Uminho | 2016
Bruno Daniel Pereira Fernandes
Real-Time Healthcare Intelligence in Organ Transplantation

Universidade do Minho
Escola de Engenharia

Bruno Daniel Pereira Fernandes

Real-Time Healthcare Intelligence in Organ Transplantation

Outubro de 2016

Universidade do Minho
Escola de Engenharia

Bruno Daniel Pereira Fernandes

Real-Time Healthcare Intelligence in Organ Transplantation

Dissertação de Mestrado
Mestrado Integrado em Engenharia Biomédica
Ramo de Informática Médica

Trabalho efetuado sob a orientação do
Professor Doutor António Carlos da Silva Abelha

Outubro de 2016

Declaração

Nome: Bruno Daniel Pereira Fernandes

Endereço eletrónico: bdpf5@hotmail.com

Cartão de Cidadão: 14370076

Título da Dissertação: Real-Time Healthcare Intelligence in Organ Transplantation

Orientador: Professor Doutor António Carlos da Silva Abelha

Coorientador: Professor Doutor José Carlos Ferreira Maia Neves

Ano de conclusão: 2016

Designação do Mestrado: Mestrado Integrado em Engenharia Biomédica

Ramo: Informática Médica

Universidade do Minho, ___ / ___ / ___

Assinatura: ____

Acknowledgements

If not for some very influential people, this dissertation would never have happened.
First of all, I would like to thank my supervisor, Prof. Dr. António Abelha, for his availability, encouragement, and guidance provided through the development of such an important project as this dissertation is, for me.
Similarly, I would like to thank Prof. Dr. José Machado for his confidence and patience over the year. I want to express my most sincere gratitude for every opportunity and knowledge provided, indispensable to expand my experience, but above all, for giving me the necessary freedom to explore and learn from my decisions.
I am indebted to all the clinicians that constitute the Gabinete Coordenador de Colheita e Transplantação at Centro Hospitalar do Porto, Hospital de Santo António, EPE who kindly shared their valuable time, insights and knowledge and without whom this dissertation would not have been concluded. At each stage of this project, whether in the early beginning or in the final phase, every meeting had its relevance in discussing and providing valuable inputs so the project could move forward.
I would also like to thank the Information Systems Service, also at Centro Hospitalar do Porto, Hospital de Santo António, EPE, for the opportunity to develop such a crucial system in a real healthcare context and for the moments of cheerfulness making it a place worth working on.
To my friends with whom I shared a lot of stories and adventures. For all their support and advice, often disguised as silly talks, I am truly grateful.
Finally, I owe a big thank you to my family for showing me that the simplest things in life are often the truest. I am especially grateful to my parents, for encouraging me in all of my pursuits and inspiring me to follow my dreams.
I am genuinely grateful to all of you!

Abstract

Organ transplantation is the best and often the only treatment for patients with end-stage organ failure. However, the universal shortage of deceased donors and the international variation in donation and transplantation activities result in a worrying situation that must be addressed. As in most countries, Portugal has implemented donation programs to answer the increasing need for transplants with the objective to identify all the possible and potential donors admitted to hospitals. These donors constitute the largest share of organ donors in Portugal, but identifying a patient that may progress to brain death could be a complex task and cadaveric organs must be transplanted in a short period of time in order to achieve satisfactory results.

Therefore, the urgent need of intelligent solutions that are able to support the decision-making process is crucial in critical areas as the organ transplantation is.

The aim of this dissertation is firstly the knowledge acquisition on the potential organ donor criteria for further detection and secondly the design and implementation of a software platform to assist the inefficient process of identification of potential organ donors. This will result in an increase of control of the screening method and consequently optimize the workflow of the pre-transplantation process.

Accordingly, and after several meetings with the transplant team, a prior identification pattern was structured and used to characterize the development of the proposed solution, named Organite. Organite is defined as a system to support the transplantation process, based on Business Intelligence technologies. It is responsible for the collection, management, storage, and signaling of potential organ donors using information from the disparate Health Information Systems to provide real-time tracking of patients and optimize the transplant team's workflow.

The developed platform is currently implemented at Centro Hospitalar do Porto, Hospital de Santo António, EPE and displays a steady and competent behavior providing consequently a way to have more control of the information needed for the decision-making process. As a result, the number of transplantation records at Centro Hospitalar do Porto, Hospital de Santo António, EPE are expected to show more profitable outcomes.

Keywords: *Decision-Making Process, Organ Transplantation, Potential Organ Donors Detection*

Resumo

O transplante de órgãos é a melhor e muitas vezes a única forma de tratamento para pacientes com casos de insuficiência terminal de órgãos. No entanto, a escassez universal de dadores falecidos e a variação internacional nas atividades de doação e transplantação resultam numa situação preocupante que deve ser abordada.

Tal como em muitos países, Portugal tem implementado programas de doação para responder à crescente necessidade de transplantes com o objetivo de identificar todos os possíveis e potenciais dadores em morte cerebral internados em hospitais. Estes dadores em morte cerebral constituem a maior parcela dos dadores de órgãos em Portugal. Contudo, identificar um paciente que pode evoluir para morte cerebral pode ser uma tarefa complexa e a transplantação de órgãos de cadáveres deve ocorrer num curto período de tempo a fim de alcançar resultados satisfatórios.

