guarantee the integration and interoperation between the different HIS, both of which are fundamental to guarantee the flow of information in health institutions. In this context, the concept of interoperability arises, which consists of the possibility of continuous communication and exchange of information, thus making information always accessible, facilitating the work of health professionals [15].

In order to improve and facilitate interoperability, the information transferred between HIS is normalized, thus avoiding misunderstandings and different message structures. The standards used can be divided into different categories depending on their purpose: communication standards, standards for the representation of clinical information and standards for imaging [15].

The most frequently used communication standard is Health Level 7 (HL7), which defines a series of guidelines for communication between HIS, thus making messages consistent and uniform, facilitating their processing and automatic analysis [11], [15].

On the other hand, as an example of a standard for clinical representation, the Systematized Nomenclature of Medicine Clinical Terminology (SNOMED-CT) appears, because it represents this

information in a consistent and efficient way, which allows its automatic processing and analysis [15], [16].

Regarding the medical imaging standards, the most widely used is Digital Imaging and Communications in Medicine (DICOM), which makes all types of medical image, of any modality, and of any equipment for exams, be stored in the same format, which allows them to be viewed on any device [15].

The reformulation of HIS in order to make them more homogeneous is not possible due to technical and financial restrictions. In this context, the Medical Information Integration, Dissemination and Archive Agent (AIDA), developed by a research group at the University of Minho, aims to send, receive, record and maintain the information contained in heterogeneous systems in the time feasible, thus enabling the control of the information flow with a certain level of autonomy. To ensure the platform's interoperability, AIDA's proactive agents, through messages in Extensible Markup Language (XML), guarantee the communication bridge between information sources existing in health institutions [13]–[15].

The security and confidentiality of information is also ensured by AIDA, and respects certain ethical and legal standards. In general, AIDA provides access to information present in all health institution information systems to authorized authorities, guaranteeing its security and availability [15].

Cardoso et al. [15] developed a module to add to AIDA that aims to monitor the intelligent agents and to create new ones.

Marcos, Maldonado, Martínez-Salvador, Boscá, & Robles [17] developed a system that uses archetypes to deal with the interoperability of Clinical Decision Support Systems (CDSS)s and EHRs.

2.3 CHUP Information System

At Centro Hospitalar Universitário do Porto (CHUP), AIDA is responsible for ensuring communication between the Information Registration Systems Hospital Patient Management System (Sistema de Gestão de Doentes Hospitalares - SONHO), Medical Support System (Sistema de Apoio Médico - SAM), Support System for Nursing Practice (Sistema de Apoio à Prática de Enfermagem - SAPE) and Electronic Clinical Process (Processo Clínico Eletronico - PCE) with

complementary systems, such as the Laboratory Information System (LIS), Department Information System (DIS), Radiology Information System (RIS) and also with web services , as can be seen in Figure 1.

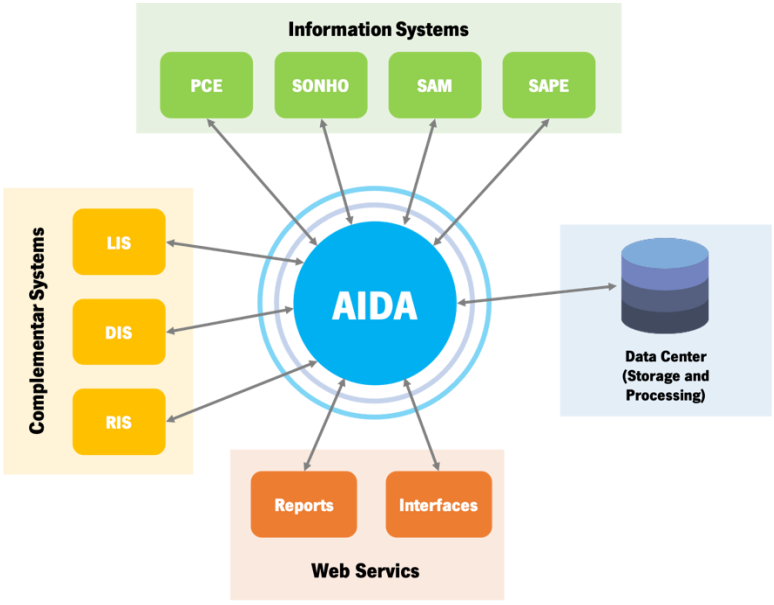


Figure 1 - AIDA modules (adapted from [13]*)*

The main objective of the PCE is to enable the registration of all data related to the patient, from personal data to pathology and diagnosis, in a safe, consistent, efficient, clear and structured way. In this way, the patient's clinical situation will be in constant record, thus allowing continuity of service provision [18], [19].

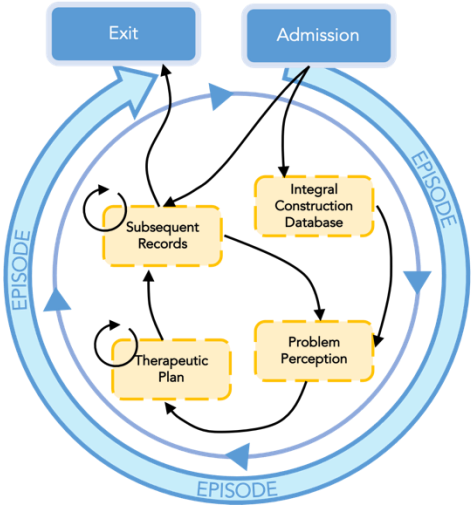


Figure 2 - Clinical information registration process carried out by AIDA-PCE (adapted from[21]*)*

The AIDA-PCE was implemented in the CHUP, and follows a method oriented to the problem known as Problem Oriented Medical Record (POMR), where clinical information is recorded for the

resolution of specific problems. Each record made is recorded containing the patient's symptoms, doctor's observations, diagnosis and the treatment plan the patient is subject to [20], [21].

In Figure 2, it is possible to understand how the AIDA-PCE records the clinical information of an episode, an episode being the collection of all operations related to the patient from his admission to the health institution until his departure [14], [21].

2.3 OpenEHR

The openEHR was developed to promote interoperability between EHR and presents a two-level approach: (1) the reference model and (2) the archetype model. The reference model defines the logical structures of EHR and demographic data. The archetypes are the definition of a distinct domain-level concept in the form of structured and constrained combinations of the classes of the reference model. The great advantage of using openEHR in EHR, besides guaranteeing interoperability, is that it allows the health professional to model the archetypes without any technological knowledge of the final EHR systems, thus creating more customizable systems and tailor-made for professionals. The use of external health terminologies, such as SNOMED CT, is also allowed in the openEHR [4]–[6], [22].

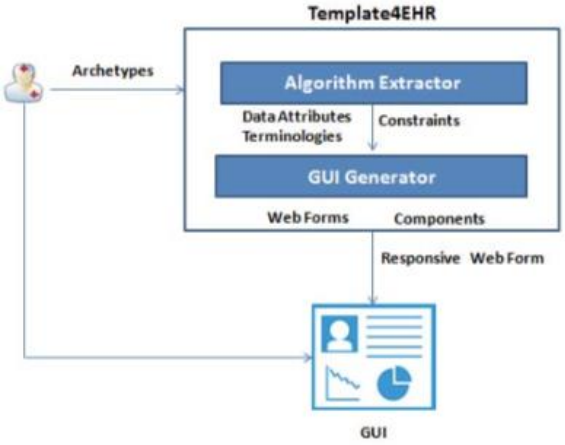


Figure 3 - Template4EHR architecture

Araujo et al. [23] developed a customizable Graphical User Interface (GUI) building tool for EHR using archetypes, named *Template4EHR*. The authors specified two algorithms to achieve the intended objective. The first algorithm, the "Algorithm Extractor", extracts from archetypes data attributes, terminologies and constraints. The second algorithm, the "GUI Generator", generates

the GUI using as input the data resulted from the first algorithm. In Figure 3 it is illustrated the *Template4EHR* architecture.

The authors also do a proof of concept, a case study to evaluate the generated GUIs. They compared the GUIs generated by the *Template4EHR* with the GUIs generated by three more systems: *Archetype Editor, Template Design* and *EhRScape Framework*. The conclusion was an 81.22% satisfaction of the professionals that took part of the study. The weak point to highlight in this study is that they do not explain how the data entered in the GUI are saved.

Schuler, Garde, Heard, & Beale [22] developed a different approach to the generation of GUI based on archetypes. The authors use the Mozilla XML User Interface Language (XUL). The authors obtained good results although some problems were encountered in using XUL, they defend that XML-based GUI languages have advantages. Therefore, the research question made "Is it possible the creation of GUI from OpenEHR archetype?" was answered positively.

Kazmi [24] made a literature review to focus on the research question "The use of EHR in medical consultations affects doctor-patient communication?". This is an important topic to be addressed, however the effect on doctor-patient communication of EHRs in the exam room has not been sufficiently explored. After a filter the author analyzed 13 articles. As results the author obtained that the use of EHR in the clinical environment has weak points in terms of behaviors, the reduction of eye contact, communication, and provision of emotional support that result in keyboarding and screen gaze by the physician. In the author opinion it should be realized more studies in this area, and he highlights the need of creation of EHR systems that are not only usable but also that the principal consideration should be patient experience.

Almeida, Frade, & Cruz-Correia [25] developed a tool that export data to standard table formats from any openEHR repository. They have to deal with some difficulties namely in a web service protocols level. This article is very important to the present project because illustrates a way to extract the data from the developed platform to be used after in the machine learning process.

2.4 Terminologies

2.4.1 SNOMED CT

The Systematized Nomenclature of Medicine – Clinical Terms (SNOMED CT) terminology is composed by concepts and by relationships that links the concepts that have related meaning. A feature of SNOMED CT is the presence of synonyms, where a concept may be associated with

different descriptions that represent the same clinical idea. SNOMED CT is becoming a frequently used terminology due to its complexity of terms and association between them [4], [5], [26].

Wasserman & Wang [27] try to answer the question "SNOMED CT is a relatively complete standardized terminology on which to base a vocabulary for the clinical problem list?". The authors implemented the SNOMED CT as a base vocabulary, however if the user could not find the desired term, he can submit the new term. For every "absent" term it was made a rigorous search with the SNOMED CT base vocabulary, if the term is really "missing" it was added to the vocabulary and assigned one of four mapping types (that represents the relationship between the SNOMED CT and the new entry). The case study presented good results, 88.4% of the terms were found in SNOMED CT, being that only 20 represented significant concepts were truly missing on the SNOMED CT from the 145 "missing" terms. Concluding affirmatively to the research investigation.

Meizoso García, Iglesias Allones, Martínez Hernández, & Taboada Iglesias [28] presents the research question "Is it possible to develop an approach that automatically bind the OpenEHR archetypes term to external terminology SNOMED CT, in a semantic level?". The authors find similarities in terms through the use of information about structural and semantic proximity. As the results 74.6% of the terms were linked with 96.1% precision. With this case study the authors claim that the human participation in mapping process can be reduced.

A good example of an SNOMED CT system implemented with success was the system developed by Duarte et al.[26]. The pathological anatomy department was chosen as pilot service, and, in the end, they concluded the feasibility of extend the system to the whole hospital aiming to provide alerts and prevention systems and reduce medical errors.

2.4.2 ICD-9-CM

The International Classification of Diseases 9th Revision Clinical Modification (ICD-9-CM) codification consists of a list of codes that correspond to diagnoses and procedures performed in a hospital. It was created by the Department of Health & Human Services of the United States of America (USA) and the Centers for Medicare and Medicaid Services as an adaptation of the International Classification of Diseases, Ninth Revision (ICD-9) drawn up by the World Health Organization (WHO) with the goal of tracking mortality statistics worldwide [29].

However, the classification and terminology used in some cases was considered obsolete, and the existence of inaccurate and limited data led to the need to update this type of codification, thus emerging the International Classification of Diseases 10th Revision Clinical Modification (ICD-10 - CM). Presenting the same hierarchical structure as the ICD-9-CM, but with the presentation of the complete title at all levels, with the requalification of some diseases and the restructuring of some chapters [30].

2.4.3 ICD-10-CM

International Classification of Diseases, 10th Revision, Clinical Modification (ICD-10-CM) is the modification made by the United States to the international classification of Diseases, 10th Revision (ICD-10) developed by WHO This came to replace the ICD-9-CM, since this version it was no longer able to accommodate major diseases and procedures [31], [32].

In the new version there are almost five times more codes to list diagnostics, however, despite the greater variety of codes available, it is a challenge to establish optimal definitions [33].

The ICD-10-CM is used for example to allocate resources in medical centers or to set capitation rates, therefore, it is very important to guarantee the accuracy of this classification [34].

2.5 Diagnosis-Related Groups

The Diagnosis-Related Groups (DRGs) consists of a system for classifying patients into relatively homogeneous groups in terms of clinical characteristics and resource consumption, essentially relating the type of patients treated with resource consumption. It was initially developed at Yale University, USA in the late 1960s, beginning of the 1970s, and has been used since 1983 by Medicare in the United States as a financing basis for hospitalized patients treble [35]–[37].

In 1984, the feasibility of implementing the DGRs in Portugal began to be studied. The first tests were carried out as a basis for financing hospitalization in 1989. Subsequently, in 1990, the Case-Mix Index (ICM) concept was used for the first time in National Health Service (*Serviço Nacional de Saúde* - SNS) hospitals as a basis for calculation of inpatient financing. In 2009, DGRs corresponded to 51% of total SNS hospital financing [35].

The DGRs requires the collection of a Minimum Data Set (MDS) so that it is possible to group an inpatient episode into one of the 25 Major Diagnostic Categories (MDC) and within these in one of the almost 669 DGRs available. Thus, the MDS includes, as follows [35], [36]:

- Main Diagnosis - responsible diagnosis for the admission of the patient to the hospital;
- Other diagnostics;
- Procedures - a hierarchy of procedures performed on the patient during the internment;
- Gender;
- Age - divided into pediatric DGRs when age is less than or equal to 17 years and adult DGRs when age is more than 17 years;
- Bird weight (newborn);
- Destination after discharge – patients can be transferred, discharged against medical order or deceased.

Each DGRs group is associated with a relative weight and a weighting coefficient. An exception threshold is also defined, that is, a normal range of hospital stay. Thus, taking into account the length of stay of each patient and the range of normality, it is possible to convert hospitalization episodes into equivalent patients. Thus, a financing basis can be defined and make it possible to calculate the production of hospital resources [36].

The Portugal National Database of DGRs, headquartered in the *Administração Central de Sistema de Saúde* (ACSS) is updated monthly with the information of the DGRs of all hospitals of the SNS [35].

Some countries such as Germany, France, the United Kingdom, Scandinavian countries, Australia or Canada have relied on the original 1983 United States DGR and created their own systems with their own episode grouping and algorithms. In the USA, the "All Patient Diagnosis Related Groups" (AP-DRG) grouper is in force, having been adopted in Portugal [35].

2.5 Machine Learning

The use of machine learning is growing exponentially in the health area and its correct implementation has brought several benefits for health professionals, as well as improving the efficiency and quality of medical care.

Devarakonda et al. [3] try to prove that the use of machine learning to generate a problem list is a better solution than the use of the problem lists generated by the physicians. To prove this research question, it was made a case study, where 15 random de-identified patient records were selected, and three problem lists were generated, the ones created by physicians (P), the ones

generated by Watson (W), the cognitive computer system developed by the authors, and the ones existing in the EHR (E). The physicians evaluated on a 10-point scale and the primary outcome was pairwise comparisons of P, W, and E. It was evaluated 732 Watson generated problems and 444 problems in EHR system. The P was a better score, however as they were the ones to evaluate their own lists this result can be influenced, despite this, in 89% of the assessments W identified at least one important problem that physicians missed. Furthermore, W was rated higher than E. Therefore, the authors concluded that with the use of systems like the Watson it is possible to improve the accuracy of the problem list. However, the authors should eliminate the influence factor in the case study by dividing the sample of physicians into groups, one aimed to the problem list creation and the other group to score that lists created. Or, instead of separating into groups, each physician has evaluated the list of another physician.

Coimbra et al. [38] aimed to improve the quality-of-service provided in a neonatology service and reduce the inherent financial costs through a diagnosis decision support system with a case-based problem solving methodology to computing. The research question was "*Is it possible to predict the length of stay of preterm infants in a neonatology service?*". Regarding the results, the model presented an accuracy of 84.9% in predicting the length of stay and the computational time was reduced around 21.3%.

Morais et al. [39] presented as goal "*Predict the need of neonatal resuscitation given some health conditions of both the newborn and the mother, and also the characteristics of the pregnancy and the delivery using Data Mining (DM) models induced with classification techniques*". The tool used to induce the DM was WEKA software and was followed the CRISP-DM methodology. The results were satisfactory once some models achieved sensitivity results higher than 90% and accuracy and specificity results higher than 98%.

3. RESULTS

3.1 Study I – New Approach to na openEHR Introduction in a Portuguese Healthcare Facility

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

WorldCIST'18 20182018, AISC 747, pp. 205–211, 2018
Springer Nature 2018

3.1.1 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.

3.1.2 Introduction

An Electronic Health Record (EHR) storages a large amount of medical data, data that must be available throughout the lifetime of a patient. Besides the effort and cost the solution must protect information when data loss occurs and at the same time be persistent and reliable across the years. The problem is often not the quantity of available data, instead the major issue is the fact that most of the information is made up to free text serving for nothing more than registering and consulting information. Since 2004, the openEHR foundation has published a series of design specifications for semantically interoperable and future-proof EHR systems. The main feature of the openEHR design is the separation between clinical concerns and technical design, the so called, two-level modelling [I-1]. The first level, Reference Model (RM), represents the technical concerns (information structure and data types). The second level of the model handles the clinical domains

(representation of communication of the semantics) [I-2]. This enables the construction of stable EHR systems without specific clinical content necessary in different fields [I-1].

The use of archetyping in openEHR enables new relationships between information and models. An archetype stands for a computable expression of a domain in the form of structured constraint statements, so openEHR archetypes are based on the openEHR Reference Model [I-2]. These can be composed into larger structures called the templates [I-2].

The purpose of the present document is to demonstrate the initial steps taken in the implementation process of an openEHR based EHR in a Portuguese major healthcare unit. The solution composes the creation of templates (modification of archetypes and translation), a system for validation and storage of the previous and the subsequent creation of web forms with basis on that operational templates. The system will also feature the generation and storage of information and exchange of information using HL7 Version 3 guidelines and HL7 V3 CDA. HL7 International specifies several flexible standards, guidelines, and methodologies by which various healthcare systems can communicate with each other. Such guidelines or data standards are a set of rules that allow information to be shared and processed in a uniform and consistent manner. These data standards are meant to allow healthcare organizations to easily share clinical information. Theoretically, this ability to exchange information should help to minimize the tendency for medical care to be geographically isolated and highly variable.

After the introduction, section two is entitled “Background” and is based on an intensive review of the literature on the theme, as well as, in the opinion of several authors, using them as motivations and strengths of the present work. Subsequently, the main section is presented as “System overview” and initially describes the proposed system in a general way, and then each component is described in detail. In the section “Discussion”, through the SWOT analysis, a primary evaluation of the described system was carried out and, in the last section “Conclusion and Future Work”, the main conclusions of the work accomplished, as well as, future steps are mentioned.

3.1.3 Background

Structuring large amount of information is not easy, as noted by several failed attempts reported in the past. According to Rector in “Clinical Terminology: Why is it so hard?” one of main reasons is that structuring requires standardization of the structured element that usually is implemented

in a top-down manner [I-3]. The openEHR architecture presents a new and interesting approach that empowers the local user to decide how they want to structure the EHR [I-3].

Although reported in various informatic research projects with considerable success, reports from real large scale implementations are still scarce [I-4]. This statement highlights the importance and difficulty of this kind of projects. Ellingsen et al. [I-4] reported the first efforts to implement large-scale EHR in Western hospitals (in Norway) conforming to the openEHR architecture [I-4], offering an insight into the first two years of the process, and the socio-technical challenges they met along the way. Wollersheim et al. [I-5] in "Archetype-based electronic health records: a literature review and evaluation of their applicability to health data interoperability and access" presents an overview of the current archetype literature relevant to Health Information Managers [I-5]. In this paper, different developments are presented, concerning different settings of healthcare: elder patient care, complementary and alternative medicine, nursing, discharge summary or even security concerns focused on the security of MEDIS. The same work also presents a distribution of works by country, with Australia assuming a leading role in the openEHR implementation.

One of these Australian works belongs to Murat Gok, entitled "Introducing an openEHR-Based Electronic Health Record System in a Hospital", provides a road map for future implementations of an openEHR system in the Austin Hospital Emergency Department, Melbourne, including history, architecture and relationships to other standards. These relations play a central role on the present work, R. Qamar and A. Rector presented in "Semantic Mapping of Clinical Data to Biomedical Terminologies to Facilitate Data Interoperability" interesting points of view on this subject, stating that interoperability of clinical systems requires integration of data models, such as HL7 messages or openEHR archetypes, with terminologies such as SNOMED-CT or ICD-X [I-6].

A key to the success of an EHR system is inevitably its usability. If the Graphical User Interface (GUI) of a system is not intuitive and appealing there will be greater resistance from healthcare professionals, leading to under-use or misuse of the system [I-7]. Template4EHR (http://template4ehr.azurewebsites.net/) and EhrScape Framework (https://www.ehrscape.com/) are examples of tools that dynamically generate GUI from archetypes for Health applications [I-8].

3.1.4 System Overview

The purposed system consists in the implementation of an openEHR based EHR in a Portuguese major healthcare unit and it can be divided into three major components, as shown in Figure 4. The first, the information workflow, focuses on the creation and edition of archetypes and templates. The second component, data generation, uses the operational templates previously created and generates dynamical web forms. The last part of the system, HL7 Binding, uses HL7 V3 guidelines and HL7 V3 CDA to generate, storage and exchange the information. An overview of each of these components is presented below.