Portanto, a necessidade urgente de soluções inteligentes que sejam capazes de apoiar o processo de tomada de decisões é crucial em áreas tão críticas como a transplantação de órgãos é.

O objetivo desta dissertação é, em primeiro lugar, a aquisição de conhecimento sobre os critérios de decisão de um potencial dador de órgãos para posterior deteção e, em segundo lugar, a conceção e implementação de uma plataforma de software para auxiliar o processo até agora ineficiente de identificação de potenciais dadores de órgãos. Isto deverá permitir ter mais controlo sobre o método de deteção e, consequentemente, otimizar o fluxo de trabalho do processo de pré-transplantação.

Por conseguinte, e após várias reuniões com a equipa do Gabinete Coordenador de Colheita e Transplantação, um padrão de identificação prévio foi estruturado e utilizado para caracterizar o desenvolvimento da solução proposta, chamada Organite. A plataforma Organite é definida como um sistema para suportar o processo de transplantação, baseado em tecnologias de Business Intelligence. É, assim, responsável pela recolha, gestão, armazenamento e sinalização de potenciais dadores recorrendo a dados provenientes de vários Sistemas de Informação Hospitalar para oferecer acompanhamento dos pacientes em tempo real e otimizar o fluxo de trabalho da equipa do Gabinete Coordenador de Colheita e Transplantação.

A plataforma assim desenvolvida está implementada atualmente no Centro Hospitalar do Porto, Hospital de Santo António, EPE e apresenta um comportamento estável e eficiente proporcionando, consequentemente, uma forma de ter mais controlo sobre a informação necessária ao processo de tomada de decisão. Como resultado, espera-se que o número de transplantações efetuadas no Centro Hospitalar do Porto, Hospital de Santo António, EPE apresente valores mais rentáveis.

Palavras-chave: *Deteção de Potenciais Dadores de Órgãos, Processo de Tomada de Decisão, Transplantação de Órgãos*

Contents

List of Tables

List of Figures

Abbreviations and Acronyms

AIDA Agency for Integration, Diffusion and Archive of Medical Information. ix, 14, 15

API Application Programming Interface. 23, 50, 54

BD Brain Dead. v, 3, 4, 27, 28, 33, 34, 40, 41, 73

BI Business Intelligence. v, vii, 3, 5, 7–10, 13, 16, 18, 22, 51, 53, 54, 71, 72

CDSS Clinical Decision Support System. 5, 7–9, 13, 15, 16, 45

CE CT Cranio-Encephalic Computerized Tomography. 34–37, 40–42, 73, 74

CHD Coordenadores Hospitalares de Doação. 6

CHP Centro Hospitalar do Porto, Hospital de Santo António, EPE. iii, v, vi, viii, 6–8, 13–15, 25, 34, 37, 42, 45, 47, 54, 59, 65, 71, 74, 75

CNT Coordenação Nacional de Transplantação. 6

CRISP-DM CRoss Industry Standard Process for Data Mining. 21

CSS Cascading Style Sheets. 49

DM Data Mining. 13, 18, 19, 21, 74

DOM Document Object Model. 50, 51

DSR Design Science Research. 8, 71, 73

DSS Decision Support System. 16, 40, 72

DW Data Warehouse. 18, 25

ER Emergency Room. 14, 34, 54, 55, 57, 59

ETL Extract, Transform and Loading. 8, 18, 48, 54, 72

GCCT Gabinete Coordenador de Colheita e Transplantação. iii, vii, viii, 6, 7, 25, 27, 34, 37, 40–42, 45, 53, 59, 60, 62, 71, 73–75

GCS Glasgow Coma Scale. 74

HIS Health Information Systems. v, 4, 5, 7–9, 13–15, 17, 33, 45, 47

HTML HyperText Markup Language. 49–51

HTTP Hyper-Text Transfer Protocol. 22, 23, 46

ICD-9-CM International Classification of Diseases, Ninth Revision, Clinical Modification. x, xvi, 34, 35, 40–42, 55, 61, 62, 64, 73, 95

ICU Intensive Care Unit. 5, 33

IDE Integrated Development Environment. 37, 48

IPST, IP Instituto Português do Sangue e da Transplantação, IP. 5, 6

IT Information Technology. 7, 8, 67, 68

JSON JavaScript Object Notation. 24, 46, 47

KDD Knowledge Discovery in Databases. 9, 18, 19, 73

MAS Multi-Agent Systems. 14

MVC Model-View-Controller. 9, 24, 47, 49, 50

NLP Natural Language Processing. 19, 20, 36

ODBC Open Database Connectivity. 54

OLAP Online Analytical Processing. 18, 19, 51

ORM Object-Relational Mapping. 49

PCE Processo Clínico Eletrónico. 14, 15, 54, 59

PL/SQL Procedural Language/Structured Query Language. 47, 48

RENNDA Registo Nacional de Não Dadores. 5, 75

REST REpresentational State Transfer. 9, 22–24, 45, 46, 50

SAM Sistema de Apoio Médico. 54

SAPE Sistema de Apoio à Prática de Enfermagem. 54

SOA Service Oriented Architectures. 14

SQL Structured Query Language. 19, 47, 48

SVG Scalable Vector Graphics. 51

SWOT Strengths, Weaknesses, Opportunities and Threats. 9, 10, 67

TDM Term-Document Matrix. 21, 22

URI Uniform Resource Identifier. 23, 24, 47, 50

XML eXtensible Markup Language. 48

Part I

Introduction

Chapter 1

Introduction and Motivation

The present dissertation project of the Integrated Master in Biomedical Engineering - expertise in Medical Informatics - aims to describe the development and deployment of a clinical decision support platform in the field of transplantation, based on Business Intelligence.