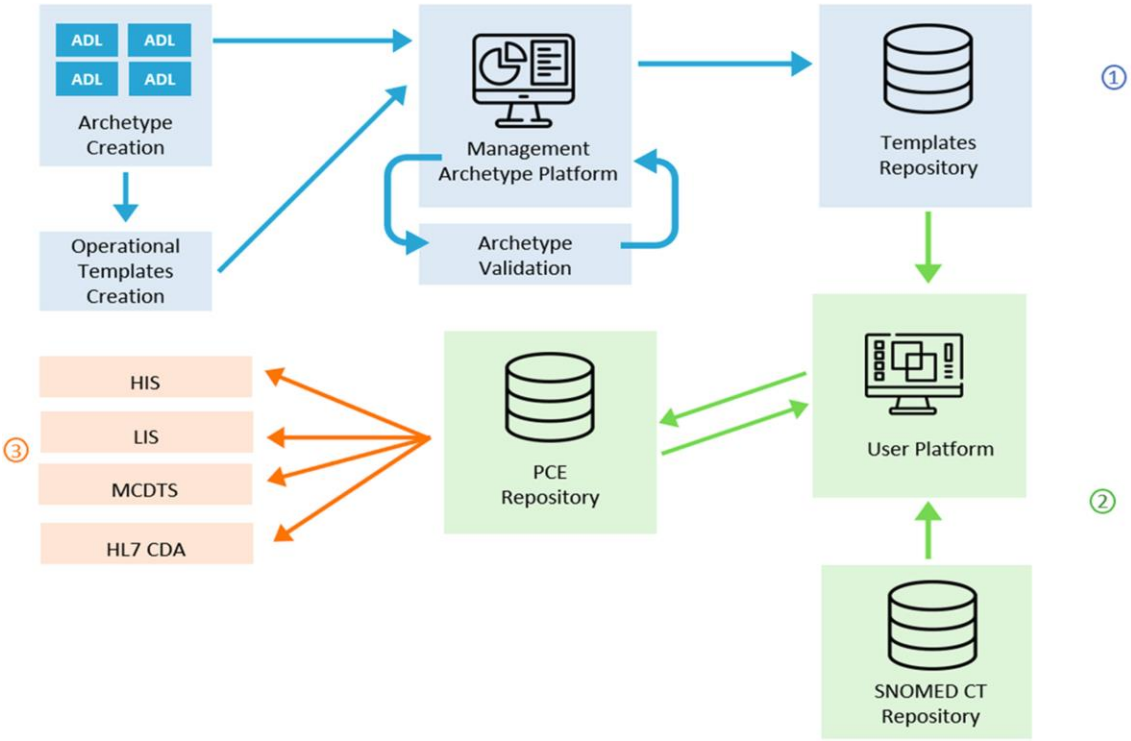


Figure 4 - The platform architecture.

3.1.4.1 Information Workflow

On the clinical domain, concepts can be organized through the archetypes, i.e., sets of independent data, which may be more or less specialized, and composed or decomposed for a new use. Blood pressure, glucose and diagnosis are some examples of the archetypes. These are written in Archetype Definition Language (ADL). The workflow of information is illustrated in Figure 5.

When new data is inserted in a specific case, such as clinical reports or specific messages, is the system requires the use of templates, previously defined and composed of various archetypes. In this case, Medical Observations template combine three independent archetypes, but the Diabetic Checkup use only two. In this new approach, Operational Templates (OPT) will be used because they have a flexible structure that facilitates the implementation processes, using extensible markup language (XML) or java object notation (JSON). With respect to this, the chosen strategy to save the OPTs in a specific database shall encounter consistency and reasoning. Furthermore, these OPTs can be compiled and transformed into specific artifacts, ready to be used by software developers, messaging systems implementers or data managers [I-9].

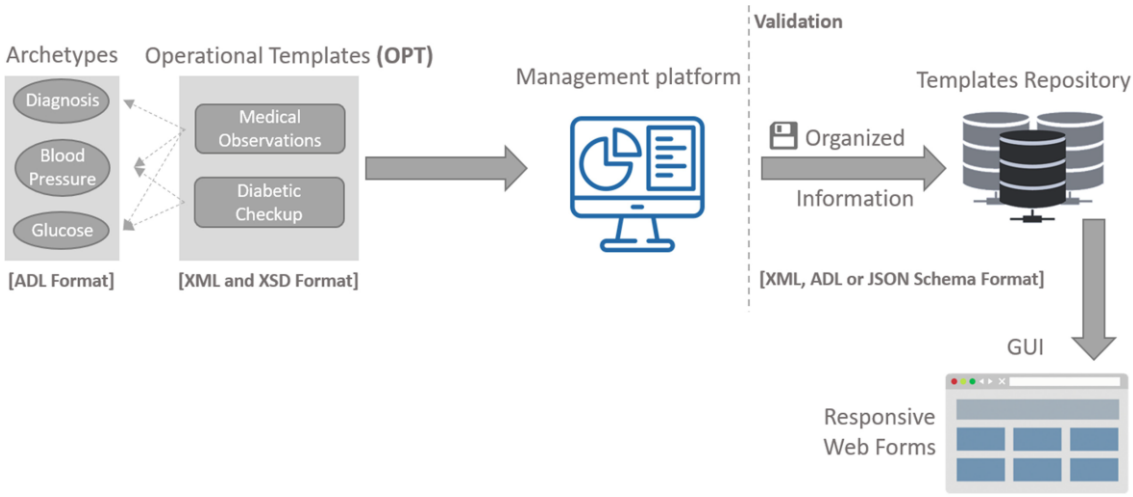


Figure 5 - Workflow of the information in the proposed system.

The proposed system will contain a multi-function support platform that allows one to submit new templates or consult existing ones. When a new template is submitted, it must be approved manually or automatically. In manual mode, a designated professional will accept or not the submitted templates, respecting the governance rules. After this validation, the template will be saved in the repository with a compatible format for the next processes, as aforementioned. In this step, the information conjugated from different HISs generates structured knowledge (Concepts and Definitions), ready to be use and, posteriorly, to create dynamic and responsive web forms. These web forms constitute the User Platform presented in Figure 4 .

3.1.4.2 Data Generation

The second part of the system consists in the dynamic generation of GUI from Operational Templates.

A web service will be responsible for reading the operational templates, extract the data and generate the GUIs, creating a webform platform.

It will use two repositories: one for storage of the OPTs (the label and definition of what is being clinically observed, and another for storage of the data inserted by healthcare professionals (the values, or results of the observation).

Semantic interoperability is ensured by binding the SNOMED CT terminology and the informational structures, represented by archetypes.

3.1.4.3 HL7 Binding

Health Level Seven version 3 (HL7 V3) defines standards for messages that are exchanged in the healthcare workflow. Relying on an object-oriented principle, where all messages are derived from the Reference Information Model (RIM) that together with data types and vocabularies are serialized in XML syntax defined by the Model Interchange Format (MIF). Version 3 of HL7 uses XML, in an evolution from v2, where "|" was used to separate the fields of a message. Although both openEHR and HL7 V3 rely on reference models, they are different, so they must be mapped to allow the conversion of information from one architecture to the other. An HL7 V3 message, as all messages, begins with a trigger event. Figure 6 represents the approach intended for the health unit.

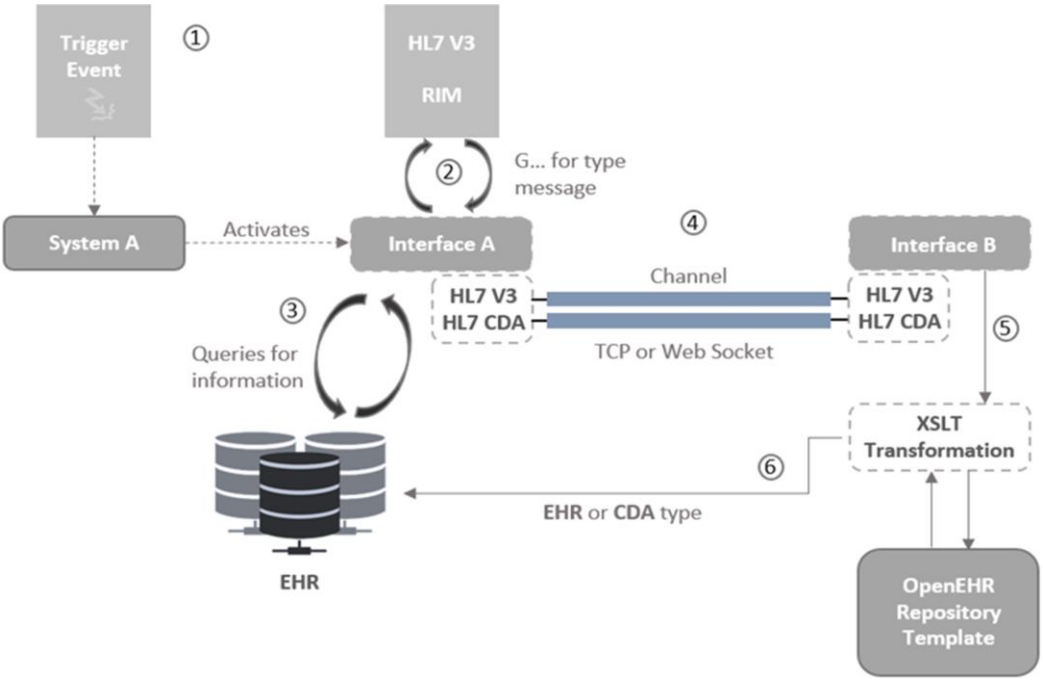


Figure 6 - Interoperability approach for openEHR based EHR

The trigger event (1) deploys a chain reaction that leads to the share of information or documents from one health system to another. The creation of an interface for TCP or Web sockets will enable the byte stream across the internet. After the trigger event, the interface must assemble the HL7 message v3 or CDA using the EHR in openEHR format, through query methodologies like ADL (2, 3). A XML-like document is created, converted to byte stream and sended to the interface B. This point will require a XSLT transformation in order to convert the XML associated with the HL7 v3 or CDA message to openEHR compatible content. This is based on the templates stored in a openEHR repository, designed specifically for that purpose (5). After that step, the information is ready for storage in openEHR-validated archetypes or templates.

3.1.4.4 System Communication: Multi-agent Systems

As the volume of clinical data is very high, it was necessary to build consistent and organized data models. In this case, each repository of the system will be updated through multi-agent systems (MAS) and, consequently, the front-end applications will be synchronized. Being multi-agent architectures a field of research of distributed artificial intelligence, this technology is intrinsically connected to the concepts that define a distributed architecture, while being distinct in the definition of an agent versus the properties of the general middle-wares of many others

distributed architectures. A MAS is a computer system with several autonomous agents that that interact with each other to perform certain tasks. Its main characteristics are the autonomous capacity to make decisions, as well as the capacity to interact with other agents through social interaction protocols, reaching the desired levels of coordination and cooperation [I-10].

Each health unity already aggregates information under the same roof, through the Agency for Integration, Diffusion and Archive of Medical Information (AIDA) system. AIDA is an agent-based platform with the purpose of ensuring the interoperability among HIS, i.e., is an example of a MAS [I-11]. On other hand, different health units communicate through HL7 v2 agents.

3.1.5 Discussion

In order to analyze the pros and cons of the proposed system a SWOT analysis was made, as described in this section (Table 1).

Table 1 - SWOT analysis for the proposed system.

Strengths	Weakness	Opportunities	Threats
Better structure of data and interoperability	Connection to the hospital intranet is required	Modernization and organizational development	User resistance of adopting a new system by healthcare professionals
User-friendly and intuitive interface	Manual creation of archetypes by health professionals	Economic benefit of using an open source solution (openEHR)	

3.1.6 Conclusion and Future Work

Verifying the SWOT analysis, it is possible to conclude that the proposed system is expected to grant significant added value for the Portuguese healthcare unit where it is to be implemented. For a better understanding of the benefits of this new system it is important to design and perform a study that compares this new system with the one previously implemented in the healthcare unit.

In the future could be added to the system a decision support tool with the aim to help the healthcare professionals for example, when a healthcare professional fills the diagnosis field in the form the symptoms field should be automatically filled in.

3.1.7 References

[I-1] Beale, T., Heard, S.: Archetype definitions and principles. In: Ocean Informatics, p. 15 (2007)

[I-2] openEHR Foundation: openEHR - Design Principles. In: Design Principles (2017)

[I-3] Rector, A.L.: Clinical terminology: why is it so hard? Methods Inf. Med. 38(4–5), 239–252 (1999)

[I-4] Ellingsen, G., Christensen, B., Silsand, L.: Developing large-scale electronic patient records conforming to the openEHR architecture. Procedia Technol. 16(2212), 1281–1286 (2014)

[I-5] Wollersheim, D., Sari, A., Rahayu, W.: Archetype-based electronic health records: a literature review and evaluation of their applicability to health data interoperability and access. Health Inf. Manag. J. 38(2), 7–17 (2009)

[I-6] Sciences, P., Qamar, R., Rector, A.: Semantic mapping of clinical model data to biomedical terminologies to facilitate. In: HealthCare Computing Conference 2007, p. 257 (2007)

[I-7] Schuler, T., Garde, S., Heard, S., Beale, T.: Towards automatically generating graphical user interfaces from openEHR archetypes. Stud. Health Technol. Inform. 124, 221–226 (2006)

[I-8] De Araujo, A.M.C., Times, V.C., Da Silva, M.U., Alves, D.S., De Santana, S.H.C.: Template4EHR: building dynamically GUIs for the electronic health records using archetypes. In: Proceedings of the 16th IEEE International Conferences on Computer and Information Technology (CIT 2016), pp. 26–33 (2017). 6th International Symposium on Cloud Service Computing (IEEE SC2 2016), 2016 International Symposium on Security Privacy in Social Networks

[I-9] Oliveira, R.: Transformation of clinical data from HL7 messages to openEHR compositions (2016)

[I-10] Oliveira, D., Duarte, J., Abelha, A., Machado, J.: Step towards interoperability in nursing practice. Int. J. Pub. Health Manag. Ethics (IJPHME) 3(1), 26–37 (2018)

[I-11] Peixoto, H., Santos, M., Abelha, A., Machado, J.: Intelligence in interoperability with AIDA. In: 20th International Symposium on Methodologies for Intelligent Systems, World Intelligence Congress, Macau. LNCS, vol. 766. Springer, Heidelberg (2012)

3.2. Study II - Improving the Codification of Hospital Discharges with an ICD-9-CM Single-page Application and its Transition to ICD-10-CM/PCS

Cecília Coimbra, Marisa Esteves, Filipe Miranda, Filipe Portela,
Manuel Filipe Santos, José Machado, António Abelha

International Journal on Advances in Life Sciences,
vol 10 no 1 & 2, year 2018

3.2.1 Abstract

In recent years, in Centro Hospitalar do Porto (CHP), there has been felt an increasing need for a computerized clinical coding tool to aid in the codification of the episodes of hospital discharges from patients admitted to its healthcare units. The process was slow and performed manually by the coding professionals, so there was neither the centralization nor the unification of the information and processes associated with the clinical coding of a hospital discharge. Hereupon, in the context of this study, the aim of the present work was to design and develop a clinical coding tool for ICD-9-CM to support the clinical practice in healthcare units. It additionally included its subsequent transition to the newer ICD-10-CM/PCS coding version. In short, the codification of hospital discharge processes enables the grouping of episodes into diagnosis-related groups (DRGs). The main motivation for the implementation of this classification system is that it provides a financial and patient classification system to contain the costs and waste associated with healthcare services. Thereby, a single-page application (SPA) for ICD-9-CM was designed in order to help health professionals of CHP in their daily work, namely the clinical coding of the episodes of hospital discharges, and it was subsequently updated to the ICD-10-CM/PCS coding version that predominantly improved specificity in describing clinical situations. The main advantages and contributions of the development and use of this Web application are the centralization of information and tasks associated with the coding of hospital discharges, the increase of productivity, and the reduction of wastes of time. Consequently, the ambition is sought to mainly improve the quantity and the quality of work performed by coding professionals at CHP

3.2.2 Introduction

The health sector represents a delicate situation for the professionals and the systems responsible for the storage and the processing of clinical data. The main problem with those processes is not in the lack of data but in the diversity and the complexity of the healthcare sector. Since a hospital offers a wide range of services for each patient and clinical condition, it leads to the creation of hundreds or even thousands of specific and unique situations for each one of them. It is then of utmost importance to find a way to measure a hospital productivity and aggregate the multiple activities performed at a healthcare unit.

This specific situation led to the adaptation of the DRGs to this scenario. These kind of classification systems rely on a prior coding system that translates all the diagnoses, procedures, external causes, and morphologies into universal codes like ICD-9-CM and ICD-10-CM/PCS, increasing the semantic interoperability and highly reducing the ambiguity of a discharge report. DRGs are clinically coherent and similar groups that are expected to use the same level of hospital resources.

The introduction of new Web frameworks and solutions for Web development resulted in a new wave of codification platforms for these types of coding systems. With modern layout and intelligent helping tools, it is possible to reduce considerably the codification errors, but also increasing the efficiency associated with the realization of those processes.

Therefore, this paper presents an insight into the development, implementation, and impact analysis of a Web application directed to the ICD-9-CM codification in a major Portuguese hospital located in the north of the country – CHP [II-1]. On the other hand, it also includes its subsequent update to the newer ICD-10-CM/PCS coding version. This scientific research has been undergoing since the end of 2016.

Thereby, the focus of this paper is mainly to highlight the differences between the prior method (by hand) and the new one with a computerized clinical coding tool, and the transition of the system from the ICD-9-CM coding version to ICD-10-CM/PCS.

In Section II, the state of the art and similar works related to this topic are described. Thereafter, in Section III, the research methodologies adopted are presented in detail. Section IV – "Single-page Application" – presents the work developed and its main results regarding the two versions

of the Web tool, followed by a brief Strengths Weaknesses Opportunities and Threats (SWOT) analysis in Section V. In Section VI, the conclusion and future work conclude briefly this paper.

3.2.3 State of the Art

The present section intends to highlight the main theoretical topics addressed throughout this manuscript, and the theory behind the realization of this work, as well as the main studies from the scientific community regarding medical codification, including the subsections "Diagnosis-related Groups", "ICD-9-CM Clinical Coding", "ICD-10-CM/PCS Clinical Coding", and "Interoperability: AIDA and AIDA-PCE".

3.2.3.1 Diagnosis-related Groups

A patient classification system is a method in which the main objective is to group patients or disease episodes in order to make it possible to identify their similarities and differences, and therefore allowing that those who belong to the same class are treated similarly.

In this context, DRGs consist of a patient classification system of patients hospitalized in acute hospitals that was developed in response to the rising costs and waste in the healthcare industry [II-2]–[II-4]. It groups patients into classes that are clinically consistent and similar in terms of resource consumption [II-2], [II-5]. Developed at the Yale University in the United States of America (USA) in the 60's, it is used since 1983 by Medicare to calculate the compensation in cases of hospitalization [II-6].

This classification system allows defining the set of goods and services that each patient receives according to his needs and the pathology that led to his hospitalization, as well as the defined treatment process. Thus, it is possible to relate the type of patients treated with the resource consumption [II-2]–[II-4],[II-7].

On the other hand, the concept of clinical coherence defines that the pathologies of the patients included in each DRG are related to an organ or system, or even with the etiology, and that the care provided is similarly the same for all the patients in that DRG [II-7]. A predetermined amount of money is disbursed to hospitals for the treatment of patients belonging to a given DRG, regardless of actual costs associated with the healthcare services provided to them [II-3], [II-4], [II-6], [II-8].

Thereby, the main motivation behind the grouping of patients from health institutions into DRGs is that it provides a financial and patient classification system that uses the diagnoses, surgical

interventions, age, gender, destination after discharge, and other related factors, as grouping criteria [II-2], [II-4], [II-8]. They were introduced in several countries, including Portugal, as a strategy for cost containment, planning, budgeting, management, and follow-up of the healthcare services provided to patients, reducing the disparities and errors [II-3], [II-6], [II-7].

The DRG requires a minimal dataset (MDS) in order to attribute one of the 25 main diagnosis categories to the discharge report [II-2]. So, the MDS includes, as follows [II-9]:

- The main diagnosis responsible for the patient admission;
- Other diagnoses;
- Procedures performed on the patient during the internment;
- Gender, age, and height;
- Destination after discharge (transferred, death or discharged against medical order).

Each DRG group has an associated relative weight and weighting coefficient, as well as an exception threshold for the number of hospitalization days that helps convert each case into equivalent patients [II-9].

Wilm Quentin and colleagues in "Hospital Payment Based on Diagnosis-related Groups Differs in Europe and Holds Lessons for the United States" highlights the differences between the original DRG and the one that countries like France, England or even Portugal implement. This adaption is the basis of most European countries method to finance hospitals, proving to be less cost worthy with a high quality of services [II-3].

Carina Fourie et al. [II-6] present in "Systematically Evaluating the Impact of Diagnosis-related Groups on Healthcare Delivery: A Matrix of Ethical Implications" a study of ethical implications and importance of the DRGs in diverse Swiss hospitals [II-6].

On other hand, in order to highlight the diversity of the subject, Yantao Xin presented a comparison of the amount of medical waste generated in major healthcare units using as basis the DRGs [II-4].

3.2.3.2 ICD-9-CM Clinical Coding

For the purpose of coding hospital discharges in terms of diagnoses and procedures, in order to allow the subsequent grouping of those episodes in DRGs, the clinical coding "ICD-9-CM" is applied [II-8]. In short, it consists of a set of diagnosis and procedure codes used for the classification and coding of the morbidity and mortality information for statistical purposes, and for

the indexing of hospital records by disease and surgical interventions. The information is then used for storage and research purposes.

Thus, the DRG classification process is performed by a coding professional who must know the system and the classification structure of the ICD-9-CM clinical coding, understand the organization of the indexes and their use in the coding of diseases and procedures, as well as how to apply correctly the principles and rules of the clinical coding ICD-9-CM [II-8].

Each time a patient is discharged from a healthcare unit, a discharge report is issued from the daily logs by the physician in charge of the patient. The ICD-9-CM clinical coding is a perfect fit to encode diagnoses, medical procedures, external causes, and morphologies, consisting in a universal list of codes recognized in any country across the world.

As the name states, it is an adaptation of the ICD-9 codification system defined and implemented by the U.S. Department of Health in collaboration with the Medicare and Medicaid Service Centers [II-10]. With more than 13,000 diagnoses and 3,500 procedures, it is essential to develop tools or systems focused on the codification process.

Some Web applications like “Find-A-Code” or the work of Marisa Teresa Chiaravalloti et al. [II-11] in “A Coding Support System for the ICD-9-CM Standard” are examples of systems developed for codification purposes [II-11]. The late one processes text in natural language using text mining algorithms, returning a list of possible codes for each case.

After the codification process, every discharge report can be read and perfectly understood in every country that adopts the same terminology.

For the purpose of the present work, and as stated by international directives, the ICD-9-CM codes are the basis of the DRGs decision [II-3]. Thereafter, the ICD-9-CM SPA coding tool was updated with the newer ICD-10-CM/PCS coding version, initially released in order to replace the ICD-9-CM version.

3.2.3.3 ICD-10-CM/PCS Clinical Coding

Firstly, it is important to note that ICD-10, unlike its precursor (ICD-9), is divided into two different parts, namely: [II-12], [II-13]

- ICD-10-CM for diagnosis coding;
- ICD-10-PCS for inpatient procedure coding.

When ICD-10 was defined, it was projected that ICD-10-CM would become a standard for all USA healthcare settings, whereas ICD-10-PCS would be required in inpatient settings only [II-12].

The ICD-10-CM/PCS clinical coding was mainly designed to offer notable advantages over ICD-9-CM, such as generating higher-quality clinical data. This would result in major improvements in the quality and the use of data in a significant number of healthcare settings, including driving a better healthcare management and improving significantly outcomes [II-12].