That said, this first chapter serves the purpose of framing the proposed topics as well as describing the main motivation that led to their development and identify the objectives to be achieved. A brief characterization of the work done is presented and, finally, the structure of the dissertation is defined, in order to ease its understanding.

1.1 Overview

Organ transplantation is often the best and sometimes the last resort for patients with end-stage organ failure and even though organ donation and transplant rates vary widely across the globe, it is agreed that organ shortage of deceased donors can not keep pace with the ever-increasing demand for transplants. [1–3]

In the field of organ donation, an early detection of potential donors is crucial. There are many factors that influence the number of potential donors but, regardless of that, it is known that the main source of all deceased donations come from Brain Dead (BD) donors, implying that most potential deceased donors must be localized in the Intensive Care Units (ICUs) and Emergency Departments (EDs) in patients who are victims of stroke or other severe brain events. [1, 4]

A comprehensive review of organ donation considers BD patients as the ideal multiorgan donors. As a result, the major source of solid organs for transplantation are provided by BD patients. [1, 5]

Particularly focusing on BD donation, the main objective is to monitor deceased organ donation potential, evaluate performance and identify key areas for improvement. [6]

In Portugal, as in most countries, many systematic approaches and programs to increase organ donor rates and also to develop deceased donation to its maximum potential have been implemented. [1, 2, 6]

While this may be true, aside from exploiting the concept of donor rate, a more meaningful measure would be the proportion of possible donors that become actual donors, the so-called conversion rate. [1]

Clinical medical records hold a wealth of information that may prove to be valuable if used to full advantage. These increasingly electronic clinical records include the patient's medical history, prescriptions, physician notes, etc. collected from the various Health Information Systems. [7–9]

To grant accurate and timely information, Health Information Systems (HIS) that provide access to electronic patient records play an expressive role in optimizing the support of adequate decision-making. Therefore, it is no surprise that in hospitals, the complexity of environments where people and information are distributed expects considerable coordination and communication among the professionals that work in such settings. [10]

Actually, these systems must go along with the fundamentals of ubiquity and quality-of-care and be embedded in some forms of intelligent mechanisms in order to be useful for medical, clinical and administrative staff. [10–13]

Indeed, the exchange and share of clinical knowledge among medical information systems is an important feature to improve healthcare systems, quality of diagnosis, but mainly, to improve quality in patient treatment. [13]

Individuals who are responsible for making decisions in organizations are aware that timely and precise information is powerful enough to improve business performance. [14] Overall, a decision support system is described as a system that intends to support business decision makers in semi-structured or unstructured decision situations working with decision makers as adjuncts, extending their capabilities but not replacing their judgment. As a matter of fact, this type of systems was aimed at decisions that required judgment or could not be completely supported by algorithms. [15]

The development of this type of systems is being encouraged to support new knowledge construction, processes improvement, pattern recognition and interpretation in large amounts of data, localization and development of organizational competencies and experience sharing. [14]

That said, and knowing that coding, storing and transmitting knowledge in organizations is not new, recently the organizational and managerial practice has become more knowledge-focused, and hospitals are not indifferent to that. [16, 17]

To best describe the decision support solution's architecture and practice, the concept of Business Intelligence (BI) is gaining ground. [18] BI characterizes a variety of activities to collect all the necessary data to make solid business decisions such as creating a data warehouse and/or data mart to store the data and handling front-end analytical tools. Altogether, this set of tools delivers end-users a high-level solution so they can make better, informed decisions counting on reports, predictions and/or analytical views. [19]

Thus, BI is mainly referred to as a broad category of applications and technologies for gathering, storing, analyzing and providing access to data in order to deliver decision makers all the tools they need to obtain competitive advantages through quick business understanding and critical insights. [20, 21]

All things considered, it is possible to construct complex and rigorous Clinical Decision Support Systems (CDSSs) based on BI technologies with data from the various HIS that, when built upon a firm basis, can take the clinical decision-making to a whole new level.

1.2 Motivations

Instituto Português do Sangue e da Transplantação, IP (IPST, IP) defines transplant as a "transference of cells, tissues or living organs from one person - the donor - to another - the receiver - or a part of the body to another in order to restore a lost function".

The transplant of one or more organs can be crucial for the survival of a patient whose vital organs stopped functioning correctly or it is biologically impossible to recover their normal function. When a case is referenced, the patient is placed on a waiting list and is evaluated individually by a transplant team that assures the principles of equality and equity according to well-defined organs distribution criteria.

The Portuguese Law considers every Portuguese citizen who does not register in life in Registo Nacional de Não Dadores (RENNDA) as a potential organ donor, that is, the Portuguese legal framework considers presumed consent, unless the patients register themselves on the appropriate authority. Still, in order for a person's organs to be considered valid for donation, his/hers death must occur at a hospital's Intensive Care

Unit (ICU), since it is in this unit that the harvesting and the correct preservation of organs are performed. After carrying out all the necessary proofs, a medical team is responsible for determining if the deceased patient can be a donor and which organs can be used for donation.

According to statistical data published by IPST, IP, in 2015 the number of brain death donors was 319, 30 more than in 2014 (289) which represents an increase to numbers superior to any of the 4 previous years. This increase was reflected in a total of 896 organs harvested in 2015, 8% more than in 2014 (829). As for the first semester of 2016, the number of deceased donors reached 189, 27 more than in the same period of 2014, achieving the highest number of donors registered in the same period since 2009. The number of harvested and transplanted organs also rose to the highest values in the last seven years. [22]

To regulate and standardize the donation activities, procurement and transplantation of organs, tissues, and cells of human origin, there is in Portugal the Coordenação Nacional de Transplantação (CNT). CNT is an organic unit belonging to IPST, IP and has been conceived and structured at three different but interlinked levels: National (CNT), Regional (Gabinetes Coordenadores de Colheita e Transplantação (GCCTs)) and Local (Coordenadores Hospitalares de Doação (CHD), Unidades de Colheita de tecidos e células, Unidades de Transplantação/Aplicação de órgãos, tecidos e células).

The first two levels work as an interface between the technical and the political strata and act in support of the deceased donation process. The Local level (or Hospital level) of coordination is represented by a network of officially authorized procurement hospitals that are directly in charge of effectively developing the deceased donation process. At the Local level, the transplant coordinator appointed at each procurement hospital has been considered a key element for the model, having a unique profile, conceived to develop a proactive donor detection program and effectively converting potential into actual organ donors.

Centro Hospitalar do Porto, Hospital de Santo António, EPE (CHP) is represented as a GCCT and, as well as the other GCCTs, counts on multidisciplinary teams to carry out organs, tissues and cells procurement. The role of the transplant coordinator is performed by medicine professionals with specific training to provide prompt identification and evaluation of potential organ and tissue donors for transplantation.

Indeed, it is this identification of potential donors that creates some complications mostly due to the high number of factors to take into account for this detection. The complexity inherent to this process challenges and slows down its effectiveness. Performing multiple manual tasks at the same time becomes limiting in terms of time and resources management. Meanwhile, a possible donor can actually die while his potential of donation is still being evaluated, resulting in loss of donation potential in hospitals. [3]

As a result, the use of Information Technology (IT) solutions can provide an added value in facilitating the process of potential donors referral using the different existing HIS. By applying the right set of information technologies, it is possible to develop CDSSs, systems that use the information provided by the various HIS for decision-making in more or less critical situations supporting the decision made by healthcare providers. Working together, this can lead to an increase in the donation potential in hospitals and therefore, address the organs needs by contributing to a timely and accurate detection.

Thereby, the motivation for the development of such a project focuses on these principles, seeking to respond to failures that the system displays by creating a CDSS platform in the field of transplantation based on BI.

1.3 Objectives

For purposes of increasing the availability of organs in CHP and trying to solve the problem of crippling delay in the potential donor referral, the author proposes to continue the existing platform developed in the context of a previous MSc in Biomedical Engineering dissertation, named **Organite**. As so, the author proposes to contribute actively to its continuous improvement and optimization. [23]

Therefore, the main goal of this project is to acquire knowledge on how to design and develop new features resulting in a high-level solution applicable in clinical context. That must result in a useful and accurate tool so that the GCCT team at CHP can improve their results.

As a result of the work previously done, a repository of potential organ donors and the mechanisms responsible for feeding it in real-time are already defined and will be used in this dissertation.

Given the complexity of a project with this magnitude, it was decided that it would be better to divide the project into several specific objectives:

1. Regarding the potential organ donors identification model:

    - Perform a longitudinal study from the actual transplanted at CHP in the past five years for feature extraction based on the frequency analysis in order to define a more accurate potential donor pattern;

    - Discuss the observed results with the GCCT team at CHP in order to describe the structure of the model to apply.

2. Regarding the repository of potential organ donors:

   - Analyze and study the existent HIS at CHP to ensure their interoperability, as well as the architecture of the previously defined repository;
   - Correct any possible situations that are not being captured correctly or that may be excluded by computer error;
   - Introduce new potential donor detection filters defined in the previous stage;
   - Automatically classify medical records in order to categorize patients according to their status (death, medical discharge, etc.) using Extract, Transform and Loading (ETL) procedures.

3. Regarding the web platform:

   - Develop an intuitive user interface allowing a real-time graphical tracking of potential donors;
   - Develop an authorization module to sign in and log in users with different permissions;
   - Define a notification system to alert responsible members whenever a new possible donor with a certain level of potential is identified;
   - Provide BI tools for statistical analysis and study of the data collected.