Supporters of ICD-10-CM/PCS defend that it incorporates greater specificity than its precursor in describing healthcare problems, but also clinical data, offers the addition of information relevant to ambulatory and managed care encounters, and impressively expanded injury codes that reflect the location of the injury [II-12], [II-14]–[II-16]. Thus, they praise its undeniable ability to provide a more detailed description of clinical situations, and substantially increasing the level of detail that can be captured [II-12], [II-13], [II-16], [II-17].

Furthermore, the structure of ICD-10-CM/PCS enables the possibility of greater expansion of code numbers, including risk factors that are regularly encountered in a primary care setting, instead of allowing the classification of diseases and injuries only [II-14].

In addition, while ICD-9 uses numeric codes (e.g., 001-999), ICD-10 uses an alphanumeric classification system which consists essentially of 1 letter followed by up to 3 numbers at a 4-character level (e.g., A00.0-Z99.9) [II-17], [II-18].

Thus, the alphanumeric format of ICD-10 offers a better structure than ICD-9, which allows a significant space for future revision without disruption of the numbering system [II-19].

On the other hand, the overall number of codes and diagnoses has increased significantly: about 17000 codes for ICD-9-CM to more than 155000 codes for ICD-10-CM/PCS [II-12], [II-13], [II-15]–[II-17], [II-20]. Thereby, the newer version includes previously unavailable codes since it enables a more detailed description of clinical situations, including for instance codes to distinguish between different types of diabetes or even the location on the patient's body of the health condition (e.g., left or right limb) [II-12].

Thus, while the main axis for ICD-9 is the nature of the health condition itself, the main axis of ICD-10 is the body region of the health condition with the highest level of specificity reachable [II-21].

In the long run, due to this improved level of detail, it is expected that ICD-10-CM will decrease medical fraud and abuse [II-12]. For instance, since even the location on the patient's body is registered in the codification process, it will reduce the possibility of coding professionals repeatedly reporting the same procedure on the same location of the body.

3.2.3.4 Interoperability: AIDA and AIDA-PCE

Nowadays, with the continuous growth of the clinical information stored into the hospital information systems (HISs), one of the major interests in the Medical Informatics field is to ensure interoperability between different information systems in health institutions [II-22], [II-23]. Thus, interoperability is increasingly considered a requirement in HISs rather than an option in order to implement an adequate communication and cooperation between distinct systems [II-24].

In short, the concept of interoperability can be defined as the ability of a system to communicate and share information with another system that arises in order to overcome the heterogeneity and distribution of several different sources of information. In the healthcare industry, the main goal of interoperability is to connect applications and data so that they can be shared across the organization and distributed to health professionals [II-22], [II-24].

Thereby, in this context, there is a need to implement dynamic platforms, such as multi-agent systems that allow the access and sharing of information between different information systems, in order to connect them, standardize distributed clinical systems, and thus reduce the delays normally generated in the process of sharing information [II-25].

The register of clinical information in CHP is ensured by a few HISs, namely the "Sistema de Apoio ao Médico" (SAM) –Medical Support System, the "Sistema de Gestão de Doentes Hospitalares" (SONHO) – Hospital Patient Management System, the "Sistema de Apoio à Prática de Enfermagem" (SAPE) – Nursing Support System, and the "Processo Clínico Eletrónico" (PCE) – Electronic Medical Record [II-23], [II-26].

In this context, the AIDA ("Agência para Integração, Difusão e Arquivo de Informação Médica") platform emerges. Some Portuguese health systems are equipped with the AIDA platform, including CHP, which uses proactive intelligent agents that ensure the interoperability between different and heterogeneous HISs, and other entities such as the complementary systems, including SAM, SONHO, SAPE, PCE, RIS ("Radiology Information System"), LIS ("Laboratory Information System"), DIS ("Department Information System"), and AIS ("Administrative Information System"), among others [II-22]–[II-26].

AIDA is a complex system consisting of simple and specialized subsystems, defined as intelligent agents, which are responsible for tasks such as the communication between heterogeneous systems, the sending and receiving of information (for example, clinical reports, medical images, and prescriptions), as well as the management and storage of data [II-22].

Thus, directly from AIDA it is possible to integrate, disseminate, and archive large sets of data from different sources (for example, services, departments, healthcare units, computers, and medical devices) [II-22]. In this way, the AIDA platform provides an easy access and sharing of registered information, facilitating medical research and the application and development of other computational tools, such as the ICD-9-CM and ICD-10-CM/PCS clinical coding tools described in this manuscript, in order to optimize the healthcare services provided by the health institution.

The electronic health record (EHR) is responsible for the safe and organized storage of all the information regarding a patient, from personal data to diagnoses and procedures [II-27]–[II-30].

The constant update is vital in this scenario, so the same research team from the Algoritmi Research Centre (University of Minho, Braga, Portugal) that developed AIDA put together the AIDA-PCE. Following the Problem Oriented Medical Record (POMR), all the patient information regarding symptoms, medical observations, diagnoses, and treatment plans are stored inside that structure.

Although these systems allow the insertion of free text and other non-universal information, the AIDA and AIDA-PCE present innovative and novel solutions to accomplish interoperability. On the other hand, the incorporation of the ICD-9-CM codes, and subsequently the ICD-10-CM/PCS codes, into the EHR represents an important feature in order to accomplish a cross-border medical record. The storage of ICD-9-CM and ICD-10-CM/PCS coded discharge reports saves space on the AIDA-PCE databases and reduces medical errors.

The next section presents the main research methodologies followed to implement this work.

3.2.4 Research Methodologies

The realization of any study in the field of Information Technologies (IT) includes the scrutinized research and analysis of the set of methodologies and Technologies available and feasible in the design of the defined IT solutions. The choice of the most appropriate methods and tools is mostly based on the advantages pointed out, as well as on associated limitations and compliance issues with related systems.

Thus, the achievement of the SPA for ICD-9-CM, and its transition from the ICD-9-CM codification to the newer ICD-10-CM/PCS version, are based on the research methodology Design Science Research (DSR), mostly used in the construction and evaluation of useful and rigorous IT solutions. Each of the design phases presented in this study included the choice and use of the

most appropriate methodologies, technologies, and tools for the definition and elaboration of the desired solution. Finally, a Proof of Concept (PoC) was also carried out corroborating the viability and usefulness of the clinical coding tool for ICD-9-CM designed and developed, and its update to the newer ICD-10-CM/PCS version, which consisted essentially of a SWOT analysis (Section V).

Thereby, a brief description of these two research methodologies are presented in this section, namely Design Science Research and Proof of Concept, in subsections A and B, respectively.

3.2.4.1 Design Science Research

In the area of ITs, the main objective of the use of the research methodology Design Science Research (DSR) is the construction and evaluation of objects, also called "artefacts", that allow professionals to process organizational information and develop actions to solve a problem [II-31], [II-32].

Thus, the methodology that drove the realization of this project is the DSR. It consists of a rigorous method of scientific research used to develop successful artefacts [II-33]. It focuses on the IT artefact with a high priority in its relevancy in its application domain. Thus, in the context of solving real world business problems, it is critical to try to improve the relevance and usefulness of the artefact [II-34], [II-35]. The designed appliance must correspond to a viable technological solution for solving important and relevant business problems, and its usefulness, quality, and effectiveness must be rigorously demonstrated through well-executed evaluation methods. In addition, research should provide clear and verifiable contributions, and should be based on the application of rigorous methods in its construction and evaluation process [II-34], [II-35].

In Figure 7, the research methodology DSR is outlined, that is, its different interconnected steps that synthesize the steps to be followed through the DSR in the construction of scientific IT artefacts, namely the steps of “Identify Problem & Motivate”, “Define Objectives of a Solution”, “Design & Development”, “Demonstration”, “Evaluation”, and “Communication”. These are the phases adopted in the design of this case study.

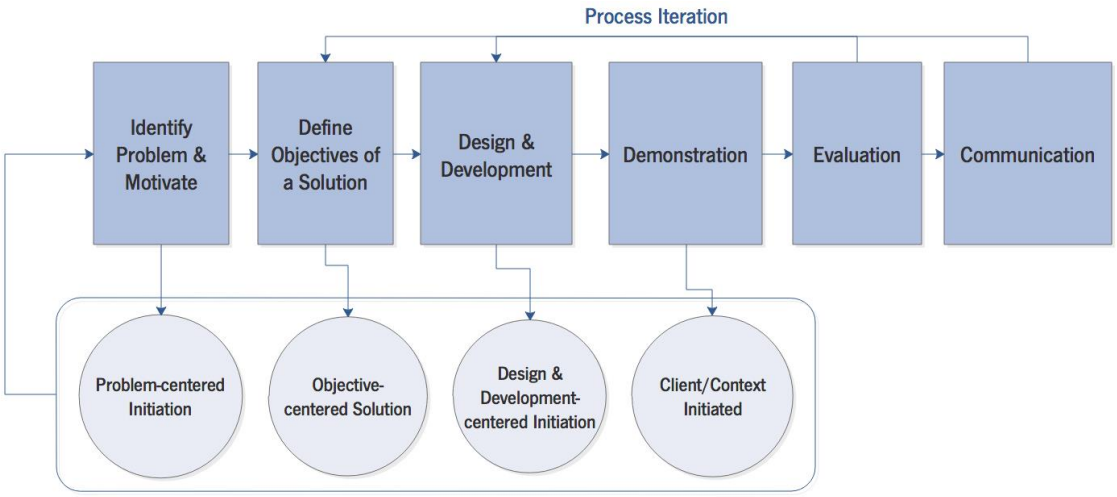


Figure 7 - Schematic representation of the research methodology Design Science Research (adapted from [II-33])

In short, in the first steps, the problem and the motivation are defined, as well as the objectives of the solution found. Then the artefact is designed and developed, directed to an important business problem to be solved that must be relevant to the solution of the same. Its development must follow a rigorous scientific process based on the knowledge and the theory already explored. Finally, the solution must be demonstrated, evaluated, communicated, and propagated efficiently to the target audience [II-31], [II-36].

Thus, the case study described in this manuscript follows the DSR research methodology because the IT solution defined meets the needs of health professionals of CHP, that is, a new clinical coding tool to support clinical practice assisting in their work, meeting the challenges currently existing at CHP's facilities.

Therefore, it provides the health institution with na appropriate and well-founded solution, based on methods and technologies that have already been explored and adapted to solve the problem in question, and also stimulate new knowledge for the organization and the scientific community. Thus, the development of this project additionally included the dissemination of the IT artifact to CHP's professionals, as well as the writing of scientific papers.

Finally, it is important to note that the clinical tool developed, including its update to the ICD-10-CM/PCS version, was duly evaluated through a SWOT analysis. It should also be pointed out that the application of the Proof of Concept research methodology to prove the feasibility,

usefulness, and usability of the tool, which included a SWOT analysis, is briefly described in Subsection B.

3.2.5 Single-Page Application

The codification of the discharge reports was made manually in CHP, making the process too slow and with a high error probability. Therefore, it emerged the need to create a process that would reduce the codification time. Thus, it was developed a Single Page Application (SPA) through which the health professionals are able to perform the codification process, and at the same time to consult patients' data, such as the discharge report, the personal information, and the hospital services where the patient was admitted.

The first version was created based on the ICD-9-CM coding version. However, there was a need to adopt a more advanced coding system, namely the ICD-10-CM/PCS codification.

3.2.5.1 SPA Version 1

The purposed layout in this scenario encompasses a solution with three main components: the patient information, the codification area, and the discharge report. The codification area is divided into five frames: diagnoses, external causes, procedures, tumor morphology, and observations. All the boards, with exception of the observations board, which is the only one that allows free text insertion, are composed by rows divided mainly in priority, description, and code.

The dynamic and aided search leads to a faster process than the already existing one. When a word is typed in the description field, a list of all the ICD-9-CM codes is presented, and when the user picks one of them on the description, the respective code is automatically filled. On the other hand, when the user enters the code, the description is also automatically filled.

When a codification is finished, the user sends it to an evaluator. If the codification fails in this evaluation phase for any reason, it is sent again to the list of codifications that needs to be done. Thus, the application was developed with three different modules. One to be used in the first codification, another to be used when a discharge report was already codified and for some reason failed in the evaluation phase, and one final module to be used when the user only has view permissions. In this last module, the users are not able to change any field present in the codification.

The application development leaned on the LAMP architecture. It uses Linux as the operative system, Apache as the Web server, MySQL as the relational database management system (RDBMS), and PHP as the object-oriented language.

In order to develop a fluid and dynamic application, the AngularJS framework greatly contributed. The modularity and extensibility of this JavaScript framework allows the development of diverse and futuristic applications.

The database storages all the data related to the codification codes, the discharge reports, user information, and all the information generated by the codification process. The RESTful Web service mediates the communication between the SPA and the database.

3.2.5.2 SPA Version 2

In the second version of the SPA, all the SPA architecture remained the same. The difference from the first version to the second was only the use of a different codification system, which consequently forced a layout change in the codification area. In this version, the codification area only presents three frames: diagnoses, procedures, and observations.

The dynamic and aided search is still present in this version.

Once the ICD-10-CM/PCS codification presents a hierarchical structure, it is easier to find a code since the order is followed, so, it was added the help button in the code insertion. When the user selects this button, it opens a new frame with dynamic search. This frame is divided into seven parts: section, body system, root operation, body part, approach, device, and qualifier. At first, all components are empty except for the component section. Since this is the root of coding, it represents the first character of the code. When the user selects a section, the body system is automatically filled with options that are related with the section selected. In turn, when the body system is selected, it automatically fills the root operation options. When the three components section, body system, and root operation are selected, a board with all the other components is filled. And after all seven parts have been selected, the code is finally complete and, thereafter, it is sent to the main frame to the row where the user clicked in the help button. It should additionally be noted that a process of validation for the code exists.

Although this second version also use MySQL as the RDBMS, the tables of the database needed to be changed in order to be able to store the new coding system.

The next section presents the SWOT analysis of the Web application developed.

3.2.6 SWOT Analysis

To test the viability, the utility, the quality, and the efficiency of the application, a PoC was necessary, in this case, a SWOT analysis. This analysis allows to analysis the strengths, weaknesses, opportunities, and threats of the application [II-39].

With the update of the coding system for the ICD-10-CM/PCS codification, a weakness of the project was eliminated, thus, becoming a more up-to-date project.

Therefore, it was made a SWOT analysis of the second version of the project.

Strengths:

- High usability, intuitive, and easy to learn (user-friendly);
- Easy access to the data of patients, as well as the hospital services in which the patient was;
- High scalability;
- Easy of reissue of coded discharge reports;
- Decrease of the codification time of the discharge reports;
- Decrease of human error;
- Easy adaptability to different health institutions.

Weaknesses:

- Requires internet connection.

Opportunities:

- Modernization and organizational development;
- Increasing expectation of the hospital administration to obtain methods that facilitate the hospital financing calculation;
- Provide the tool to help in the calculation of the hospital financing.

Threats:

- Lack of acceptance to resort to new technologies by health professionals.

The next section presents the conclusion and future work of this study.

3.2.7 Conclusion and Future Work

Finally, the realization of this case study allowed the development of a clinical practice tool, namely a user-friendly clinical coding tool for ICD-9-CM, and its subsequent transition to the newer ICD-10-CM/PCS coding version. The Web application is currently implemented in a production machine of CHP, and it is currently being used by the coding professionals of the hospital in order to perform the clinical coding of the episodes of hospital discharges from patients admitted to CHP. This will then facilitate the grouping of processes into DRGs, that is, a financial system that can manage the costs and waste associated with healthcare services. In the coming years, the expansion of the Web application is expected.

Trained professionals using the ICD-9-CM and ICD-10-CM/PCS clinical coding tools reported significant differences in time consumption and committed errors when using a computerized system to perform their tasks. It represents an asset to its users, since it facilitates the work of health professionals, and increases their capacity and speed of work by reducing the number of tasks required to perform a certain codification. In this way, the development of the clinical tool allows the centralization of a set of tasks and information in a SPA, greatly benefiting its users.

When comparing the SPA for ICD-9-CM codification with its update to the newer ICD-10-CM/PCS version, health professionals defend that the advantages are much greater, including a better specificity in describing clinical situations, leading to an improved level of detail. They also defend that it is easier to find a code in the new update due to the help bottom. Thus, they do not need to remember all the codes or

their entire description, making it more intuitive.

Regarding future work, the addition of a Business Intelligence (BI) module in the clinical coding tool for ICD-10-CM/PCS is foreseen, that is, the addition of a module with clinical and performance indicators [II-40]. Its principal aim is the visualization of indicators that show the association between the number of coded processes and each coding professional, as well as the temporal evolution of the number of processes encoded by each coding professional. Thus, the main objective of the insertion of this module is to study and analyze the performance of the coding professionals, that is, to identify, for example, the coding professionals who codify the most, and also those who are coding the least. In this way, it is tried to encourage even more the increase of the production of health professionals at CHP.

Finally, in early 2018, the research team already began the implementation of the system in more health institutions in Portugal. It is thus confirmed our strong desire to continue to expand the system across the country due to its undeniable advantages, which were defended throughout this paper.

3.2.8 References

[II-1] C. Coimbra et al., "Improving Diagnosis-related Groups with a Computerized Clinical Coding Tool for ICD-9-CM Codification," in eTELEMED 2017: The Ninth International Conference on eHealth, Telemedicine, and Social Medicine, 2017, pp. 33–36.

[II-2] M. M. Bellanger, W. Quentin, and S. S. Tan, "Childbirth and Diagnosis Related Groups (DRGs): Patient Classification and Hospital Reimbursement in 11 European Countries," Eur. J. Obstet. Gynecol. Reprod. Biol., vol. 168, no. 1, pp. 12–19, May 2013.

[II-3] W. Quentin, D. Scheller-Kreinsen, M. Blumel, A. Geissler, and R. Busse, "Hospital Payment Based On Diagnosis-Related Groups Differs In Europe And Holds Lessons For The United States," Health Aff., vol. 32, no. 4, pp. 713–723, Apr. 2013.

[II-4] Y. Xin, "Comparison of Hospital Medical Waste Generation Rate Based on Diagnosis-related Groups," J. Clean. Prod., vol. 100, pp. 202–207, 2015.

[II-5] M. Barrento, "Sistema de Business Intelligence Aplicado aos Grupos de Diagnósticos Homogéneos," Universidade Nova de Lisboa, 2010.

[II-6] C. Fourie, N. Biller-Andorno, and V. Wild, "Systematically Evaluating the Impact of Diagnosisrelated Groups (DRGs) on Healthcare Delivery: A Matrix of Ethical Implications," Health Policy (New. York)., vol. 115, no. 2–3, pp. 157–164, 2014.

[II-7] R. Busse et al., "Diagnosis Related Groups in Europe: Moving Towards Transparency, Efficiency, and Quality in Hospitals?," BMJ, vol. 346, 2013.

[II-8] P. Jorge, "Plataforma de Dados de Saúde – Portal Institucional," Instituto Superior de Engenharia do Porto, 2013.

[II-9] S. B. Rimler, B. D. Gale, and D. L. Reede, "Diagnosis-related Groups and Hospital Inpatient Federal Reimbursement," RadioGraphics, vol. 35, no.6, pp. 1825–1834, Oct. 2015.

[II-10] M. Rouse and D. Fluckinger, "ICD-9-CM (International Classification of Diseases, Ninth Revision, Clinical Modification)," 2014. [Online]. Available: http://searchhealthit.techtarget.com/definition/ICD-9- CM. [Accessed: 31-Jan-2017].

[II-11] M. T. Chiaravalloti, R. Guarasci, V. Lagani, E. Pasceri, and R. Trunfio, "A Coding Support System for the ICD-9-CM Standard," Proc. - 2014 IEEE Int.Conf. Healthc. Informatics, ICHI 2014, pp. 71–78, 2014.

[II-12] M. Topaz, L. Shafran-Topaz, and K. H. Bowles, "ICD-9 to ICD-10: Evolution, Revolution, and Current Debates in the United States," Perspect. Heal. Inf. Manag. / AHIMA (American Heal. Inf. Manag. Assoc., vol. 10, no. Spring, 2013.

[II-13] R. Mills, R. Butler, E. McCullough, M. Bao, and R. Averill, "Impact of the Transition to ICD-10 on Medicare Inpatient Hospital Payments," Medicare Medicaid Res. Rev., vol. 2, no. 2, pp. 1–13, 2011.

[II-14] A. Hazlewood, "ICD-9 CM to ICD-10 CM: Implementation Issues and Challenges," The American Health Information Management Association, 2017. [Online]. Available: http://library.ahima.org/doc?oid=59978#.WUqLXMZ OrVo. [Accessed: 21-Jun-2017].

[II-15] J. A. Hirsch et al., "ICD-10: History and Context," Am. J. Neuroradiol., vol. 37, no. 4, pp. 596–599, 2016.

[II-16] A. G. Fiks and R. W. Grundmeier, "Elucidating Challenges and Opportunities in the Transition to ICD-10-CM" Pediatrics, vol. 134, no. 1, pp. 169–170, 2014.

[II-17] J. Armstrong, P. McDermott, G. R. Saade, and S. K. Srinivas, "Coding Update of the SMFM Definition of Low Risk Cesarean Delivery from ICD-9-CM to ICD-10-CM," Am. J. Obstet. Gynecol., 2017.

[II-18] N. Jetté et al., "The Development, Evolution, and Modifications of ICD-10," Med. Care, vol. 48, no. 12,pp. 1105–1110, 2010.

[II-19] S. Bowman, "Why ICD-10 Is Worth the Trouble," The American Health Information Management Association, 2017. [Online]. Available: http://library.ahima.org/doc?oid=78673#.WUqMIMZ OrVo. [Accessed: 21-Jun-2017].

[II-20] R. Caskey et al., "The Transition to ICD-10-CM: Challenges for Pediatric Practice," Pediatrics, vol.134, no. 1, pp. 31–36, 2014.

[II-21] J. R. Lowe, G. Raugi, G. E. Reiber, and J. D. Whitney, "Changes in Classifications of Chronic Lower - Limb Wound Codes in Patients with Diabetes: ICD-9-CM Versus ICD-10-CM," Adv. Skin Wound Care, vol. 28, no. 2, pp. 84–92, 2015.

[II-22] L. Cardoso, F. Marins, F. Portela, A. Abelha, and J. Machado, "Healthcare Interoperability through Intelligent Agent Technology," Procedia Technol., vol. 16, pp. 1334–1341, 2014.

[II-23] H. Peixoto, M. Santos, A. Abelha, and J. Machado, "Intelligence in Interoperability with AIDA," in Foundations of Intelligent Systems, Springer Berlin Heidelberg, 2012, pp. 264–273.

[II-24] L. Cardoso, F. Marins, F. Portela, M. Santos, A. Abelha, and J. Machado, "The Next Generation of Interoperability Agents in Healthcare," Int. J. Environ. Res. Public Health, vol. 11, no. 5, pp. 5349–5371, 2014.

[II-25] A. Abelha, C. Analide, J. Machado, J. Neves, M. Santos, and P. Novais, "Ambient Intelligence and Simulation in Health Care Virtual Scenarios," IFIP Int. Fed. Inf. Process., vol. 243, pp. 461–468, 2007.

[II-26] J. Machado, A. Abelha, J. Neves, and M. Santos, "Ambient Intelligence in Medicine," in IEEE Biomedical Circuits and Systems Conference, 2006, pp. 94–97.

[II-27] M. Salazar et al., "Step towards Paper Free Hospital through Electronic Health Record," in Advances in Information Systems and Technologies, Springer International Publishing, 2013, pp. 685–694.

[II-28] J. Duarte et al., "Data Quality Evaluation of Electronic Health Records in the Hospital Admission Process," in EEE/ACIS 9th International Conference on Computer and Information Science, 2010, pp. 201–206.

[II-29] J. Duarte, C. F. Portela, A. Abelha, J. Machado, and M. F. Santos, "Electronic Health Record in Dermatology Service," in ENTERprise Information Systems, Springer Berlin Heidelberg, 2011, pp. 156–164.

[II-30] F. Portela et al., "Electronic Health Records in the Emergency Room," in IEEE/ACIS 9th International Conference on Computer and Information Science, 2010, pp. 195–200.

[II-31] S. T. March and V. C. Storey, "Design Science in the Information Systems discipline: An introduction to the special issue on Design Science Research," MIS Quartly, vol. 32, no. 4, pp. 725–730, 2008.

[II-32] M. Bilandzic and J. Venable, "Towards Participatory Action Design Research: Adapting Action Research and Design Science Research Methods for Urban Informatics," J. Community Informatics, vol.7, no. 3, 2011.

[II-33] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, "A Design Science Research Methodology for Information Systems Research," J. Manag. Inf. Syst., vol. 24, no. 3, pp. 45–77, 2007.

[II-34] a. R. Hevner, S. T. March, and J. Park, "Design Science in Information Systems Research," MIS Q., vol. 28, no. 1, pp. 75–105, 2004.

[II-35] A. R. Hevner and S. Chatterjee, "Design Research in Information Systems," Des. Res. Inf. Syst., vol.28, pp.75–105, 2004.

[II-36] V. K. Vaishnavi and W. Kuechler, Design Science Research Methods and Patterns: Innovating Information and Communication Technology, 2nd ed. CRC Press, 2015.

[II-37] 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, 2015, vol. 166, pp. 240–245.

[II-38] B. Schmidt, "Proof of Principle Studies," Epilepsy Res., vol. 68, no. 1, pp. 48–52, 2006.

[II-39] R. Pereira, M. Salazar, A. Abelha, and J. Machado, "SWOT Analysis of a Portuguese Electronic Health Record," in Collaborative, Trusted and Privacy- Aware e/m-Services, Springer Berlin Heidelberg, 2013, pp. 169–177.

[II-40] A. Brandão et al., "A Benchmarking Analysis of Open-Source Business Intelligence Tools in Healthcare Environments," Information, vol. 7, no. 4,p. 57, Oct. 2016.

3.3. Study III – Improving the Management of Hospital Discharges

Ana Cecília Coimbra, José Machado, António Abelha

The 2nd International Conference on
Modern Management based on Big Data, 2021
Status: submission process

3.3.1 Abstract

At Centro Hospitalar Universitário do Porto (CHUP), a tool for computerized clinical coding was developed to assist in the codification of hospital discharge. However, for this tool to be useful, it is necessary to have a process to manage the entire coding process. Thus, a platform was developed to help manage the coding of hospital discharge episodes. The biggest advantage of the existence of this platform is better to organize the entire coding process in order to improve the quantity and quality of work performed by CHUP's health professionals.

3.3.2 Introduction

Hospital records are quite complex to be processed, not only for their quantity, but also for the variety of data. The number of different occurrences and combinations of factors generates an enormous diversity of data. Thus, it is difficult to assess hospital productivity, therefore, the need arises to use DGRs, which in turn rely on coding systems that translate diagnoses and procedures to group episodes in groups with resembling characteristics that in turn use similar hospital resources.

In the paper "Improving the Codification of Hospital Discharges with an ICD-9-CM Single-page Application and its Transition to ICD-10-CM/PCS" [III-1] a Single-page application (SPA) was presented. The goal was to facilitate the coding process by making all the necessary data available for users to carry out the coding of each episode. It is possible to fill the fields related to the codification of diagnosis and procedures in a single page and at the same time view the general data of the patient and the discharge report. This data integration makes the coding process faster and more efficient since users do not have to consult the necessary data on several platforms at the same time.

However, this SPA does not bring much advantage if there is no platform that manages the episodes, that is, that differentiates between coded and non-coded episodes, among other things. The focus of this paper is to describe the workflow of the management system of the codification process.

3.3.3 State of the Art

In this section are presented the main theoretic topics behind the realization of this work.

The Medical Coding and DRGs systems, are both powerful tools that have been employed in the modern era to satisfy financial and statistical needs [III-2].

3.3.3.1 Diagnosis Related Groups

The Diagnosis Related Groups (DRG) is a very important and powerful tool in terms of financial and statistical indicators in the health sector. How a health institution is able to quantify all the resources that are allocated in all different situations that happens in a hospital? The DGRs are used for helping in that topic, helping in classify and divide the clinical episodes into groups with similar resources allocations [III-2]. Epidemiological information, diagnostics, clinical features and procedures are used by the DRG system to assign episodes to an individual DRG (group) [III-3].

The benefit of using the DRG system is based on the premise that since each group contains episodes with similar characteristics (course of disease, length of hospital stay (LOHS) and treatment requirements) then their costs will also be similar. With this in mind, it is thus possible to assign a certain reimbursement to a DRG group and thus reduce medical expenses and related financial charges while maintaining the quality of medical services and ensuring that resources are allocated correctly [III-3], [III-4].

However, if an error in assign a DRG to a case occurs, it can have enormous financial repercussion [III-2].

3.3.3.2 ICD-10-CM

The International Classification of Diseases Tenth Revision Clinical Modification (ICD-10-CM) it is a medical terminology which intends to standardize all medical information, so that in this way it is possible to compare clinical records anywhere in the world [III-4], [III-5].

The ICD-10 terminology is divided into two different parts: ICD-10-CM for diagnostic codes and ICD-10-PCS for procedure codes in patients. This terminology was developed with the aim of improving the existing one (ICD-9), the main difference being the quantity and specificity of the clinical data [III-6]-[III-8].

While ICD-9-CM has 14,567 diagnostic codes and 3882 procedure codes, the ICD-10-CM/PCS version has 69823 diagnostic codes and 71974 procedure codes, this was only possible by changing the format of the codes, changing from numeric (ICD-9-CM) to alphanumeric (ICD-10-CM/PCS). In general, ICD-10-CM/PCS is a more complex system that allows greater specificity and, consequently, a better classification of clinical episodes [III-9], [III-10].

If hospital units adopt this terminology in their clinical records, it makes it easier to use DGRs since having standardized records facilitates their distribution to groups.

3.3.3.3 Hospital Morbidity Information System

The *Administração Central do Sistema de Saúde* (ACSS) and the *Serviços Partilhados do Ministério da Saúde* (SPMS) developed a new Hospital Morbidity Information System (Sistema de Informação de Morbilidade Hospitalar - SMIH) that uses ICD10-CM/PCS as a coding system. The aim is the existence of a cross section across the entire National Health System (Serviço Nacional de Saúde – SNS), where clinical coding is done in ICD-10-CM/PCS of the episodes by each doctor in his working hospital, being an online platform and based on central data [III-11].

3.3.4 Research Methodologies

Design Research (DR) is essential to create products, services and systems capable of responding to human needs [III-12]. The difference between DR and Design Science Research (DSR).is that the first corresponds to research on or about design, while the second corresponds to research using design as a method or technique. Learning through the construction of artifacts is the characteristic that defines the DSR [III-13].

Figure 8 shows the DSR model that emerges from the Takeda model adaptation [III-13].

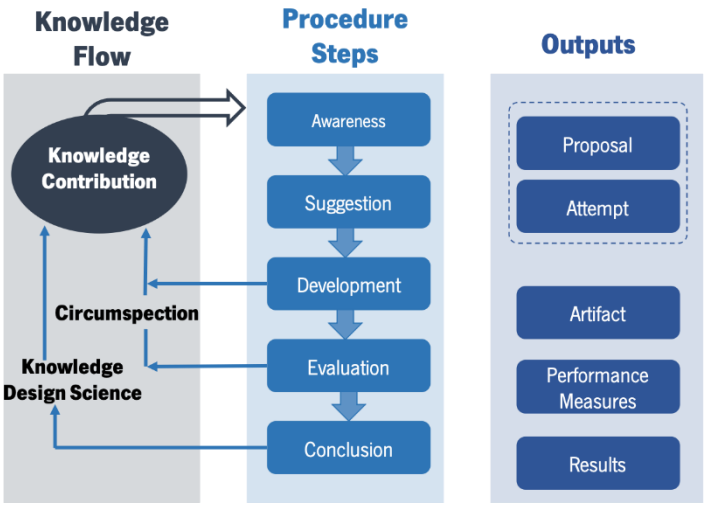


Figure 8 - DSR Model (adapted from ([III-13])).

A typical DSR model follows the following methodology:

- *Awareness:* In this phase the problem is identified, presenting a research proposal.
- *Suggestion:* Considering the proposal made in the previous phase, the objectives for its development are defined. It consists of a creative phase, where in addition to existing elements, new elements can also emerge.
- *Development:* This phase focuses on the development and implementation of the artifact.
- *Evaluation:* Evaluation of the artifact, according to the criteria outlined in the first phase and the outline of improvements, if necessary.
- *Conclusion:* In this phase it can result in two paths: if the results obtained are satisfactory then the end of the research is given, otherwise, the DSR cycle is restarted.

3.3.4.1 Proof of Concept

Proof of concept (PoC) is used to assess whether a concept or theory is valid and can be proven through the practical model. In the scope of information systems (IT) it is used to assess whether the objective for which the technology was developed has been achieved or not [III-14], [III-15].

In this paper the PoC used was SWOT analysis and it is described in section 3.4.6.

3.3.5 Management platform

This section is subdivided, the first part describes the coding process management platform, and the second part describes the platform connection with the SIMH system.

3.3.5.1 "*E-codificação*" platform

The coding process management platform is called "*e-codificação*". This platform contains all episodes that have already been discharged from the hospital.

The management platform will be used by medical coders and administrative personnel who will manage and distribute episodes by coders. However, the confidentiality of patient data is a factor that has to be taken into account, thus, the clinical information of each patient can only be viewed by authorizes users. With this in mind, three types of users with different access were created: administrator, administrative staff and coding physician.

The administrator, as the name implies is the platform administrator and has access to all the information.

The administrative staff manages episodes, distributing them by coders. This type of users able to access only general information such as the process number, medical specialty in which the episode was discharged, date of discharge, among others, without ever having access to clinical information of the patient.

On the other hand, the third type of user, the coding physician, has access to all clinical information about the episodes attributed to him, nevertheless, it is not able to access episodes that are attributed to other doctors.

An episode from entering in the platform until the coding is completed goes through different states. The possible states are:

- **Discharge to close** – episodes that are closed, but may be a lack of essential elements for coding, and therefore they cannot yet be distributed to coders;
- **Discharge closed** – episodes ready to be distributed;
- **Pending** – episodes that were flagged by the coders with any problem, whether due to lack of information or whether, for example, the impossibility to access the discharge report;
- **In coding** – episodes that are currently being coded by coding physicians;
- **In audit** – episodes that were completed by the coders, but that need to be validated to be accepted as completed;

- To recode – episodes that have codification done, however this codification needs to be redone or corrected;
- To re-audit – episodes that is not the first time that the audit is done;
- Conclude – episodes that have the codification completed;

The workflow, in terms of states, of the management platform is presented in Figure 9. In short, an episode starts with the status "Discharge to close", and if no important element is missing, it is distributed to an appropriate encoder by the type of medical specialty and type of episode (interment or ambulatory).

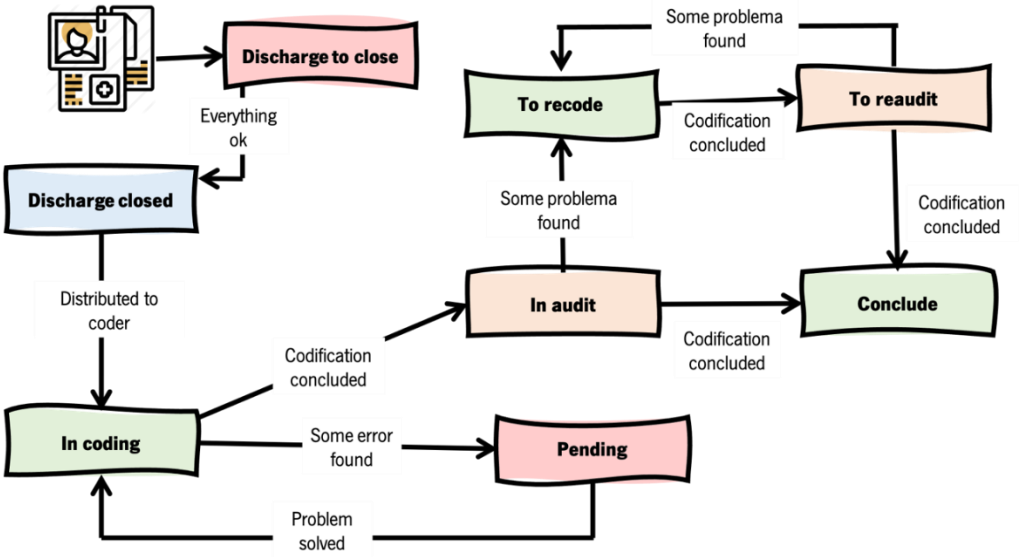


Figure 9 - Worflow e-codificação

In the "in coding" state, the encoder does as the name indicates the codification of the episode, if it finds any problem, for example, lack of anesthesia report in the discharge report file, or it is not possible to view the report, the encoder sends the episode to the "Pending" folder. In this state, episodes are reprocessed, and problems are solved and sent back to the state "in coding". In case everything is alright, at the end of the codification, it goes to "In audit" status, the episodes cannot be "concluded" immediately as it is necessary to have an evaluation of the codification done.

In this state, it is confirmed if the coding was done correctly or if it contains any error in this state. If there is an error, it is sent to "recode" to the same encoder, and once finished it goes to the "to re-code" state, and then if everything is alright, it is finally finished, otherwise it goes back to recode.

3.3.5.2 SIMH Interoperability

Although the developed system contains a section where the codification can be done, the Administração Central do Sistema de Saúde (ACSS) and the Shared Services of the Ministry of Health (Serviços Partilhados do Ministério da Saúde - SPMS) developed the "Sistema de Informação para a Morbilidade Hospitalar" (SIMH), the coding system to allow centralization of all clinical data encoded all hospitals in Portugal.

However, in addition to the new episodes being able to be encoded in SIMH it is also necessary to have a migration of the episodes that were previously encoded in "*e-codificação*".

Therefore, an agent was then created that would collect data from episodes already coded, and with that would create a JSON, with the pre-defined format, and this would be sent to SIMH through a web service.

In this way, it allows coders to maintain the system they are used to working with, (the "*e-codificação*") and acquire interoperability with the central SIMH database because all the episodes encoded in the "*e-codificação"* are, when concluded, sent to the SIMH through the agent that was developed.

3.3.6 SWOT Analysis

A SWOT Analysis was used as a proof of concepts with the purpose of testing the viability, the utility, the quality, and the efficiency of the application. Being possible this way analyzes the strengths, weaknesses, opportunities, and threats of the application [III-16].

Strengths:

- Interoperability;
- User-friendly;
- Takes into account data confidentiality;
- Decrease the human error;
- Facilitates episode distribution management to encode;
- Streamlines the coding process

Weaknesses:

- Requires internet connection.

Opportunities:

- Modernization and organizational development;
- Increasing expectation of the hospital administration to obtain methods that facilitate the hospital financing calculation;
- Provide the tool to help in the calculation of the hospital financing.

Threats:

- Lack of acceptance to resort to new technologies by health professionals.

3.3.7 Conclusion and Future Work

This paper presents a reliable solution for management of the coding process, so it is possible to follow all the steps that a coding episode goes through and facilitates the resolution of technical problems that may exist. For example, if an episode that is being coded has some data missing from the discharge report, the encoder will simply put it in a pending state detailing the problem it encountered. Since it will appear in the pending episodes tab, it will be easily found by the administrative staff, making its resolution faster and more effective.

This tool also has the advantage of not leaving any episodes forgotten, since all episodes that are discharged are immediately loaded onto the platform.

In terms of future work, the agent created to guarantee the interoperability of the system with SIMH, once it is able to read all diagnostics and coded procedures, can be adapted to be used in ETL processes to be used later in Data Mining processes.

3.3.8 References

[III-1] C. Coimbra, M. Esteves, F. Miranda, F. Portela, M. F. Santos, and J. Machado, "Improving the Codification of Hospital Discharges with an ICD-9-CM Single-page Application and its Transition to ICD-10-CM/PCS," *Int. J. Adv. Life Sci.*, vol. 10, no. 1, pp. 23–30, 2018.

[III-2] V. García Calderón *et al.*, "The impact of improving the quality of coding in the utilities of Diagnosis Related Groups system in a private healthcare institution. 14-year experience," *Int. J. Med. Inform.*, vol. 129, no. February, pp. 248–252, 2019.

[III-3] L. Chok *et al.*, "Effect of diagnosis related groups implementation on the intensive care

unit of a Swiss tertiary hospital: A cohort study," *BMC Health Serv. Res.*, vol. 18, no. 1, pp. 1–10, 2018.

[III-4] C. Lee, J. M. Kim, Y. S. Kim, and E. Shin, "The Effect of Diagnosis-Related Groups on the Shift of Medical Services From Inpatient to Outpatient Settings: A National Claims-Based Analysis," *Asia-Pacific J. Public Heal.*, vol. 31, no. 6, pp. 499–509, 2019.

[III-5] G. Hernandez-Ibarburu *et al.*, "ICD-10-PCS extension with ICD-9 procedure codes to support integrated access to clinical legacy data," *Int. J. Med. Inform.*, vol. 122, pp. 70–79, 2019.

[III-6] G. H. Utter, G. L. Cox, O. O. Atolagbe, P. L. Owens, and P. S. Romano, "Conversion of the Agency for Healthcare Research and Quality's Quality Indicators from ICD-9-CM to ICD-10-CM/PCS: The Process, Results, and Implications for Users," *Health Serv. Res.*, vol. 53, no. 5, pp. 3704–3727, 2018.

[III-7] M. Topaz, L. Shafran-Topaz, and K. H. Bowles, "ICD-9 to ICD-10: evolution, revolution, and current debates in the United States.," *Perspect. Health Inf. Manag.*, vol. 10, 2013.

[III-8] R. E. Mills, R. R. Butler, E. C. McCullough, M. Z. Bao, and R. F. Averill, "Impact of the transition to ICD-10 on medicare inpatient hospital payments," *Medicare Medicaid Res. Rev.*, vol. 1, no. 2, pp. 1–13, 2011.

[III-9] V. Alonso *et al.*, "Problems and barriers in the transition to ICD-10-CM/PCS: A qualitative study of medical coders' perceptions," *Adv. Intell. Syst. Comput.*, vol. 932, pp. 72–82, 2019.

[III-10] J. Armstrong, P. McDermott, G. R. Saade, and S. K. Srinivas, "Coding update of the SMFM definition of low risk for cesarean delivery from ICD-9-CM to ICD-10-CM," *Am. J. Obstet. Gynecol.*, vol. 217, no. 1, pp. B2-B12.e56, 2017.

[III-11] S. N. de Saúde, "ACSS apresenta o novo ecossistema sobre Morbilidade Hospitalar." [Online]. Available: http://www.acss.min-saude.pt/2018/06/04/acss-apresenta-o-novo-ecossistema-sobre-morbilidade-hospitalar/.

[III-12] P. Lee, "Design Research: What Is It and Why Do It?" [Online]. Available: https://reboot.org/2012/02/19/design-research-what-is-it-and-why-do-it/.

[III-13] V. Vaishnavi, B. Kuechler, and S. Petter, "Design Science Research in Information Systems," no. 1, pp. 1–66, 2012.

[III-14] A. B. Sergey, D. B. Alexandr, and A. T. Sergey, "Proof of Concept Center — A Promising Tool for Innovative Development at Entrepreneurial Universities," *Procedia - Soc. Behav.*

Sci., vol. 166, pp. 240–245, 2015.

[III-15] B. Schmidt, "Proof of principle studies," *Epilepsy Res.*, vol. 68, no. 1, pp. 48–52, 2006.

[III-16] R. Pereira, M. Salazar, A. Abelha, and J. Machado, "SWOT analysis of a Portuguese Electronic Health Record," *IFIP Adv. Inf. Commun. Technol.*, vol. 399, pp. 169–177, 2013.

3.4. Study IV – Building Personalized Medicine: Informed Consent

Ana Cecília Coimbra, Filipe Miranda, José Machado, António Abelha

12th International Symposium on Ambient Intelligence, 2021

Status: submission process

3.4.1 Abstract

Informed consent is an important doctor-patient communication tool, it is through them that the patient is able to understand everything that involves the procedure he is going to undergo, from results to side effects, and through that information he makes the decision to continue the procedure or not. The wide variety of procedures and problems already associated with patients make the use of a standard procedure obsolete, here the need arises to create personalized consents. This paper presents a tool developed with the aim of providing personalized informed consent to doctors in a user-friendly way.

3.4.2 Introduction

Good communication between doctor and patient has an influence on the quality of care provided, as it has repercussions on treatment, psychological function and recovery rate. Thus, for there to be a true relationship between doctor and patient, the existence of informed medical consent is essential [IV-1], [IV-2].

Anderson et al. [IV-3] point out that medical students do not feel confident in their theoretical and practical understanding of the key aspects of informed consents.

Gori et al. [IV-4] shows in their study that what the doctor thinks the patient perceives is different from what he really understands. Thus, an informed consent must contain an accurate description of the proposed procedure, the risks and benefits, advantages and disadvantages of not undergoing treatment, alternative treatments and their risks and benefits, estimated recovery time and estimated time to return to normal life [IV-2].

How can we consider whether a consent is successful or not? If the objective of the consent is the doctor is on an equal footing with the patient, we can consider from the beginning a failed consent, as it will be practically impossible to match in terms of background for the doctor and the

patient to be at the same level of professional decision. On the other hand, if the objective is to inform the patient in the best possible way so that he makes a conscious and autonomous decision, then we can consider a successful informed consent, even if the patient denies performing the procedure [IV-1].