1.4 Research Process and Activities

It may be challenging when developing medical applications to provide solutions that are both useful and accepted by its future users. As for that, and to address the above research objectives, several research activities have been conducted in order to study and specify the main concepts in such a project.

As one of the main objectives of this project is to provide a solution to a set of identified problems and to increase the utility of a CDSS, the Design Science Research (DSR) methodology was used as it is the most suited to this type of projects.

This project began with a revision of the literature and the state of the art regarding the transplantation activity in Portugal and the strategies followed in other countries to improve the situation. At the same time, a comprehensive understanding of how IT and BI tools could benefit this type of solutions was also performed.

Given the importance of those domains for the whole work, decision-driven techniques and methodologies applied to organizations were also assessed.

After a full definition of terminology, the design and development stage started with a decision on what technologies were to be used. A set of tools was evaluated and chosen according to their fitness to the project. As part of this stage, cooperation with medical staff in order to get constructive feedback during the development process, also played an important role.

Another key stage is the proof of concept, i.e. showing that the designed solution brings benefit to the work load for the staff. This was done taking the already existent solution as a term of comparison and performing a Strengths, Weaknesses, Opportunities and Threats (SWOT) analysis to assess the system's performance and acceptance.

1.5 Document Organization

This document is divided into five main parts.

The first part, the Introduction, explores superficially the research problem and specifies the project's motivations and aims.

The second part, the State of the Art, describes the background of the problem in two chapters: Chapter 2 and Chapter 3. Chapter 2 defines the technical context of the institution where this project will be applied, beginning with an introduction of the main Health Information Systems, their organization, and relevance in this project. Concepts like interoperability, Clinical Decision Support System, Business Intelligence, Knowledge Discovery in Databases, REpresentational State Transfer and Model-View-Controller will also be characterized in this chapter, and a brief history on the previous state of the system will be presented. Chapter 3 explains the main activities provided for decision-making in the transplantation area, as well as some features and limitations.

The third part, Defining the Potential Organ Donor Criteria, defines the knowledge discovery and describes the methods and technologies used in the process of inducing useful knowledge for clinical decision support, in Chapter 4. These criteria are presented as a complement to the existing ones contributing for its continuous optimization.

The fourth part, Design and Implementation, describes the technical details and design choices of the constructed system, in Chapter 5. Chapter 6 characterizes the implementation elements and presents the visualization layer and all the BI tools and procedures constructed for data representation, in the form of results. Chapter 7 evaluates and demonstrates the project's feasibility using a SWOT matrix to verify its potential for real-world application.

The fifth part, the Conclusions, expresses in Chapter 8 the final conclusions of this dissertation and enumerates several issues that may constitute an opportunity for future work.

Part II

State of the Art

Chapter 2

Conceptual Framework

This chapter aims to describe the theoretical and scientific concepts explored along this dissertation.
After identifying the main Health Information Systems (HIS) applied at Centro Hospitalar do Porto, Hospital de Santo António, EPE (CHP), where this project is implemented, the central concepts like Clinical Decision Support System (CDSS), Business Intelligence (BI), and Data Mining (DM) are also reviewed in order to promote a full understanding on how this project fits the specifications.

2.1 Health Information Systems at CHP

In medical science, where lives are at stake, a perfect scenario would be for all healthcare professionals to be perfectly skilled and informed, and for healthcare to be a field of order, knowledge and defined procedures even though uncertainty is a part of everyday life in surgical decision-making. However, one of the greatest concerns in nowadays' organizations is to introduce technology into their environments, and hospitals are not an exception. Consequently, HIS have been introduced in order to improve the quality of healthcare delivery. [24–27]
HIS can be defined as the socio-technical subsystems that comprise all management and organization information processing, as well as the healthcare professionals roles. [28, 29] The design and implementation of HIS must focus on ensuring the efficient production of information in order to provide clinical decision-making capabilities. [29, 30] As to provide complete and useful features, HIS must also allow the extraction of clinical and management indicators as a way to improve the decision-making, planning and logistics pro-

cesses. [31, 32]

The following list portrays examples of some of the HIS implemented at CHP:

- **Processo Clínico Eletrónico (PCE)** — defined as an electronic repository of patients' data, PCE stores the information about patients in a secure, organized and centralized way making it accessible whenever needed;
- **SClínico** — defined as a system to represent, manage and archive clinical information, SClínico is built upon two former applications: Sistema de Apoio ao Médico (SAM) and Sistema de Apoio à Prática de Enfermagem (SAPE) which comprised information on medical and nursing practices, respectively. Besides the previously defined functions by both systems, SClínico holds a new set of features like integration with the electronic medical prescription, for instance;
- **Sistema de Urgências** — defined as a dedicated system for registry, management and archive of clinical episodes occurring in the Emergency Room (ER).