Paterick et al. [IV-2] warns that informed consent is not feasible in all situations, during a medical emergency most patients are too ill to understand informed consent. In a study done in Swedish during episodes of acute myocardial infection 86% of the patients were not able to understand all the information given by doctors. This is where the importance of presumed consents comes in, the presumed consents are used when a patient is not able to make decisions on its own in emergency situations.

The need for personalized informed consent arises from the impossibility of having a fixed consent model since it varies depending on the procedure to be performed and the patient's comorbidity.

This paper focuses on the development of a platform created with the aim of providing personalized informed consent.

3.4.3 Research Methodologies

This platform was developed using the VueJS framework following the Design Science Rechearch methodology. VueJS was released in 2014 and is an open-source progressive JavaScript framework used to build web interfaces and single-page applications. The great advantage is the versatility of this framework, making it possible to build either very simple web apps or very complex ones, in addition to the fast-learning curve [IV-5].

In this case study, a RESTFul Web Service was used, having been developed in PHP.

3.4.3.1 Design Science Research

In terms of the development of DSR information systems it has become very popular as a research paradigm. Having been adopted by researchers in several areas, such as medical information systems, decision support systems, among others [IV-6].

The DSR methodology is based on the production of objects, also called artifacts developed with the aim of solving the problem raised [IV-7]- [IV-9].

Figure 10 shows the six steps to be followed using the DSR methodology proposed by Peffers et al. [IV-10] namely "Identify Problem", "Define Objectives of a Solution", "Design & Development", "Demonstration", "Evaluation" and "Communication".

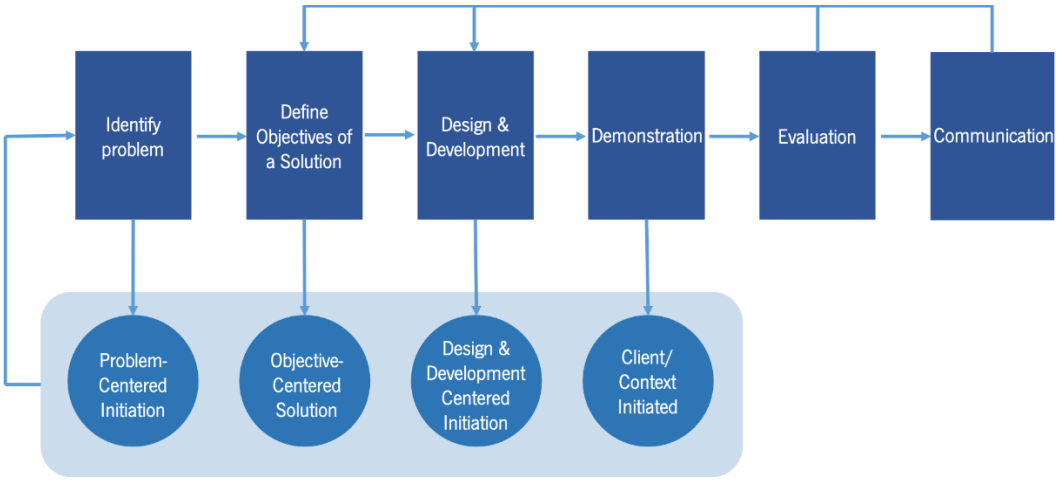


Figure 10 - DSR Model (adapted from [IV-10])

First, the problem is identified and defined, and in the next step, objectives are defined that the solution found must achieve. Having defined the problem and the objectives of the solution, step three can be started, which consists of the design and development of the artefact. Once this is developed, it must be demonstrated and evaluated and if it fulfills the objectives and solves the problem raised initially, it must then proceed to the last step and be communicated to the target audience. If the artefact, after the evaluation, does not fulfill the objectives or does not solve the initial problem, it returns to step two or three depending on the problem encountered [IV-10], [IV-11].

Thus, the use of the DSR methodology allows the development of a well-founded solution that meets the requirements of the target audience. Therefore, the case study in this paper follows the DSR methodology.

3.4.3.2 Proof of Concept

The Proof of Concept (PoC) research methodology is composed of a functional model that can be used to prove or validate a concept that has been developed through analysis or technical papers. As a result, it goes on to determine whether an idea or principle is successful and feasible, and thus capable of being used in a useful manner [IV-12].

Thus, a PoC is frequently mentioned as one of the most crucial steps in the design, development, implementation, and proposal phase of an IT solution prototype, primarily to determine if an IT solution satisfies its function, that is, if it meets the specifications and defined

objectives for which it was originally developed. Nonetheless, it enables the detection of potential vulnerabilities or errors in the built IT solution[IV-13].

To summarize, a PoC helps you to present in practice the principles, methodologies, and technology used in the development of a proposal, and thus verify the proposed solution by demonstrating its viability and utility for the intended purpose by defending its efficiency.

The proof of concept of this project is briefly described in Section 3.4.5 of this article.

3.4.4 Single-Page Application

At Centro Hospitalar Universitário do Porto (CHUP) the informed consents are universal, personalized only at the level of the type of procedure and without customization at the level of the patient. Thus, the need arose to create personalized informed consent. With this in mind, a Single Page Application (SPA) , named "*e-consentimento*", was created with the aim of making informed consents customizable, within certain standards, by the doctor who is asking for them.

The SPA is divided into two subsets, the first containing the general model of informed consent and presumed consent. In the general model, all fields are empty, so that the doctor can create his own consent.

The second subset contains consents divided by specialties and type of procedure.

Each doctor will only have access to the consent of his or her specialty, for example, the oncologist cannot issue a consent for general anesthesia.

The doctor chooses the specialty, within the available ones, that he wants to issue, chooses the procedure that will be accomplished, and a visualization of the consent is provided with the form format fill it out to facilitate the filling, and at the same time standardize the data. The first two fields "diagnosis and/or clinical description" and "description of the act/intervention" are automatic filling and follow the terminology ICD-9-CM. A list of all the ICD-9-CM codes is presented when a word is typed in these fields, the user then chooses the term and the respective ICD-9-CM code is associated with it.

The content of the rest of the form will depend on the type of consent selected. Some fields will be predefined text without the possibility of being changed, other fields will be lists that can be selected by doctors if they want it to be part of the final consent or not. For example, in terms of oncology consents, a list of possible side effects is provided, to suit the chosen treatment and the

patient, the treatments varied and where the reaction in each organism depends on the clinical conditions of each patient.

Finally, the doctor finalizes the form and presses the "accepted" button if the patient accepts consent and "not accepted" if he does not accept it, two documents are then issued one for the hospital, another for the patient to be signed with all the data entered by the doctor.

In case the patient needs a legal representative, this function is also available which the doctor can select.

3.4.5 SWOT Analysis

As a proof of concept, a SWOT analysis was carried out, thus making it possible to test the viability, the utility, the quality, and the efficiency of the application. Thus, analyzes the strengths, weaknesses, opportunities, and threats of the application.

Strengths:

- Personalized informed consents to distinct procedure of distinct specialties;
- User-friendly;
- Decrease the human error.

Weaknesses:

- Requires internet connection.

Opportunities:

- Modernization and organizational development;
- Increase the opportunity of doctors create consents taking into account the patient comorbidity;

Threats:

- Lack of acceptance to resort to new technologies by health professionals.

3.4.6 Conclusion and Future Work

This paper presents a tool developed to overcome the problem of rigid and unalterable informed consent. For example, a cancer patient, depending on the type of cancer and the diseases already associated with the patient, the side effects may be different, such as the results obtained in the execution of the treatment, hence this type of personalized consent is an asset, since the doctor can choose, for example, within the list of side effects that best suit each situation.

As for future work, there is the creation of consents associated with doctors, that is, each doctor has the model already pre-defined, without having to always choose the same options, making it faster to issue these documents and the doctor can then dedicate more of your time talking to the patient than looking at the computer.

Another proposal for future work would be the possibility for a doctor to submit a new consent proposal, if none of the existing ones is suitable, and later it will be accepted or not by someone responsible.

3.4.7 References

[IV-1] C. W. Lidz, "Informed consent: A critical part of modern medical research," *Am. J. Med. Sci.*, vol. 342, no. 4, pp. 273–275, 2011.

[IV-2] T. J. Paterick, G. V. Carson, M. C. Allen, and T. E. Paterick, "Medical informed consent: General considerations for physicians," *Mayo Clin. Proc.*, vol. 83, no. 3, pp. 313–319, 2008.

[IV-3] T. N. Anderson *et al.*, "Perception and confidence of medical students in informed consent: A core EPA," *Surg. (United States)*, vol. 167, no. 4, pp. 712–716, 2020.

[IV-4] S. Gori, M. T. Greco, C. Catania, C. Colombo, G. Apolone, and V. Zagonel, "A new informed consent form model for cancer patients: Preliminary results of a prospective study by the Italian Association of Medical Oncology (AIOM)," *Patient Educ. Couns.*, vol. 87, no. 2, pp. 243–249, 2012.

[IV-5] "The Good and the Bad of Vue.js Framework Programming." [Online]. Available: https://www.altexsoft.com/blog/engineering/pros-and-cons-of-vue-js/. [Accessed: 22-Jan-2020].

[IV-6] N. H. Thuan, A. Drechsler, and P. Antunes, "Construction of design science research questions," *Commun. Assoc. Inf. Syst.*, vol. 44, no. 1, pp. 332–363, 2019.

[IV-7] S. T. March and V. C. Storey, "Design science in the information systems discipline: An introduction to the special issue on design science research," *MIS Q. Manag. Inf. Syst.*, vol. 32, no. 4, pp. 725–730, 2008.

[IV-8] M. Bilandzic and J. Venable, "Towards Participatory Action Design Research: Adapting Action Research and Design Science Research Methods for Urban Informatics," *J.*

Community Informatics, vol. 7, no. 3, pp. 1–17, 2011.

[IV-9] A. R. Hevner, S. T. March, J. Park, and S. Ram, "Design science in information systems research," *MIS Q. Manag. Inf. Syst.*, vol. 28, no. 1, pp. 75–105, 2004.

[IV-10] K. Peffers, T. Tuunanen, M. A. Rothenberger, and S. Chatterjee, "A design science research methodology for information systems research," *J. Manag. Inf. Syst.*, vol. 24, no. 3, pp. 45–77, 2007.

[IV-11] R. Weber, *Design-science research*. Elsevier Ltd, 2018.

[IV-12] A. B. Sergey, D. B. Alexandr, and A. T. Sergey, "Proof of Concept Center — A Promising Tool for Innovative Development at Entrepreneurial Universities," *Procedia - Soc. Behav. Sci.*, vol. 166, pp. 240–245, 2015.

[IV-13] B. Schmidt, "Proof of principle studies," *Epilepsy Res.*, vol. 68, no. 1, pp. 48–52, 2006.

3.5. Study V – Medical Diagnosis Classification Using WEKA

Ana Cecília Coimbra, Filipe Miranda, José Machado and António Abelha

The 2nd International Conference on
Modern Management based on Big Data, 2021
Status: submission process

3.5.1 Abstract

The use of data mining techniques is not new—commonly it's used in various other industries, such as financial services, marketing, and manufacturing. The main goal of data mining is to find patterns in a large dataset that yield insight and expertise. Thus, in terms of healthcare, data mining methods have a wide range of uses, including diagnosing cancers, pattern recognition, and prognosticating patient health outcomes. Each patient's diagnosis at the Centro Hospitalar Universitário Universitário Porto has an ICD-10-CM code. This data can be used to build a predictive model to classify diagnosis using secondary diagnosis. Three datasets were then created to be tested using data mining techniques. As a result, the algorithm that had the best performance was the Random Tree (99.8% corrected classified instances) using the third dataset with the five main diagnoses of each patient as parameters

3.5.2 Introduction

In a medical environment, records are created on a massive scale; however, these records are frequently used primarily by health professionals to consult their patients' health records. This presents an opportunity to utilize this massive dataset to create support tools for health professionals. Thereby, data mining techniques have a diverse spectrum of applications, including diagnosing diseases, identifying patterns, and even predicting length of stay or state of health evolution of a patient [V-1–V-4].

The use of data mining techniques is not new; in fact, it is widely used in a variety of other fields, including financial institutions, marketing, manufacturing, among others. The overriding

objective of data mining is to uncover trends inside a massive data set that can be translated into relevant knowledge/information [V-1], [V-2], [V-5].

At the Centro Hospitalar Universitário do Porto (CHUP) all discharge reports are coded in the terminology ICD-10-CM, therefore, there are diagnostic records for each episode translated into this terminology.

Upon normalizing the data, they can be used in the data mining process to construct diagnostic prediction model. This is the aim of the study presented in this paper: to develop classification models for primary diagnoses using secondary diagnoses via data mining algorithms.

3.5.3 Research Methodology

The knowledge discovery in databases (KDD) process is often categorized in the following steps [V-1], [V-3]:

1. *Selection* - analyzing the database, a selection is made of the data that are relevant for the outlined objective
2. *Pre-processing* - The previously selected data is evaluated, and contradictions and missing data values are removed.
3. *Transformation* - As the name suggests, this is the process by which data is transformed. That is, the data must be structured before it can be used in the Data Mining process and thus find patterns.
4. *Data Mining* - Application of Data Mining algorithms
5. *Interpretation/Evaluation* - This phase is used to conduct an analysis and interpretation of the results. Following that, the trained Data Mining model is put to the test, with its accuracy being determined by the patterns' correct classifications. If the accuracy is less than optimal, the Data Mining model should be modified.

With regard to data mining, some of the most frequently used techniques are [V-1], [V-6]:

- *Association* - Data mining technique that is used to discover relationships between objects that are all present. These rules will assist you in forecasting one situation in relation to another. Behavioral modeling and market classification techniques are used to analyze and classify all of a customer's shopping habits and product selections. Each relationship contains laws that are multilevel, dimensional, and quantitative.
- *Classification* - Classification approaches are supervised learning methods for categorizing raw data, and supervised learning methods are used in tin data classification. There are three main

classifications available to data scientists today: decision tree, Bayesian classification, neural network, and support vector machine classification.

- *Clustering* - Used to create clusters based on similarity and to create clusters based on dissimilarity and is an unsupervised learning technique that utilizes clusters of related objects to classify them. Clustering is a widely used technique in image processing, data analysis, and pattern recognition. Linear regression, multivariate linear regression, nonlinear regression, and multivariate nonlinear regression are all forms of prediction.

3.5.3.1 Classifiers used

J48: J48 it is a simple decision tree algorithm and a supervised learning technique. The algorithm is commonly used for classification, it employs divide and conquer tactics. Reduces the entire dataset into a subset dependent on data that is already in the training dataset [V-1], [V-5], [V-7].

Random Forest: Numerous classification trees are constructed using this approach on the basis of the dataset. Tree votes are prepared according to a tree classification and classified based on a vector. classification that helps to describe the overall picture[V-1].

Support Vector Machine (SVM): It is a technique that relies on the interpretation of decision boundaries. This works to identify distinct individual instance data as belonging to different classes member objects [V-1].

Naïve Bayes: A naive Bayesian class compares an algorithm or neural network to the tree, perceptron, and network learners using rules. It implies that an attribute has a unique impact on each class [V-1], [V-5].

3.5.3.2 Metrics of performance

Several parameters will be compared in order to determine the optimal model for classifying the chosen dataset: Correctly Classified Instances, Incorrectly Classified Instances, Kappa Statistics, Mean Absolute Error, Root Mean Squared Error and time [V-1], [V-5], [V-7], [V-8].

- Correctly Classified Instances: percentage of correctly categorized data;
- Incorrectly Classified Instances: percentage of incorrect classification of data;

- Kappa Statistics: A calculation of the degree to which observers or measures of the same categorical variable agree in a nonrandom manner;
- Mean Absolute Error: average prediction error, calculated by averaging the difference between the predicted and actual values.
- Root Mean Squared Error: standard deviation of the prediction errors
- Time: time taken to train or model a dataset completely (in seconds);

3.5.3.3 Tool Used

The study presented in this paper was conducted using the WEKA software. WEKA (Waikato Environment for Knowledge Analysis) is a software package that contains a collection of machine learning algorithms for data mining tasks. Along with presenting a large collection of algorithms, it also has the advantage of making it simple to load the type of data intended for use, as these data do not have to be in a specific format, allowing for example, to load data in CSV or ARFF, among others. It also has the advantage of running on any operating system as it is written in Java [V-5], [V-7].

3.5.3.4 Dataset

Since this study's purpose is to develop a diagnostic forecast model based on other diagnoses, three similar datasets were created to determine the optimal diagnostic approach.

The first dataset was obtained from the CHUP coding platform's records; the dataset's attributes are listed in Table 2, and it contains 4322 records.

Table 2 - Attributes of the Dataset I

	Attribute	Type	Description
1	Type of episode	Nominal	Type of episode, internment, ambulatory, etc
2	provenance	Numeric	health institution where the patient comes from (can come from the institution itself)
3	Type provenance	Numeric	type of origin, urgent or scheduled
4	Destination of discharge	Numeric	destination of the patient after discharge, can be home, another institution or death
5	Length of stay	Numeric	number of days the patient has been in the hospital from the time of entry to discharge
6	Gender	Numeric	gender

7	Age	Numeric	Age
8	Diag1	Nominal	main diagnosis
9	Diag2	Nominal	secondary diagnosis associated with the patient
10	Diag3	Nominal	other diagnosis associated with the patient

The ICD-10-CM terminology codes are composed of up to seven characters, the first three of which represent the classification to which the code belongs. As a consequence, the need emerged to create a second dataset with the same attributes as the first, but with the exception of diagnostic attributes; these will contain only the first three digits of each diagnostic code, effectively generalizing these.

Consider the codes *S52.1*, *S52.2*, and *S52.3*, which denote "Fracture of the upper end of the radius", "Fracture of the shaft of the ulna", and "Fracture of the shaft of the radius", respectively. These would be represented by *S51*, which corresponds to "Fracture of forearm." While details are lost, the primary diagnosis remains, and instead of three distinct records, there are now three identical ones. This way, by lowering the degree of specificity, the percentage of the same type of diagnosis can be increased.

Table 3 - Examples of record in each dataset

Attributes	Dataset I	Dataset II	Dataset III
1 -Type of episode	INT	INT	INT
2 - provenance	0	0	0
3 - Type provenance	1	1	1
4 - Destination of discharge	114	114	114
5 - Length of stay	6	6	6
6 - Gender	2	2	2
7 - Age	9	9	9
8 - Diag1	J189	J18	J189
9- Diag2	R0689	R06	R0689
10 - Diag3	D649	D64	D649
11 - Diag4	-	-	R400
12 - Diag5	-	-	I447

The purpose of this second dataset is to determine whether data mining models perform better when diagnoses are generalized.

Once the objective of the paper is to create a model for predicting a main diagnosis through secondary diagnoses, a third dataset was created, where instead of having only the three main diagnoses, there are five main diagnoses.

Each dataset is represented by an example in the Table 3.

3.5.4 Results and Discussion

This section presents the results obtained from the data mining process for each of the algorithms chosen for each dataset.

Table 4 - Classification Results of the Dataset I

Algorithm	Correctly Classified Instances	Incorrectly Classified Instances	Kappa statistic	Mean absolute error	Root mean squared error	Time to build model (seconds)
Naïve Bayes	59.8241 %	40.1759 %	0.5803	0.0103	0.0737	0.03
SVM	20.8979 %	79.1021 %	0.1501	0.0157	0.1252	3.14
J48	71.2335 %	28.7665 %	0.7007	0.0073	0.0606	0.98
Random Tree	90.0486 %	9.9514 %	0.8968	0.0024	0.0343	1.88

The classification of dataset I is summarized in Table 4. The table demonstrates that the algorithm that produces the highest number of correctly categorized cases for dataset I is the Random Tree, which also produces the highest kappa. However, in terms of model construction time, it is not the quickest. But even so, we may consider 1.88 seconds to be quite fast.

Table 5 - Classification Results of the Dataset II

Algorithm	Correctly Classified Instances	Incorrectly Classified Instances	Kappa statistic	Mean absolute error	Root mean squared error	Time to build model
Naïve Bayes	54.2716 %	45.7284 %	0.52	0.0217	0.1122	0.01
SVM	30.8543 %	69.1457 %	0.2668	0.0277	0.1663	3.7
J48	69.8588 %	30.1412 %	0.6853	0.0156	0.0884	0.54
Random Tree	86.5085 %	13.4915 %	0.8593	0.0067	0.0575	0.46

Regarding dataset II, as represented in Table 5, the algorithm with the best results remains the Random Tree, however the proportion of correctly categorized cases is lower than the result for dataset I. With the exception of the SVM algorithm, almost all algorithms demonstrated a decline in the percentage of correctly classified cases. However, the percentage of correctly classified cases is very low.

The third dataset is the better performer since it has a greater amount of correctly categorized cases and needs fewer model construction time than the other two datasets.

Table 6 - Classification Results of the Dataset III

Algorithm	Correctly Classified Instances	Incorrectly Classified Instances	Kappa statistic	Mean absolute error	Root mean squared error	Time to build model
Naïve Bayes	76.5938 %	23.4062 %	0.7563	0.0075	0.063	0.01
SVM	21.2216 %	78.7784 %	0.1538	0.0171	0.1309	0.87
J48	76.6384 %	23.3616 %	0.7572	0.0067	0.0579	0.4
Random Tree	99.8217 %	0.1783 %	0.9982	0.0001	0.0056	0.35

Figure 11 compares the percentage of cases correctly classified by each algorithm with each dataset, allowing it simpler to understand each algorithm's efficiency. It is self-evident that the Random Tree algorithm is the best suited for the purpose of this article, having outperformed the other algorithms on all datasets.

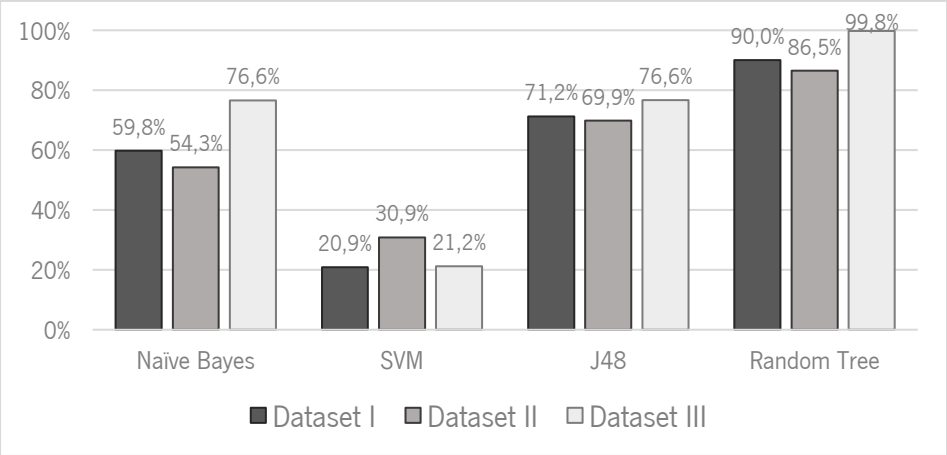


Figure 11 - Correctly Classified Instances percentage of all algorithms for each dataset

3.5.5 Conclusion and Future Work

Three similar datasets were used, with the diagnostic parameters differentiated. The first two relied on only three primary diagnoses, whereas the third relied on five primary diagnoses. The distinction between the first and second datasets is in the ICD-10-CM terminology; the first dataset included the entire diagnostic code, whereas the second only includes the first three characters.