As these systems are developed in an isolated way by different entities, their independence hinders the interaction among different systems. To solve this kind of problems, the need to implement proper communication and cooperation among these systems became evident. That's where the concepts of integration and interoperability apply, as both are important for cooperation and information flow in healthcare organizations. [26]
The objective of integration is to obtain information from several systems in order to improve their capabilities whereas the objective of interoperability focuses on a continuous communication and information exchange in different cooperating systems. [26, 33]
The Agency for Integration, Diffusion and Archive of Medical Information (AIDA) is an agent-based platform created by a group of researchers from the University of Minho, the Artificial Intelligence Group, and is currently installed in several Portuguese hospitals, like CHP, to guarantee interoperability. [26]

2.1.1 Agency for Integration, Diffusion and Archive of Medical Information

AIDA is an electronic platform that imbues many different integration features, using mainly Service Oriented Architectures (SOA) and Multi-Agent Systems (MAS) to implement interoperability and overcome difficulties in achieving uniformity of clinical systems, as well as medical and administrative complexity of different HIS. [34, 35]

With its proactive behavior, AIDA is in charge of tasks such as communicating with the heterogeneous systems, sending and receiving information (e.g. medical or clinical reports, images, collections of data, prescriptions), managing and saving the information and answering to information requests, with the necessary resources to their correct and in time-line accomplishment. [34–39]

This way it is possible to access information anytime and anywhere in the hospital thereby allowing a more efficient clinical information management. [36] Additionally, AIDA supports web-based services that enable human interaction with the system such as sending SMS via phone or email. [37]

At CHP, the system responsible for collecting, consolidating and making data available for PCE is AIDA.

2.2 Clinical Decision Support Systems

Making decisions is characterized as the process of developing and analyzing alternatives to make a **decision** which is nothing more than a choice from the available alternatives. The majority of the decisions are made in response to a **problem** and involve **judgment**, which are the cognitive aspects of the decision-making process. [40]

Healthcare organizations, as many others, are striving to make sense of the rapidly increasing volume and variety of data generated by both internal and external resources. [41] The challenge now is how to make sense of the raw data inside relational databases and produce meaningful information for executive professionals to make effective decisions in the diverse clinical areas. [42]

That's exactly where the CDSSs fit in. CDSSs are described as "any electronic system designed to aid directly in clinical decision making in which characteristics of individual patients are used to generate patient-specific assessments or recommendations that are then presented to clinicians for consideration". [43]

In their classic form, CDSSs include alerts, reminders, order sets that automatically remind the clinical of a determined action, or care summary dashboards that provide feedback on quality and performance indicators for further improvement. [44]

Such systems have been shown to improve overall healthcare quality and efficiency by enhancing the delivery of preventive care services [45, 46], and improve adherence to recommended care standards. [47] Reviews on the deployment of CDSSs showed that: [48]

- Computer-based decision support is more effective than manual processes for decision support;
- CDSS interventions that are presented automatically and fit into the workflow of the clinicians are more likely to be used;
- CDSSs that recommend actions for users are more effective than the ones that simply provide assessments;
- CDSS interventions that provide information at the time and place of decision-making are more likely to have an impact.

Nevertheless, it is important to understand that a CDSS's target is not to replace clinicians but to support their daily work by reconstructing the way clinical episodes are seen.

2.3 Business Intelligence

Cody, Kreulen *et al.* [49] refer to BI as systems that combine data from different sources with analytic tools in order to provide relevant information for the decision-making process. The aim of such systems is to improve the availability and overall quality of that information. [14]

When compared to Decision Support System (DSS), BI is defined as a broad category of applications, practices, and technologies for gathering, storing, analyzing and providing access to data to support the business users in making better decisions. This includes all the ways an enterprise can explore, access and analyze information in the data warehouse to develop insights that will utterly lead to improved, informed decisions. [40, 50]

In fact, BI is now considered a top priority for many organizations and although healthcare still lags behind other industries in adopting BI, leading practitioners agree that taking a more enterprisewide, platform-based approach to BI provides large gains to the institution. [50]

Depending on the requirements of every project, technologies can be merged and combined, creating a reliable DSS architecture. This DSS's architecture must back up all the technological requirements BI might need. After a literature analysis, the framework that best suits this project is composed of three levels, from a development point of view, from bottom to top: The Data Sources level, the Data Warehouse level, and the Business Analytics level as represented in Figure 2.1. [51] This approach was proposed by Han and Kamber [52] and Eckerson [53].

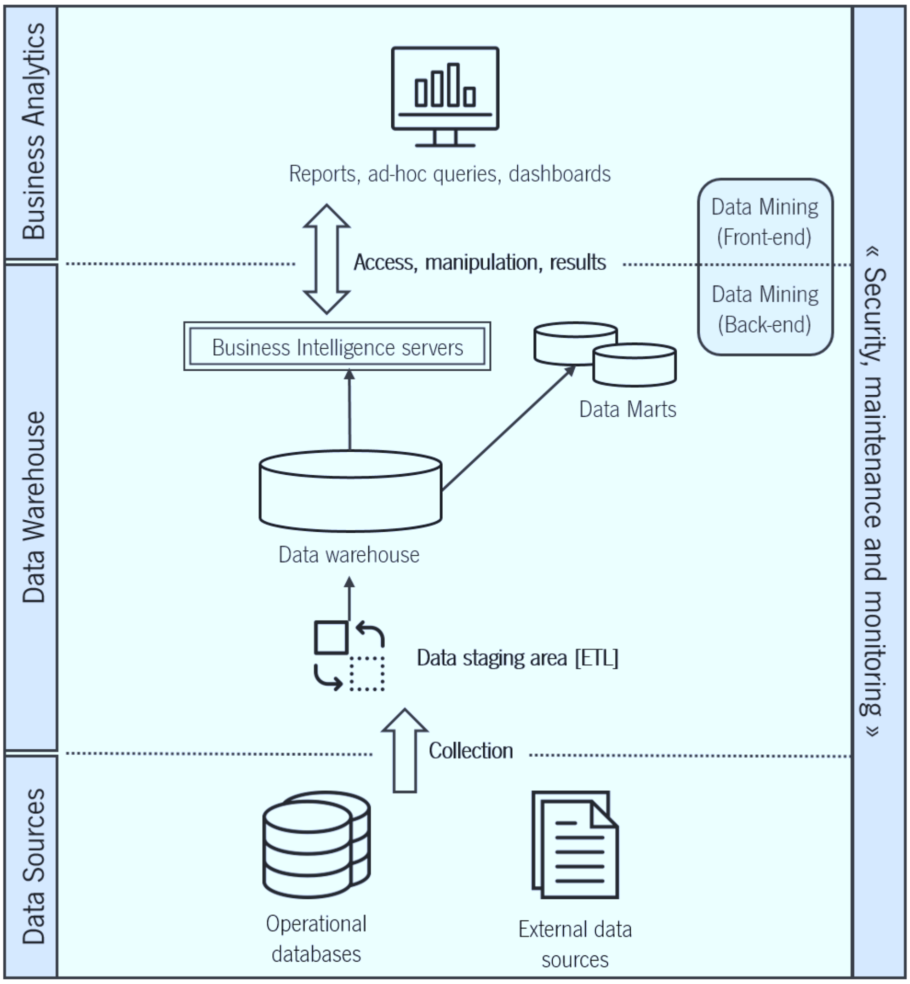


Figure 2.1: BI's support infrastructure architecture as proposed by Han and Kamber [52] and Eckerson [53] (adapted from [18]).

The Data Sources level represents all the operational environment of the application. These Data Sources are represented as databases including the data from different operational systems, responsible for feeding the Data Warehouse. They represent the origin of information intended to collect and make available through the Data Warehouse and they also represent the diverse operational activity's support systems used in an organization, in the form of HIS. [18]

The Data Warehouse level represents the storage environment that include the data validation and transformation before storing it in the Data Warehouse (DW). There always exist some kind of treatment to perform over the data before integrating it into the analytic repository. The area where this happens is called the Data Staging Area. Here, sequential search algorithms and text transformation procedures predominate, called the Extract, Transform and Loading (ETL) procedures.

Above the Data Staging Area is the DW which, as mentioned before, is the analytical repository where data is stored after correctly transformed and integrated. It is over the DW that a set of services with the objective to provide information to end-users is implemented. This set of services includes a variety of utilities such as multidimensional analysis servers that can be later accessed by Online Analytical Processing (OLAP) tools, BI servers, Data Mining engines, amongst others. [18, 54]

The Business Analytics level is characterized as the set of technologies that support real-time analytics, enabling end-users to interact and explore the data from the repository previously created. Here the user is provided with a set of dynamic analysis tools that can be used to query data, generate reports or identify trends and patterns in data. That's where the interaction with the users takes place and where they can communicate with the system and analyze the presented results. This is one of the reasons why the user interface must be specially designed for an intuitive interaction.

2.4 Knowledge Discovery in Databases and Data Mining

Knowledge Discovery in Databases (KDD) is a fast-evolving research field which development is targeted to the benefit of practical, social and economic fields, amongst others. [55]

The motivation for this growth is linked primarily to the existence of a powerful technology for collecting, storing and managing large amounts of data, also known as big data. Most of this data hold valuable information like patterns and trends that can be utilized to improve business decisions. However, considering that the data is collected into databases, it became necessary to develop automatic processes for data analysis, like DM. [55]

As mentioned above, OLAP tools ease the process of querying data as they are query-oriented, that is, the user controls what is represented by querying the system. This user-dependent approach can, however, interfere in the process of intelligently identifying patterns in data. To counter this, the utilization of automatic (or semi-automatic) computer-dependent analysis techniques for knowledge extraction on big repositories of data became evident. [55] These techniques usually don't go beyond simple Structured Query Language (SQL) queries, OLAP tools or Data Visualization mechanisms. [55]

Towards such deficiency to analyze and comprehend large volumes of data, several studies have been conducted in order to develop knowledge extraction techniques, also known as DM. This process of KDD aims at finding knowledge from a set of data so it can be used in decision-making processes. Therefore, an important requirement is that this knowledge is human-understandable, as well as useful and interesting for end-users, generally the decision-makers, in a way this knowledge can support their decision-making process. [55, 56]

The particular step of KDD of finding useful patterns and trends in data with the goal of finding new information and transform this specific information into actionable results is then assigned to DM. [56, 57]

2.5 Text Mining

As the DM techniques have been developed to deal with structured data, specific Text Mining techniques to deal with unstructured data were conceived. [55]

Text Mining is described as a process of knowledge discovery based on technologies such as Natural Language Processing (NLP), information extraction and data mining that makes it possible to discover patterns and trends semi-automatically from huge collections of unstructured text. [58–64]