Four algorithms were used to classify the three datasets: naive bayes, SVM, j48, and Random Tree. The Random Tree algorithm produced the best results across all parameters in all datasets.

At the moment, at CHUP, medical records of each episodes, with the exception of diagnoses, are in free text in reports; in the future, these reports will be structured, allowing for the easy use of attributes such as medication used, symptoms, and even laboratory analysis results. With this change, it will be possible to incorporate this data into the dataset, resulting in even more precise predicting models.

3.5.6 References

[V-1] N. Kumar and S. Khatri, "Implementing WEKA for medical data classification and early disease prediction," *3rd IEEE Int. Conf.* , pp. 1–6, 2017.

[V-2] H. C. Koh and G. Tan, "Data mining applications in healthcare.," *J. Healthc. Inf. Manag.*, vol. 19, no. 2, pp. 64–72, 2005.

[V-3] H. Kaur and S. K. Wasan, "Empirical Study on Applications of Data Mining Techniques in Healthcare," *J. Comput. Sci.*, vol. 2, no. 2, pp. 194–200, 2006.

[V-4] D. Tomar and S. Agarwal, "A survey on data mining approaches for healthcare," *Int. J. Bio-Science Bio-Technology*, vol. 5, no. 5, pp. 241–266, 2013.

[V-5] Sujata Joshi and S. R. P. Shetty, "Performance Analysis of Different Classification Methods in Data Mining for Diabetes Dataset Using WEKA Tool," *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 3, no. 3, pp. 1168–1173, 2015.

[V-6] N. Jothi, N. A. Rashid, and W. Husain, "Data Mining in Healthcare - A Review," *Procedia Comput. Sci.*, vol. 72, pp. 306–313, 2015.

[V-7] P. Yasodha and N. R. Ananthanarayanan, "Comparative Study of Diabetic Patient Data's Using Classification Algorithm in WEKA Tool," *Int. J. Comput. Appl. Technol. Res.*, vol. 3, no. 9, pp. 554–558, 2014.

[V-8] B. M. Patil, D. Toshniwal, and R. C. Joshi, "Predicting burn patient survivability using decision tree in WEKA environment," *2009 IEEE Int. Adv. Comput. Conf. IACC 2009*, no. March, pp. 1353–1356, 2009.

4. PUBLICATIONS

The publications published or accepted so far:

- Tiago Guimarães, Cecília Coimbra, Filipe Portela, Manuel Filipe Santos, José Machado, António Abelha, "Step towards Multiplatform Framework for Supporting Pediatric and Neonatology Care Unit Decision Process*"*, *Procedia Computer Science*, Volume 63, 2015, Pages 561-568. https://doi.org/10.1016/j.procs.2015.08.385
- Tiago Guimarães, Ana Coimbra, Simão Frutuoso and António Abelha. "A Multiplatform Decision Support Tool in Neonatology and Pediatric Care." In *Applying Business Intelligence to Clinical and Healthcare Organizations,* ed. José Machado and António Abelha, 272-283 (2016). https://doi:10.4018/978-1-4666-9882-6.ch014
- José Machado, Lucas Oliveira, Luís Barreiro, Serafim Pinto and Ana Coimbra. "Applying Soft Computing to Clinical Decision Support." In *Applying Business Intelligence to Clinical and Healthcare Organizations,* ed. José Machado and António Abelha, 256-271 (2016). https://doi:10.4018/978-1-4666-9882-6.ch013
- Ana Coimbra, Henrique Vicente, António Abelha, M. Filipe Santos, José Machado, João Neves and José Neves. "Prediction of Length of Hospital Stay in Preterm Infants a Case-Based Reasoning View" In: *Czarnowski I., Caballero A., Howlett R., Jain L. (eds) Intelligent Decision Technologies 2016. Smart Innovation, Systems and Technologies*, vol 56. Springer, Cham. https://doi.org/10.1007/978-3-319-39630-9_10
- Ana Morais, Hugo Peixoto, Cecília Coimbra, António Abelha, José Machado, "Predicting the need of Neonatal Resuscitation using Data Mining", *Procedia Computer Science*, Volume 113, 2017, Pages 571-576. https://doi.org/10.1016/j.procs.2017.08.287
- Cecília Coimbra, Marisa Esteves, Filipe Miranda, Filipe Portela, Manuel Santos, José Machado, António Abelha, "Improving Diagnosis-related Groups with a Computerized Clinical Coding Tool for ICD-9-CM Codification". In *Proceedings the Ninth International Conference on eHealth, Telemedicine, and Social Medicine*. Nice, France. IARIA, 2017
- Cecília Coimbra, Marisa Esteves, Filipe Miranda, Filipe Portela, M. F. Santos, José Machado and A. Abelha. "Improving the Codification of Hospital Discharges with an ICD-9-CM Single-page Application and its Transition to ICD-10-CM/PCS."In *Advances in life sciences* 10 (2018): 23-30.

- Daniela Oliveira, Ana Coimbra, Filipe Miranda, Nuno Abreu, Pedro Leuschner, José Machado, António Abelha, "Step Towards OpenEHR in a Portuguese Healthcare facility", in *Advances Intelligent Systems and Computing*, Springer, 2018 (Accepted)
- Bruno Fernandes, and Cecília Coimbra, and António Abelha. "Real-Time Healthcare Intelligence in Organ Transplantation: Real-Time Intelligence in Organ Transplantation." In *Next-Generation Mobile and Pervasive Healthcare Solutions.* edited by Jose Machado, António Abelha, Manuel Filipe Santos, and Filipe Portela, 128-152. Hershey, PA: IGI Global, 2018. http://doi:10.4018/978-1-5225-2851-7.ch009

5. Discussion, Conclusion and Future Work

This report presents the thesis plan for a 3-year PhD thesis in the Biomedical Engineering Doctoral Program.

The project was divided into six phases, with research questions being created as well as objectives for each one. It follows a pragmatic approach having as strategy the design science research.

5.1 Discussion and Conclusion

At the beginning of the document, some research questions were raised. This section aims to answer these questions.

- **Question 1: It is possible to implement an openEHR based Electronic Health Record (EHR) in a Portuguese major healthcare unit?**

The implementation of an openEHR based EHR is not impossible however, it is not a simple problem, in reality it is a very complex problem. It is understood that in a hospital there are thousands of records already made in the EHR in force, making the migration of existing data in this a complex case. There is also the adaptation of health professionals, who, in turn, are resistant to changes in technological terms, like to keep what they are used to working with. However, despite this, the implementation of such a system is possible, going through a consistent and defined structure from the beginning, also having a gradual adaptation with professionals, being able, for example, in the first phase to allow them to consult old records in the previous system.

- **Question 2: What is the importance of implementing this system?**

The main advantage of implementing a system based on openEHR is the structuring of data. With this structuring, all processes that feed on the EHR are benefited because they no longer deal with data irregularities, like for example a system having for example two records one saying "1000 grams" and the other "1 kilo", this makes the processing of this data completely different and more complex than if there is a structure that "forces" the user to always place with the same unit.

In addition, it will have the advantage of data interoperability since the system will have incorporated a message exchange protocol using HL7 version 3 guidelines. This ability to switch messages will serve to minimize the tendency for medical data to be geographically located.

- **Question 3: What does ICD-10-CM/PCS represent in health institutions?**

In health institutions, the ICD-10-CM/PCS allows universal coding of diagnoses and procedures, thus making it possible to understand these parameters in any language and in any part of the world. This is a very important aspect, as there is more and more progress towards the universality of databases, that is, data recorded in one institution could be consulted in other institutions anywhere in the world, thus facilitating the decision-making process, as access to clinical history is facilitated.

- **Question 4: What are the advantages of ICD-10-CM/PCS over ICD-9-CM/PCS?**

The increased accuracy of ICD-10-CM/PCS codes is a significant benefit. This was due to the disparity in the number of characters in the code between ICD-9-CM/PCS and ICD-10-CM/PCS, with ICD-9-CM/PCS having five digits and ICD-10-CM/PCS having seven.

In addition, the ICD10 terminology codes are alphanumeric. As a result, the ICD10 terminology will have a greater range and specificity of terms, making it a more detailed and useful terminology that can be used to feed other processes later.

Furthermore, the greater variety of codes allows the ICD10 vocabulary to evolve over time, keeping pace with advancements in medicine and technology.

- **Question 5: What is the role of ICD-10-CM/PCS in Diagnosis Related Groups (DRG)?**

In the DRG, the ICD-10-CM/PCS aims to assist in grouping episodes of hospitalization in the DRG, since the ICD-10-CM/PCS is a very complete coding, thus being able to be coded not only the diagnosis and procedures performed during hospitalization. The use of clinical terminology creates homogeneous records with the same or similar terms, facilitating the grouping of an episode of hospitalization in a DRG.

- **Question 6: What is the importance of DRG in health institutions?**

Correct calculation of hospital financial expenses is of utmost importance, so that health care can be the best possible. However, calculating individual resource expenditure per episode

internally would be too time-consuming. Thus, the DRG is a valuable tool in supporting this calculation, as it allows grouping episodes of hospitalization into groups of equal expenditure of resources, thus enabling a methodology that results in a number of episodes per DRG, with each DRG having an associated cost. It is thus possible to calculate the resources as closely as possible.

- **Question 7: What gaps are found in the codification process of the inpatient report?**

Although the page where each episode is coded is operational, there is a need for a management process in the background. Professionals need to understand which episodes to code, which ones are finished, which have associated problems. In this way, the management platform has come to facilitate, as the name implies, the management of the coding process, creating a functioning flow.

- **Question 8: How can you guarantee compliance with the data protection regime?**

Compliance with the data protection regime is guaranteed through the creation of different types of user, each user only has access to the information he or she must have, an administrator cannot see the patients' clinical information, and the coding doctors themselves are only able to consult the information of patients assigned to them.

- **Question 9: How can the platform interoperability with SIMH be guaranteed?**

The interoperability of the platform is achieved through agents that process the information registered in the database and send it in the format requested by SIMH. In this way, the coders can still perform the coding on the "*e-codificação*" platform, which in turn is sent to SIMH, respecting the SPMS request that coded hospital records must be in a centralized database for all hospitals.

- **Question 10: How are informed consent made available at CHUP?**

Informed consent is strictly made available at CHUP, that is, the doctor has access to a generalized document for that specialty and sometimes for the intended procedure and cannot change or adjust this consent to the reality of each patient.

• **Question 11: What is the importance of personalized informed consent in health institutions?**

Informed consent is very important in providing clinical care, as good communication and understanding between patient and doctor are the key to good treatment. However, it is impossible to create a rigid consent for each situation that may occur, since each procedure is different and each patient has their set of diseases, with different necessary care. Thus, personalized informed consents allow the creation of consents through a base where doctors can exclude or include what they find most annoying for each situation.

• **Question 12: Does "*e-consentimento*" have the capacity to arrest the user through an intuitive interface and easy learning to adapt it?**

The platform developed is a simple platform, with only one page where the doctor searches in hierarchical order, that is, first choose a specialty and then a procedure within that specialty, then a form appears with fixed and unalterable fields and with fields that he can select if he wants to include it or not, and even some fields where he can add other subjects that he deems appropriate, finally, he just has to press the "accept" button if the patient accepts the procedure, or "refuse" otherwise. If the patient needs a legal representative to sign the document with him or by him, there is a button where the doctor clicks and fills in the data of the legal representative. Finally, just print the two copies of the created document. It is a simple application, very intuitive and with direct access through the PCE without the need to register.

• **Question 13: Is the immediate implementation of "*e-consentimento*" viable?**

The implementation of "*e-consentimento*" is simple since it is an application without dependencies on others, having its own database. The only necessary process is to create a link in the PCE to access this new platform.

• **Question 14: Will using the entire ICD-10-CM code set or just the first three digits be more efficient?**

Since data mining is the process of identifying patterns in large amounts of data, the more standardized the data, the more easily these patterns will be discovered. Thus, it was critical to conduct this study in order to develop the most effective forecast models. The study used the same dataset with identical data and algorithms, differing only in the presentation of diagnostic parameters, with one dataset containing the complete diagnostic code and the other containing

only the first three digits. However, despite the fact that this difference sped up the algorithm, all other evaluation parameters performed worse, even though the difference was small. Thus, it can be concluded that categorization of diagnoses is unnecessary.

- **Question 15: Is it relevant how many diagnoses are used as parameters in the data mining process?**

Given that prediction models for a primary diagnosis were being developed using secondary diagnoses, it was critical to determine whether the number of diagnoses used made a difference in these models. Thus, a third dataset was created in which each episode was characterized according to five diagnoses. The classification results for this dataset were extremely positive, as it was the dataset with the best performance across all Data Mining algorithms. To summarize, it is critical to increase the number of parameters in the dataset in order to obtain more accurate models.

- **Question 16: It is possible to apply data mining algorithms using just the *"e-codificação"* platform's parameters?**

Despite the lack of clinical information in the datasets, such as laboratory results or patient symptoms, the classification process produced quite positive results, with one of the algorithms presenting a percentage of 99.8 percent correctly classified cases. What can be concluded is that it is possible to create prediction models for main diagnoses using only data obtained from the "*e-codificação*" platform.

Finally, it is possible to answer the main research question:

- **Question 17: How the use of evidence-based medicine, patient oriented medical records and intelligence, can improve the clinical problem lists?**

All case studies contribute to improving the list of problems. The first one aims at structuring data, with structured data facilitating any process that contributes to the creation of a list of clinical problems.

The second and third case studies are aimed at coding episodes and later distributing them to homogeneous groups. With the data for each episode encoded in terminology, all Data Mining processes become less complex, thus allowing the creation of more improved lists of clinical problems.

Regarding the fourth case study, filling in the diagnostic fields and possible complications allows creating a network of associations between diagnoses and potential problems that may result, thus allowing to feed decision support processes with relationships between diagnoses and complications/problems. In this way we were able to obtain a list of more comprehensive clinical problems.

Finally, the fifth study used data from the episodes coded in the studies two and three. It was possible to create models of diagnosis classification with high success in this study; these models could be used in the future to construct the clinical problem lists. This way, clinical problem lists may be revised or new diagnoses can be recommended based on current data ceasing to be a completely manual process.

5.2 Future Work

Although both five studies have good results when analyzed through the proof of concept, there are always some points that can be added in the future.

In the first study, this being the implementation of a new system of clinical records, a proposal that can be made is the development of a decision support tool in order to facilitate the registration of clinical data. This tool could, for example, fill the list of symptoms after the doctor has filled in the diagnostic field.

In turn, in the second and third study, the addition of a BI module that allows the analysis of various data that can be taken from discharge reports, such as which diagnoses are most used, which type of patient does the most determined procedure, if there is any correlation between groups of patients and certain diagnoses, among other parameters.

Furthermore, the creation of a Data Mining process is also a point that can be created as future work, using the coded data taken from discharge reports, it is possible through Data Mining to make predictions of the length of stay for each patient depending on the procedure performed taking into account the patient's characteristics.

Finally, regarding the fourth study, the creation of pre-defined consents for each doctor can be a point to think about, so the doctor could dedicate more of his consultation time to explaining the procedure to the patient than to be looking at the computer and choose all the fields you want to appear in the informed consent.

It is worth noting that, considering his or her flexibility, the doctor might discover that the informed consent available does not always match the standards that he or she finds desirable. It

would be beneficial to have the ability to create a consent that could later be reviewed and approved or rejected by a superior.

REFERENCES

[1] M. Rouse, "What is problem list? - Definition from WhatIs.com," 2011. [Online]. Available: http://searchhealthit.techtarget.com/definition/problem-list.

[2] J. L. Cardoso 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," *Comput. Methods Programs Biomed.*, vol. 134, pp. 267–287, 2016.

[3] M. V. Devarakonda, N. Mehta, C. H. Tsou, J. J. Liang, A. S. Nowacki, and J. E. Jelovsek, "Automated problem list generation and physicians perspective from a pilot study," *Int. J. Med. Inform.*, vol. 105, no. December 2016, pp. 121–129, 2017.

[4] C. Martínez-Costa, M. Menárguez-Tortosa, J. T. Fernández-Breis, and J. A. Maldonado, "A model-driven approach for representing clinical archetypes for Semantic Web environments," *J. Biomed. Inform.*, vol. 42, no. 1, pp. 150–164, Feb. 2009.

[5] J. L. Allones, M. Taboada, D. Martinez, R. Lozano, and M. J. Sobrido, "SNOMED CT module-driven clinical archetype management," *J. Biomed. Inform.*, vol. 46, no. 3, pp. 388–400, 2013.

[6] S. Kobayashi, E. Kimura, and K. Ishihara, "Archetype model-driven development framework for EHR web system," *Healthc. Inform. Res.*, vol. 19, no. 4, pp. 271–277, 2013.

[7] S. A. Matney, J. J. Warren, J. L. Evans, T. Y. Kim, A. Coenen, and V. A. Auld, "Development of the nursing problem list subset of SNOMED CT," *J. Biomed. Inform.*, vol. 45, no. 4, pp. 683–688, 2012.

[8] W. L. Galanter, D. B. Hier, C. Jao, and D. Sarne, "Computerized physician order entry of medications and clinical decision support can improve problem list documentation compliance," *Int. J. Med. Inform.*, vol. 79, no. 5, pp. 332–338, 2010.

[9] D. M. Hartung, J. Hunt, J. Siemienczuk, H. Miller, and D. R. Touchette, "Clinical implications of an accurate problem list on heart failure treatment," *J. Gen. Intern. Med.*, vol. 20, no. 2, pp. 143–147, 2005.

[10] R. Haux, "Health information systems - Past, present, future," *Int. J. Med. Inform.*, vol. 75, no. 3-4 SPEC. ISS., pp. 268–281, 2006.

[11] R. Lenz and M. Reichert, "IT support for healthcare processes - premises, challenges,

perspectives," *Data Knowl. Eng.*, vol. 61, no. 1, pp. 39–58, 2007.

[12] E. Ammenwerth, S. Gräber, G. Herrmann, T. Bürkle, and J. König, "Evaluation of health information systems - Problems and challenges," *Int. J. Med. Inform.*, vol. 71, no. 2–3, pp. 125–135, 2003.

[13] H. Peixoto, M. Santos, A. Abelha, and J. Machado, "Intelligence in Interoperability with AIDA," 2012, pp. 264–273.

[14] J. Machado, A. Abelha, J. Neves, and M. Santos, "Ambient intelligence in medicine," *2006 IEEE Biomed. Circuits Syst. Conf.*, pp. 94–97, 2006.

[15] L. Cardoso, F. Marins, F. Portela, M. Santos, A. Abelha, and J. Machado, "The next generation of interoperability agents in healthcare," *Int. J. Environ. Res. Public Health*, vol. 11, no. 5, pp. 5349–5371, 2014.

[16] I. H. T. S. D. Organisation, "What is snomed ct?" [Online]. Available: http://www.snomed.org/snomed-ct/what-is-snomed-ct. [Accessed: 08-Jul-2018].

[17] M. Marcos, J. A. Maldonado, B. Martínez-Salvador, D. Boscá, and M. Robles, "Interoperability of clinical decision-support systems and electronic health records using archetypes: A case study in clinical trial eligibility," *J. Biomed. Inform.*, vol. 46, no. 4, pp. 676–689, 2013.

[18] M. B. Martins, "Processo Clínico Electrónico-Levantamentos de Processos no Hospital da Prelada," 2011.

[19] W. Bonney, "Applicability of Business Intelligence in Electronic Health Record," *Procedia - Soc. Behav. Sci.*, vol. 73, pp. 257–262, 2013.

[20] M. Miranda, A. Abelha, M. Santos, J. Machado, and J. Neves, "A group decision support system for staging of cancer," *Lect. Notes Inst. Comput. Sci. Soc. Telecommun. Eng.*, vol. 1 LNICST, pp. 114–121, 2009.

[21] J. Duarte, C. F. Portela, A. Abelha, J. Machado, and M. F. Santos, "Electronic health record in dermatology service," *Commun. Comput. Inf. Sci.*, vol. 221 CCIS, no. PART 3, pp. 156–164, 2011.

[22] T. Schuler, S. Garde, S. Heard, and T. Beale, "Towards automatically generating graphical user interfaces from openEHR archetypes.," *Stud. Health Technol. Inform.*, vol. 124, no. October 2015, pp. 221–226, 2006.

[23] A. M. C. de Araujo, V. C. Times, M. U. da Silva, D. S. Alves, and S. H. C. de Santana, "Template4EHR: Building Dynamically GUIs for the Electronic Health Records Using

Archetypes," in *2016 IEEE International Conference on Computer and Information Technology (CIT)*, 2016, pp. 26–33.

[24] Z. Kazmi, "Effects of exam room EHR use on doctor-patient communication: A systematic literature review," *Inform. Prim. Care*, vol. 21, no. 1, pp. 30–39, 2013.

[25] J. Almeida, S. Frade, and R. Cruz-Correia, "Exporting Data from an openEHR Repository to Standard Formats," *Procedia Technol.*, vol. 16, pp. 1391–1396, 2014.

[26] J. Duarte, S. Castro, M. Santos, A. Abelha, and J. Machado, "Improving Quality of Electronic Health Records with SNOMED," *Procedia Technol.*, vol. 16, pp. 1342–1350, 2014.

[27] H. Wasserman and J. Wang, "An applied evaluation of SNOMED CT as a clinical vocabulary for the computerized diagnosis and problem list.," *AMIA Annu. Symp. Proc.*, vol. 2003, pp. 699–703, 2003.

[28] M. Meizoso García, J. L. Iglesias Allones, D. Martínez Hernández, and M. J. Taboada Iglesias, "Semantic similarity-based alignment between clinical archetypes and SNOMED CT: An application to observations," *Int. J. Med. Inform.*, vol. 81, no. 8, pp. 566–578, 2012.

[29] "Icd-9-cm (international classification of diseases, ninth revision, clinical modification)." [Online]. Available: http://searchhealthit.techtarget.com/definition/ICD-9-CM. [Accessed: 14-Nov-2018].

[30] "Basic difference between icd-9-cm and icd-10-cm." [Online]. Available: https://mychie.org/%0Dsystem/files/Basic_Differences_ICD-9-CM_ICD-10-CM.pdf. [Accessed: 14-Nov-2018].

[31] P. Wu *et al.*, "Mapping ICD-10 and ICD-10-CM codes to phecodes: Workflow development and initial evaluation," *J. Med. Internet Res.*, vol. 21, no. 11, pp. 1–13, 2019.

[32] J. W. Sun *et al.*, "Adaptation and Validation of the Combined Comorbidity Score for ICD-10-CM," *Med. Care*, vol. 55, no. 12, pp. 1046–1051, 2017.

[33] G. H. Utter, G. L. Cox, O. O. Atolagbe, P. L. Owens, and P. S. Romano, "Conversion of the Agency for Healthcare Research and Quality's Quality Indicators from ICD-9-CM to ICD-10-CM/PCS: The Process, Results, and Implications for Users," *Health Serv. Res.*, vol. 53, no. 5, pp. 3704–3727, 2018.

[34] M. del M. Roldán-García, M. J. García-Godoy, and J. F. Aldana-Montes, "Dione: An OWL representation of ICD-10-CM for classifying patients' diseases," *J. Biomed. Semantics*,

vol. 7, no. 1, pp. 1–16, 2016.

[35] "No Title." [Online]. Available: http://portalcodgdh.min-saude.pt/index.php/Grupos_de_%0DDiagn%7Bó%7Dsticos_Homog%7Bé%7Dneos_(GDH). [Accessed: 28-Nov-2018].

[36] "No Title." [Online]. Available: http://portalcodgdh.min-saude.pt/images/2/2e/CodificacaoClinica%25%0D26DesempenhoHospitalar.pdf. [Accessed: 28-Nov-2018].

[37] W. Quentin, D. Scheller-Kreinsen, M. Blümel, A. Geissler, and R. Busse, "Hospital payment based on diagnosis-related groups differs in Europe and holds lessons for the united states," *Health Aff.*, vol. 32, no. 4, pp. 713–723, 2013.

[38] A. Coimbra *et al.*, "Prediction of Length of Hospital Stay in Preterm Infants a Case-Based Reasoning View," in *Intelligent Decision Technologies 2016*, vol. 56, I. Czarnowski, A. M. Caballero, R. J. Howlett, and L. C. Jain, Eds. Springer International Publishing, 2016, pp. 115–128.

[39] A. Morais, H. Peixoto, C. Coimbra, A. Abelha, and J. Machado, "Predicting the need of Neonatal Resuscitation using Data Mining," *Procedia Comput. Sci.*, vol. 113, pp. 571–576, 2017.

[40] A. R. Hevner, S. T. March, J. Park, and S. Ram, "Design science in information systems research," *MIS Q. Manag. Inf. Syst.*, vol. 28, no. 1, pp. 75–105, 2004.

APPENDIX

Appendix A – Platform "*e-codificação*"

Appendix A.1 – Database from "*e-codificação*" platform

The database used by the coding platform gathers information related to the ICD-10-CM/PCS code, as well as information related to the discharge summary. The tables used are:

- ***waltas_move*** - in this table the start and end dates of a certain coding are recorded, having also stored the coding doctor's id;
- ***lista_transf_altas*** - allows monitoring the patient's movement between services during the hospitalization episode. Once, it records the dates of entry and exit of each service that the patient went through;
- ***wlist_altas*** – from this table is used information related to the patient, such as the episode, the id_high and the code of the doctor who passed a certain discharge;
- ***processos_anteriores*** - in this table, all the versions related to a certain encoding are recorded. The information for each encoding is recorded in a JavaScript Object Notation (JSON) file, the path to that file being recorded in this table, as well as the specification of the encoding version;
- ***lista_estados*** - this table contains the list of possible states;
- ***AIDA_JAVA_BO_GPDF*** - helps the generation of hospital discharge pdf;
- ***icd10cm_lista_proc*** - contains ICD-10 codes for diagnostics;
- ***icd10cm_lista_diag*** - contains ICD-10 codes for procedures.

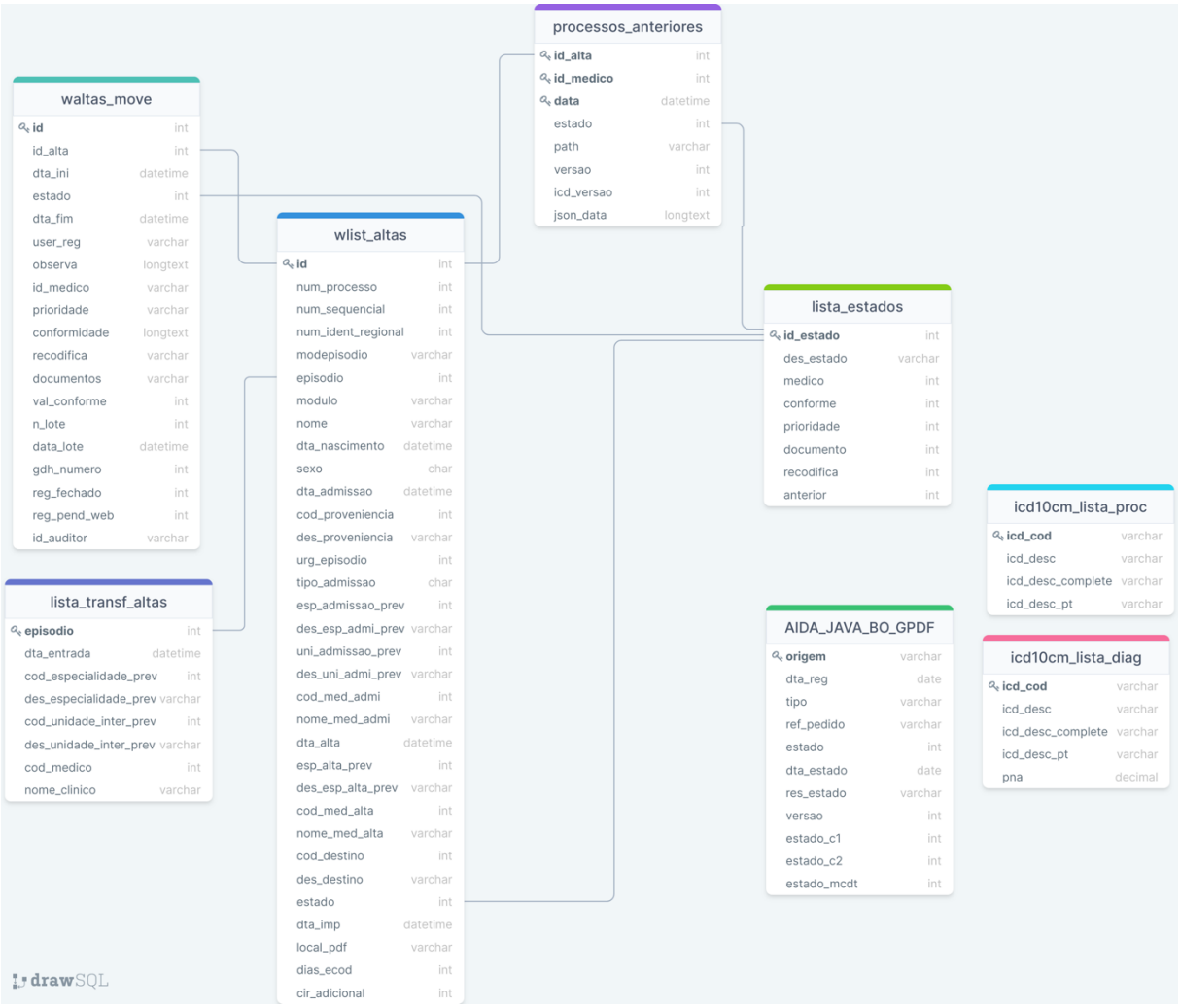


Figure 12 - "E-codificação" database

Appendix A.2 – Platform "*e-codificação*" images

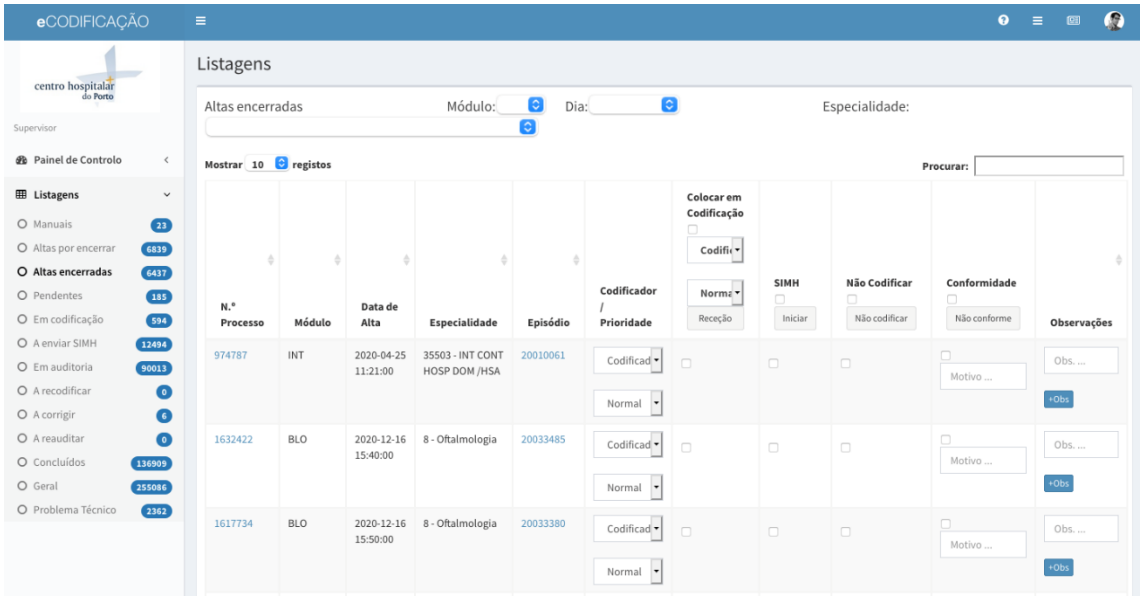


Figure 13 - "e-codificação" management platform

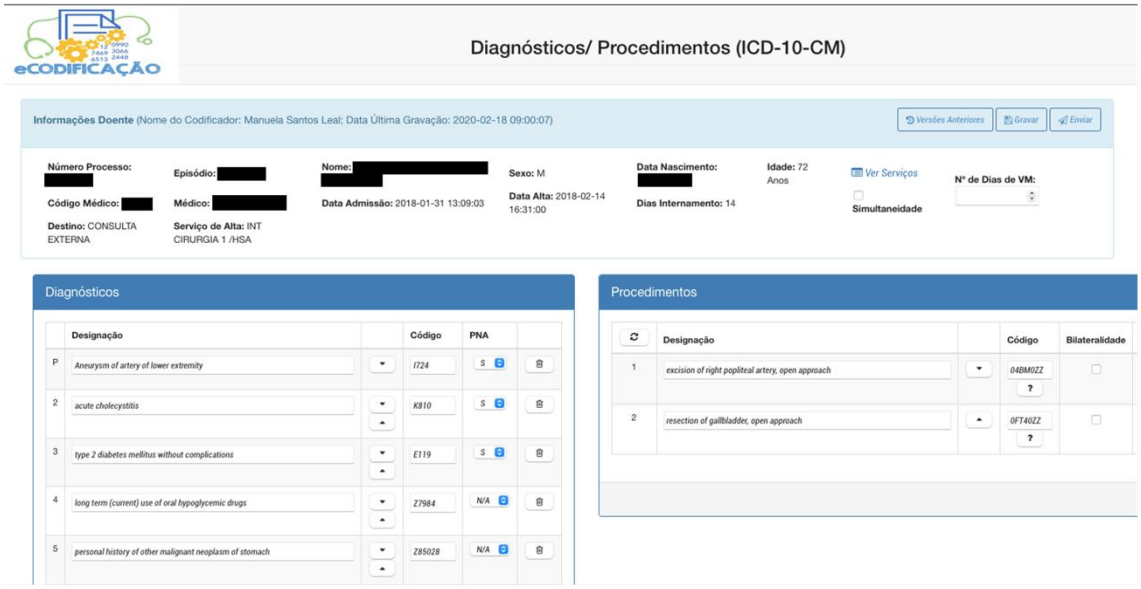


Figure 14 - "e-codificação" codification platform

Appendix A.3 – Excel file with the parameters to send episodes from "*e-codificação*" to SIMH

Table 7 - Excel file with the parameters to send episodes from "e-codificação" to SIMH

Episódio									
Árvore de campos possíveis para qualquer tipo de episódio		**Visível no: E - Envio C - Consulta**	**Tipo dados**	**Obrigatório vir preenchido por integração**	**Lista de Valores**	**Descrição**	**Regras**	**Possível de editar no SIMH ***	**Possível de editar via *WebService* Operação U - Atualização episódio**
operacao		E	String	S	N U D	Novo episódio Atualização de episódio Apagar	Não é possível realizar qualquer operação quando o episódio se encontra no estado 7 - Fechado;	N.A.	N.A.
estado		E / C	Integer		0 1 2 5 6 7 8	0 - Por Codificar 1 - Não Codificável 2 - Rascunho 5 - Em Auditoria 6 - Finalizado 7 - Fechado 8 - Rascunho SI	0: Se não preenchido (por codificar) 1: Só em modo de consulta, não são aceites no envio. 2: Indica que o episódio já tem codificação, final ou não, e depois via SIMH podem mudar o estado para finalizado; 5: Só em modo de consulta, não são aceites no envio. 6: Só em modo de consulta, não são aceites no envio. 7: Só em modo de consulta, não são aceites no envio. 8: Quando o sistema fonte altera para 1 (Por codificar) ou 2 (Rascunho) um episódio que estava nos estados 5 (Em Auditoria) ou 6 (Finalizado).	S	S
dataEstado		C	String				Formato dd-MM-yyyy HH:mm:ss	N	N (automático)
nSessao		E / C	Integer			nSessao = num da sessão associada ao episódio indicado em nEpisodio	Só se aplica para tipoEpisodio = CON - Ambulatório Médico - Consulta Externa MCDT - Ambulatório Médico - MCDT HDI - Ambulatório Médico - Hospital Dia Se o campo não estiver preenchido o número da sessão será atribuido automáticamente e será igual ao número do episodio.	N	N
dataCodificacao		E / C	String				Formato dd-MM-yyyy HH:mm:ss Só será registada se o episódio tiver o estado a 2	N	N
tipoEpisodio		E / C	String	S	INT BLO CON MCDT HDI	Internamento Cirurgia de Ambulatório Ambulatório Médico - Consulta Externa Ambulatório Médico - MCDT Ambulatório Médico - Hospital Dia		N	N

*(Todos os campos podem vir preenchidos, mas os campos marcados com 'N' não são editáveis)

Episódio									
Árvore de campos possíveis para qualquer tipo de episódio		**Visível no: E - Envio C - Consulta**	**Tipo dados**	**Obrigatório vir preenchido por integração**	**Lista de Valores**	**Descrição**	**Regras**	**Possível de editar no SIMH ***	**Possível de editar via *WebService* Operação U - Atualização episódio**
utente	nome	E / C	String	S				N	S
	nEpisodio	E / C	String	S				N	N
	nProcesso	E / C	String	S				N	N
	nUtente	E / C	Integer	S			nUtente obrigatório = S na sheet EFR No caso de recém-nascidos, se não tiver N° de utente, coloca-se o da mãe	N	S
	nBeneficiario	E / C	String	N				N	S
	sexo	E / C	Integer	S	1 2 3	Masculino Feminino Hermafrodita		N	S
	dataNascimento	E / C	String	S			formato dd-MM-yyyy HH:mm:ss (relativamente ao tempo podem enviar 00:00:00)	N	S
	nacionalidade	E / C	String	S	Códigos da *sheet* Nacionalidade			N	S
	codDistrito	E / C	String	nacionalidade = 'PT'	Códigos COD_DIST da *sheet* DistrConcFreg		2 digitos do distrito	N	S
	codConcelho	E / C	String	nacionalidade = 'PT'	Códigos COD_CONC da *sheet* DistrConcFreg		2 digitos do concelho	N	S
	codFreguesia	E / C	String	nacionalidade = 'PT'	Códigos COD_FREG da *sheet* DistrConcFreg		2 digitos da freguesia (todas a ativas + as inativadas na ultima revisão em 2013)	N	S
	codEfr	E / C	Integer	S	Códigos da *sheet* EFR			N	S

*(Todos os campos podem vir preenchidos, mas os campos marcados com 'N' não são editáveis)

Episódio									
Árvore de campos possíveis para qualquer tipo de episódio		**Visível no: E - Envio C - Consulta**	**Tipo dados**	**Obrigatório vir preenchido por integração**	**Lista de Valores**	**Descrição**	**Regras**	**Possível de editar no SIMH ***	**Possível de editar via *WebService* Operação U - Atualização episódio**
medicos	nOrdemMedCod	E / C	Integer	N				S	S
	responsavel	E / C	String	S				N	S
	nOrdemMedResp	E / C	Integer	S				N	S
	nOrdemMedAudit	E / C	Integer	N				S	S
	nMecanografico	C	String	N		Pode ser a mesma pessoa que o médico codificador. É a ultima pessoa a registar alguma alteração sobre o episódio.	O nº mecanografico obtido é o associado ao utlizador no registo do mesmo.	S	S
admissao	provenienciaDe	E / C	String		Códigos da *sheet* Unidades Saude	Se não for preenchido = 'Da própria Instituição'		N	S
	provInterna	E / C	Integer	Se tipoEpisodio = INT	1 2 3 4	Consulta Externa Hospital de Dia Cirurgia de Ambulatório Urgência	Obrigatório Se for 'Da própria Instituição' (campo provenienciaDe não preenchido)	N	S
	tipo	E / C	Integer	S	1 2	Urgente Programada	Urgente (só para tipoEpisodio = INT)	N	S
posAlta	destino	E / C	Integer	S		Ver códigos na *sheet* Destino Após Alta		N	S
	transferidoPara	E / C	String			Ver códigos na *sheet* Unidades Saude	Obrigatoriedade de preenchimento conforme sheet Destino Após Alta. O código "000000" - "Desconhecido", só se aplica quando o destino após alta está preenchido com o código "213" - "Internamento noutro Hospital";	N	S
	motivo	E / C	Integer		10 20	Para seguimento Por falta de recursos	Regras de Obrigatoriedade na sheet Destino Após Alta	N	S

*(Todos os campos podem vir preenchidos, mas os campos marcados com 'N' não são editáveis)

Episódio									
Árvore de campos possíveis para qualquer tipo de episódio		**Visível no: E - Envio C - Consulta**	**Tipo dados**	**Obrigatório vir preenchido por integração**	**Lista de Valores**	**Descrição**	**Regras**	**Possível de editar no SIMH ***	**Possível de editar via *WebService* Operação U - Atualização episódio**
cirurgia	dataInicio	E / C	String				Formato dd-MM-yyyy HH:mm:ss; Obrigatório se existir algum dos códigos da *sheet* Intervenção Cirúrgica	N	S
	dataFim	E / C	String						
	tipo	E / C	String	S	1 2 3 4 5	GPRSNSCIR - MRC GPRSNSCIR - MRA PROGRAMADA BASE PROGRAMADA ADICIONAL URGENTE		N	S
servicos	index	E / C	Integer	S	1,2,3,4,5,6	Valor da ordem do serviço	O valor 6 é entendido como o ultimo serviço	N	S
	descricao	E / C	String	S			Descrição interna do Hospital	N	S
	codigo	E / C	String	S			Código interno do Hospital	N	S
	codEspecialidade	E / C	String	S			Deve respeitar o formato indicado na Circular Normativa da ACSS N° 20/2015/DPS/ACSS	N	S
	dataAdmissao	E / C	String	S			formato dd-MM-yyyy HH:mm:ss	N	S
	dataAlta	E / C	String	S			formato dd-MM-yyyy HH:mm:ss	N	S
Devem ser colocados os serviços num máximo de 6, se existirem mais do que 6, registam os 5 primeiros e o ultimo.									

*(Todos os campos podem vir preenchidos, mas os campos marcados com 'N' não são editáveis)

Episódio									
Árvore de campos possíveis para qualquer tipo de episódio		**Visível no: E - Envio C - Consulta**	**Tipo dados**	**Obrigatório vir preenchido por integração**	**Lista de Valores**	**Descrição**	**Regras**	**Possível de editar no SIMH ***	**Possível de editar via *WebService* Operação U - Atualização episódio**
outrosDados	pesoNascenca	E / C	Integer		100 <= Peso <= 9000 (valores em gramas)	Peso à nascença	Obrigatório se existir algum dos códigos da sheet Peso à Nascença	S (só quando IDADE < 29 dias)	S
	semanaGestacao	E / C	Integer		1 <= Semanas <= 45	Número de semanas de gestação	Obrigatório se existir algum dos códigos da *sheet* Semanas de Gestação		
	diasVentlMecInv	E / C	Integer			Dias de ventilação mecânica invasiva	Obrigatório consoante as regras da *sheet* Dias de Ventilação Mecânica	S (só em episódios de INT)	S
	diasCuidInt	E / C	Integer			N° dias nos cuidados intensivos	Obrigatório se existir algum serviço associado ao serviço indicado na *sheet* Dias Cuidados Int	S (só em episódios de INT)	S
diagnosticos	index	C	Integer			Valor da ordem do diagnostico	Obrigatório se codigo preenchido	S	N
	codigo	C	String				Código de diagnostico válido para a versão do ICD em causa	S	N
	pna	C	String		S N D I NA	Sim Não Desconhecido Indeterminado clinicamente Isento de registo de PNA	Só obrigatório em Internamento (tipoEpisodio = INT)	S	N
Devem ser acrescentados os restantes registos respeitando a estrutura acima									
procedimentos	index	C	Integer			Valor da ordem do procedimento, só para controlo do envio	Obrigatório se codigo preenchido	S	N
	codigo	C	String				Código do procedimento válido para a versão do ICD em causa	S	N
Devem ser acescentados os restantes registos respeitando a estrutura acima									

*(Todos os campos podem vir preenchidos, mas os campos marcados com 'N' não são editáveis)

Table 8 - Auxiliar Table - Nationality

CODIGO	DESC_PAIS
MS	Monserrate
MT	Malta
MU	Maurícias
MV	Maldivas
MW	Malawi
MX	México
MY	Malásia
MZ	Moçambique
NA	Namíbia
YE	Iémen
YT	Mayotte
ZA	África do Sul
ZM	Zâmbia
ZW	Zimbabwe
VG	Ilhas Virgens (Britânicas)
VI	Ilhas Virgens (Estados Unidos)
VN	Vietname
VU	Vanuatu
WF	Wallis e Futuna (Ilhas)
WS	Samoa
XX	Nascido a Bordo
DD	República Democrática Alemã
GG	Guernsey

CODIGO	DESC_PAIS
IM	Ilha de Man
JE	Jersey
PZ	Zona do Canal do Panamá
QT	Índias Ocidentais
SU	União Soviética
TP	Timor
YD	Iémen do Sul
YU	Jugoslávia
ZR	Zaire
AA	Bombaim-União Indiana
RS	Sérvia
ME	Montenegro
XK	Kosovo
CX	Ilha Christmas
CY	Chipre
CZ	República Checa
DE	Alemanha
DJ	Jibuti
DK	Dinamarca
DM	Domínica
DO	República Dominicana
DZ	Argélia
EC	Equador

CODIGO	DESC_PAIS
EE	Estónia
EG	Egipto
EH	Sara Ocidental
ER	Eritreia
ES	Espanha
ET	Etiópia
FI	Finlândia
FJ	Ilhas Fiji
FK	Ilhas Falkland (Malvinas)
FM	Micronésia (Estados Federados da)
FO	Ilhas Faroé
FR	França
GA	Gabão
GB	Reino Unido
GD	Granada
GE	Geórgia
AD	Andorra
AE	Emiratos Árabes Unidos
AF	Afeganistão
AG	Antigua e Barbuda
AI	Anguila
AL	Albânia
AM	Arménia

CODIGO	DESC_PAIS
AN	Antilhas Holandesas
AO	Angola
AQ	Antárctica
AR	Argentina
AS	Samoa Americana
AT	Áustria
AU	Austrália
AW	Aruba
AX	Ilhas Aland
AZ	Azerbaijão
BA	Bósnia-Herzegovina
BB	Barbados
BD	Bangladesh
BE	Bélgica
BF	Burkina Faso
BG	Bulgária
BH	Barém
BI	Burundi
BJ	Benim
BM	Bermudas
BN	Brunei Darussalam
BO	Bolívia
BR	Brasil

CODIGO	DESC_PAIS
BS	Bahamas
BT	Butão
BV	Ilha Bouvet
BW	Botswana
BY	Bielorússia
BZ	Belize
CA	Canadá
CC	Ilhas Cocos (Keeling)
CD	Congo (República Democrática do)
CF	Centro-Africana (República)
CG	Congo
CH	Suiça
CI	Costa do Marfim
CK	Ilhas Cook
CL	Chile
CM	Camarões
CN	China
CO	Colômbia
CR	Costa Rica
JP	Japão
KE	Quénia
KG	Quirguizistão
KH	Camboja

CODIGO	DESC_PAIS
KI	Kiribati
KM	Comores
KN	São Cristóvão e Nevis
KP	Coreia, República Popular Democrática da
KR	Coreia, República da
KW	Kuwait
KY	Ilhas Caimão
KZ	Cazaquistão
LA	Laos (República Democrática Popular do)
LB	Líbano
LC	Santa Lúcia
LI	Liechtenstein
LK	Sri Lanka
LR	Libéria
LS	Lesoto
LT	Lituânia
LU	Luxemburgo
LV	Letónia
LY	Líbia (Jamahiriya Árabe da)
MA	Marrocos
MC	Mónaco
MD	Moldova (República de)
MG	Madagáscar

CODIGO	DESC_PAIS
MH	Ilhas Marshall
MK	Macedónia (Antiga República Jugoslava da)
ML	Mali
MM	Myanmar
MN	Mongólia
MO	Macau
MP	Ilhas Marianas do Norte
MQ	Martinica
MR	Mauritânia
CS	Sérvia e Montenegro
CU	Cuba
CV	Cabo Verde
NC	Nova Caledónia
NE	Niger
NF	Ilha Norfolk
NG	Nigéria
NI	Nicarágua
NL	Países Baixos
NO	Noruega
NP	Nepal
NR	Nauru
NU	Niue
NZ	Nova Zelândia

CODIGO	DESC_PAIS
PE	Peru
PF	Polinésia Francesa
PG	Papuásia-Nova Guiné
PH	Filipinas
PK	Paquistão
PL	Polónia
PM	São Pedro e Miquelon
PN	Pitcairn
PR	Porto Rico
PS	Território Palestiniano Ocupado
PT	Portugal
PW	Palau
PY	Paraguai
QA	Catar
RE	Reunião
RO	Roménia
RU	Rússia (Federação da)
RW	Ruanda
SA	Arábia Saudita
SB	Ilhas Salomão
SC	Seychelles
PE	Peru
PF	Polinésia Francesa

CODIGO	DESC_PAIS
SD	Sudão
SE	Suécia
SG	Singapura
SH	Santa Helena
SI	Eslovénia
SJ	Svaldbard e a Ilha de Jan Mayen
SK	Eslováquia
SL	Serra Leoa
SM	San Marino
SN	Senegal
SO	Somália
SR	Suriname
ST	São Tomé e Príncipe
SV	El Salvador
SY	Síria (República Árabe da)
SZ	Suazilândia
TC	Ilhas Turcas e Caicos
TD	Chade
TF	Territórios Franceses do Sul
TG	Togo
TH	Tailândia
TJ	Tajiquistão
TK	Tokelau

CODIGO	DESC_PAIS
TL	Timor Leste
TM	Turquemenistão
TN	Tunísia
TO	Tonga
TR	Turquia
TT	Trindade e Tobago
TV	Tuvalu
TW	Taiwan (Província da China)
TZ	Tanzânia, República Unida da
UA	Ucrânia
UG	Uganda
UM	Ilhas Menores Distantes dos Estados Unidos
US	Estados Unidos
UY	Uruguai
UZ	Usbequistão
VA	Santa Sé (Cidade Estado do Vaticano)
VC	São Vicente e Granadinas
VE	Venezuela
GF	Guiana Francesa
GH	Gana
GI	Gibraltar
GL	Gronelândia
GM	Gâmbia

CODIGO	DESC_PAIS
GN	Guiné
GP	Guadalupe
GQ	Guiné Equatorial
GR	Grécia
GS	Geórgia do Sul e Ilhas Sandwich do Sul
GT	Guatemala
GU	Guam
GW	Guiné-Bissau
GY	Guiana
HK	Hong-Kong
HM	Ilha Heard e Ilhas Mcdonald
HN	Honduras
HR	Croácia
HT	Haiti
HU	Hungria
ID	Indonésia
IE	Irlanda
IL	Israel
IN	Índia
IO	Território Britânico do Oceano Índico
IQ	Iraque
IR	Irão (República Islâmica)
IS	Islândia

CODIGO	DESC_PAIS
IT	Itália
JM	Jamaica
JO	Jordânia

Table 9 - Auxiliar Table - EFR

Código	Descrição	Num Utente Obrigatório
935601	SERVICO NACIONAL DE SAÚDE	S
935610	Migrante RESIDENTE COM N.º DE UTENTE	S
935612	ACORDO - BRASIL	N
935618	CONVENCãO - CABO VERDE (BOLSEIROS)	N
935619	EVACUADOS ANGOLA	N
935620	EVACUADOS CABO-VERDE	N
935621	EVACUADOS GUINE-BISSAU	N
935622	EVACUADOS S. TOME E PRINCIPE	N
935623	EVACUADOS MOCAMBIQUE	N
935624	EM SITUACãO IRREGULAR - CUIDADOS URGENTES E VITAIS	N
935625	MENOR EM SITUACÃO IRREGULAR	N
935626	REQUERENTE DE ASILO OU ESTATUTO DE REFUGIADO	N
935628	NACIONAIS DA NORUEGA, DINAMARCA E REINO UNIDO	N
935640	ADSE-SNS	S
935641	SAD-GNR-SNS	S
935642	SAD-PSP-SNS	S
935643	IASFA-SNS	S
935631	Imprensa Nacional Casa da Moeda	S
935629	SNS -Inscritos Açores	S
935630	SNS-Inscritos Madeira	S
935611	Regulamento CE - Doc Portátil S2	N
935613	Convenção - Andorra	N
935614	Convenção - Cabo Verde	N
935615	Convenção - Quebec	N
935616	Convenção - Marrocos	N
935617	Convenção - Tunísia	N
999991	Seguradoras	N
999992	SAMS	N
999993	Hospitais SNS	N
999994	Hospitais não SNS	N
999995	Outros	N
999996	Independentes	N
935008	Hospital Santo Espirito EPE	N
935007	Hospital Divino Espirito Santo EPE.	N
935003	Unidade Saude Ilha Flores	N
935010	Unidade Saude da Ilha Graciosa	N
935012	Unidade Saude Ilha Corvo	N
935014	Unidade Saude da Ilha do Pico	N
935024	Unidade Saude Ilha São Jorge	N
935028	Unidade Saude Ilha Faial	N
935029	Unidade Saude Ilha Santa Maria	N
935030	Unidade Saude Ilha Terceira	N
935001	ARS Norte	N
935002	ARS Centro	N
935006	ARS LVT	N
935004	ARS Alentejo	N
935005	ARS Algarve	N
935009	Hospital da Horta, EPE	N

Table 10 - Auxiliar Table - Destino Apos Alta

Descrição	Codigo	Obriga ao registo do motivo de transferência	Permite registo do motivo de transferência	Obriga ao registo do Transferido Para	Permite registo do Transferido Para
PARA O DOMICILIO	100	N	N	N	N
Hospital Dia	111	N	N	N	N
Consulta Externa do Hospital	112	N	N	N	N
Consulta Externa de Outro Hospital	113	N	N	N	N
Centro de Saude / Medico de Familia	114	N	S	N	S
Para outra instituição sem internamento	200	N	S	N	S
Internamento em Hospital do SNS	211	S	S	S	S
Centro de Saude (com internamento)	212	S	S	N	S
Internamento noutro Hospital	213	S	S	S	S
SERVICO DOMICILIARIO	300	N	N	N	N
SAIDO CONTRA PARECER MEDICO	400	N	N	N	N
FALECIDO	500	N	N	N	N
Cuidados continuados integrados	600	S	S	N	S
Unidade de Convalescenca	611	S	S	N	S
Unidade de media duracão e reabilitação	612	S	S	N	S
Unidade de longa duracão e manutenção	613	S	S	N	S
Unidade de cuidados paliativos	614	S	S	N	S
Unidade de AVC	615	S	S	N	S
Unidade C.C.Int. de Saúde Mental	616	S	S	N	S
Unidade C.C.Int. sem outra especificação	617	S	S	N	S

Table 11 - Auxiliar Table - Peso à nascença

ICD-10-CM (target)	Long_Descp ICD-10-CM
Z3800	Single liveborn infant, delivered vaginally
Z3801	Single liveborn infant, delivered by cesarean
Z381	Single liveborn infant, born outside hospital
Z382	Single liveborn infant, unspecified as to place of birth
Z3830	Twin liveborn infant, delivered vaginally
Z3831	Twin liveborn infant, delivered by cesarean
Z384	Twin liveborn infant, born outside hospital
Z385	Twin liveborn infant, unspecified as to place of birth
Z3861	Triplet liveborn infant, delivered vaginally
Z3862	Triplet liveborn infant, delivered by cesarean
Z3863	Quadruplet liveborn infant, delivered vaginally
Z3864	Quadruplet liveborn infant, delivered by cesarean
Z3865	Quintuplet liveborn infant, delivered vaginally
Z3866	Quintuplet liveborn infant, delivered by cesarean
Z3868	Other multiple liveborn infant, delivered vaginally
Z3869	Other multiple liveborn infant, delivered by cesarean
Z387	Other multiple liveborn infant, born outside hospital
Z388	Other multiple liveborn infant, unsp as to place of birth

Table 12 - Auxiliar Table - Semanas de Gestação

ICD-10-CM (target)	Long_Descp ICD-10-CM
Z370	Single live birth
Z371	Single stillbirth
Z372	Twins, both liveborn
Z373	Twins, one liveborn and one stillborn
Z374	Twins, both stillborn
Z3750	Multiple births, unspecified, all liveborn
Z3751	Triplets, all liveborn
Z3752	Quadruplets, all liveborn
Z3753	Quintuplets, all liveborn
Z3754	Sextuplets, all liveborn
Z3759	Other multiple births, all liveborn
Z3760	Multiple births, unspecified, some liveborn
Z3761	Triplets, some liveborn
Z3762	Quadruplets, some liveborn
Z3763	Quintuplets, some liveborn
Z3764	Sextuplets, some liveborn
Z3769	Other multiple births, some liveborn
Z377	Other multiple births, all stillborn
Z379	Outcome of delivery, unspecified
Z3800	Single liveborn infant, delivered vaginally
Z3801	Single liveborn infant, delivered by cesarean
Z381	Single liveborn infant, born outside hospital
Z382	Single liveborn infant, unspecified as to place of birth
Z3830	Twin liveborn infant, delivered vaginally
Z3831	Twin liveborn infant, delivered by cesarean
Z384	Twin liveborn infant, born outside hospital
Z385	Twin liveborn infant, unspecified as to place of birth
Z3861	Triplet liveborn infant, delivered vaginally
Z3862	Triplet liveborn infant, delivered by cesarean
Z3863	Quadruplet liveborn infant, delivered vaginally
Z3864	Quadruplet liveborn infant, delivered by cesarean
Z3865	Quintuplet liveborn infant, delivered vaginally
Z3866	Quintuplet liveborn infant, delivered by cesarean
Z3868	Other multiple liveborn infant, delivered vaginally
Z3869	Other multiple liveborn infant, delivered by cesarean
Z387	Other multiple liveborn infant, born outside hospital
Z388	Other multiple liveborn infant, unsp as to place of birth

Table 13 - Auxiliar Table - Dias Ventilação Mecânica

Procedimento	Designação	Campo DV	Tempo de internamento	Notas
Não preenchido	NA	Indiferente	Indiferente	Não pode ter nenhum dos seguintes procedimentos 5A1935Z 5A1945Z 5A1955Z
5A1935Z	Respiratory Ventilation, Less than 24 Consecutive Hours	=< 24 horas E diferente de não preenchido	Indiferente	Validação entre Procedimento e DV e vice versa
5A1945Z	Respiratory Ventilation, 24-96 Consecutive Hours	> 24 horas e < 4 dias E diferente de não preenchido	Indiferente	Validação entre Procedimento e DVM e vice versa
5A1955Z	Respiratory Ventilation, Greater than 96 Consecutive Hours	>= 4 dias E diferente de não preenchido	>= 4 dias	Validação entre Procedimento e DVM e vice versa

Table 14 - Auxiliar Table - Dias cuidados intensivos

Código	Descrição
1106	UNIDADES DE CUIDADOS INTENSIVOS

Appendix B – Platform "e-consentimento"

Appendix B.1 – Database from "e-consentimento"

The database used by the informed consent platform gathers information related to the ICD-10-CM/PCS code, as well as information related to the discharge summary. The tables used are:

- ***icd9cm_lista*** – contains the respective ICD-9-CM codes for diagnostics and procedures;
- ***prefeitos*** – in this table, the consent models informed by specialties and by theme are organized;
- ***assinadosecon*** – contains the history of informed consents issued, making it possible to view previous consents later;
- ***especialidades*** - all specialties are organized;
- ***utilizadores*** – list of users, as well as their respective characteristics, such as specialties, thus making it possible to limit the issuing of informed consents from each user to their specialty only.

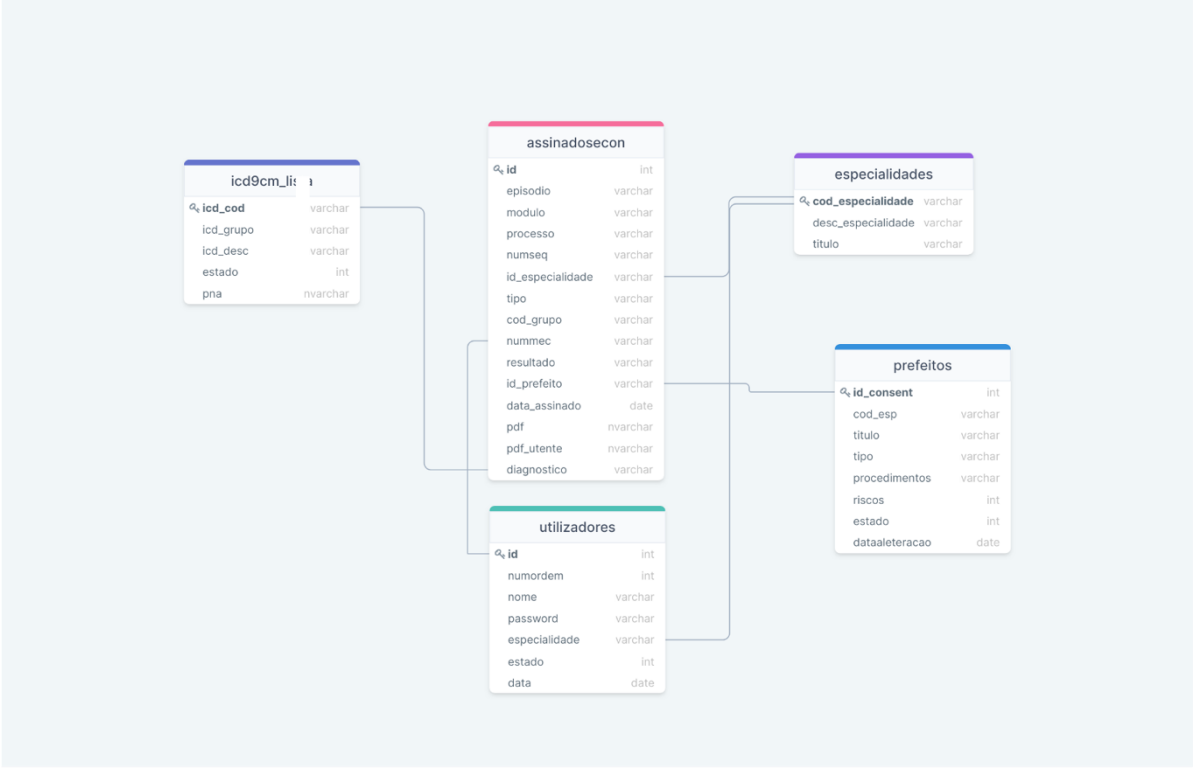


Figure 15 - "E-consentimento" database

Appendix B.2 – Platform "*e-consentimento*" images

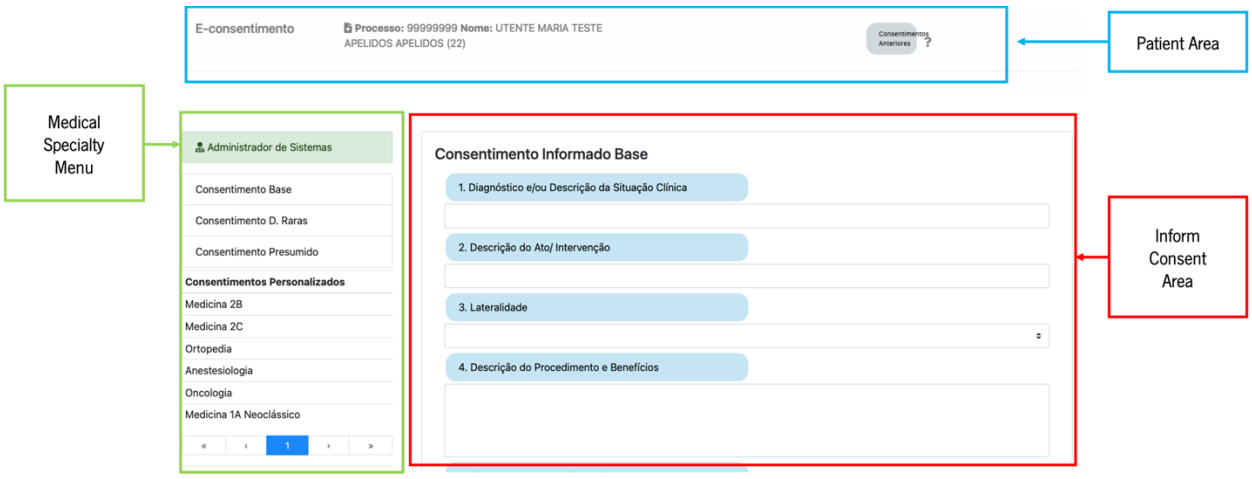


Figure 16 - "E-consentimento" general view

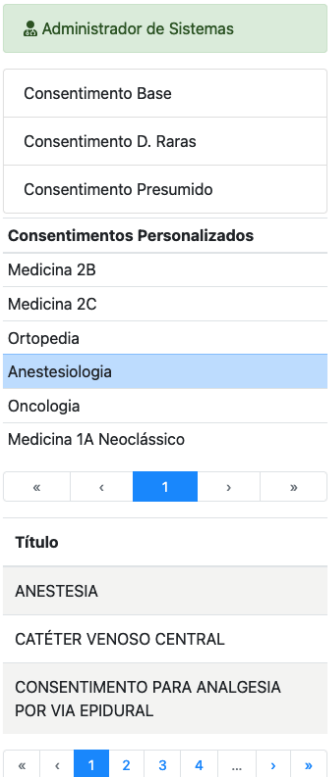


Figure 17 - "E-consentimento" medical speciality menu

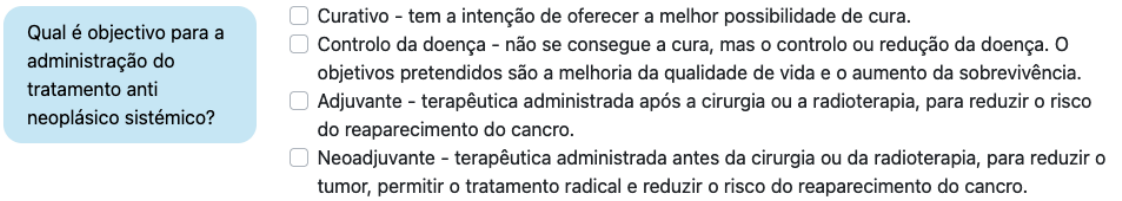


Figure 18 - "E-consentimento" select option

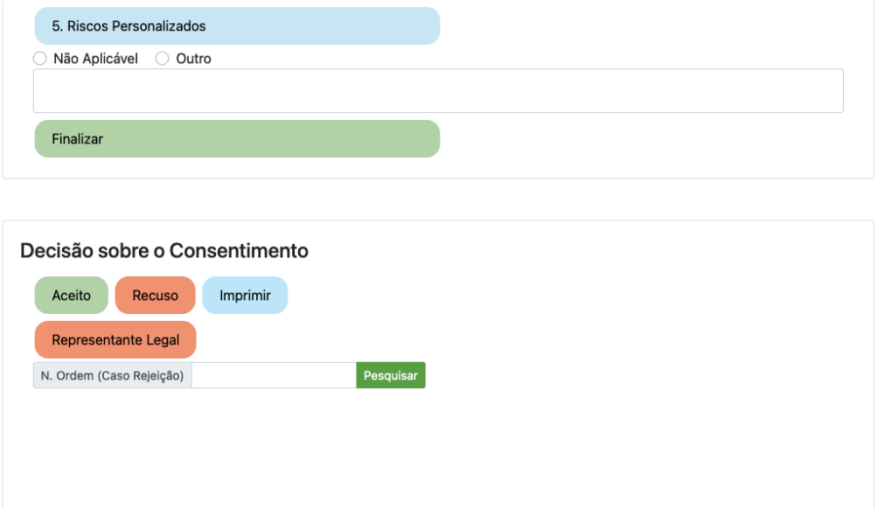


Figure 19 - "E-consentimento" final steps