A Text Mining system, as well as a data mining one, relies on preprocessing routines, pattern-discovery algorithms and presentation-layer elements such as visualization tools to ease how users skim through the results obtained. [63, 65, 66] In contrast to data mining, Text Mining systems' preprocessing operations focus on the identification and extraction of representative features from natural language documents. These preprocessing operations are responsible for modeling unstructured text stored in document collections (implicit knowledge) into structured content (explicit knowledge). [65, 66]

With Text Mining, researchers intend to discover new patterns and knowledge aggregates that may not be so obvious or known using elements for frequency analysis, identification of documents containing certain attributes for further statistical analysis, etc. The information obtained through these elements is used to create new characteristics from data for predictive modeling. [67]

Moreover, Text Mining builds upon many disciplines concerned with the handling of natural language covering techniques and methodologies from the areas of information retrieval, information extraction and corpus-based computational linguistics. [65]

Term extraction is the most common form of Text Mining. This technique maps information from unstructured data into a structured format. The simplest data structure in Text Mining is the feature vector which is a weighted list of words. The most important words on a document are listed along with their relative importance. This way, the text within documents is reduced to a list of terms and weights. The entire semantics is not considered yet the key concepts are identified. This is called a statistical analysis, where the importance of a term is given for its frequency within the documents. [55, 68, 69]

2.5.1 Text Mining in Healthcare

In healthcare, physicians often express opinions in terms of words that contain useful information that can be further used to develop intelligent models to improve the healthcare process. Hence, Text Mining proves to be an effective tool in healthcare datasets if correctly applied within clinical record archives leading to an accurate prediction of future outcomes in various healthcare settings. [8] Hirschman *et al.* [70] notes in his review of milestones in biomedical Text Mining research that the field began by focusing on three approaches to process text: linguistic context of text, pattern matching, and word co-occurrence. This last one, word co-occurrence, proves to be very effective on occurrence statistics-based predictions.

As the output of a NLP system is intended to be employed by a healthcare application, it must hold adequate recall, precision and specificity for the intended clinical application, but it should also be possible to adjust its performance according to the needs of the application. Any application involving NLP must undergo a performance evaluation before being deployed to ensure it is appropriate to address such clinical settings. [63]

One evident challenge in the Text Mining field lies on the pertinent declaration of negatives in text which makes it harder to determine whether a finding within narrative medical records is present or absent. Indeed, the most frequently described findings are the ones that deny these findings on the patient. [71] Thus, the development of methods to automatically discriminate between terms that are mentioned as being present and terms that are negated provide a crucial improvement for this type of systems where negation in predicate logic is well defined and syntactically present.

2.5.2 The Process of Text Mining

Text Mining studies should follow a structured methodology based on best practices. For that, a standard process model similar to CRoss Industry Standard Process for Data Mining (CRISP-DM) - the industry standard for DM projects - is needed. Nonetheless, a specific process model for Text Mining should include much more elaborate data preprocessing methods. [15]

At a very high-level, the Text Mining process can be expressed in a three consecutive tasks diagram as depicted in Figure 2.2.

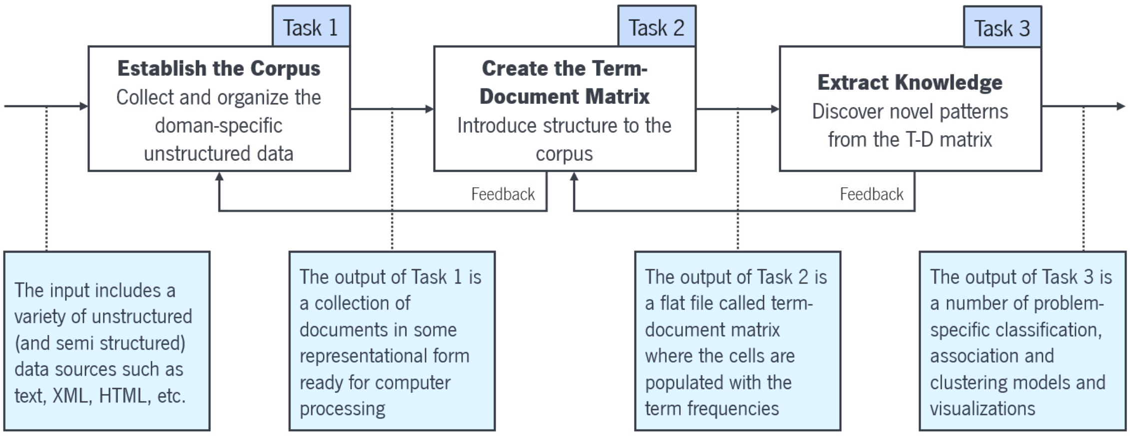


Figure 2.2: The three-step Text Mining process (adapted from [15]).

Task 1's purpose is to collect all documents related to a domain of interest for analysis. Once the document files are collected, they should be transformed and organized in such manner that they are all in the same representational form to be read by the Text Mining software. [15]

Task 2's goal is to convert the list of organized documents (the corpus) into a Term-Document Matrix (TDM) where the cells are filled with the most appropriate indices being the most common a list of frequencies for each term used in the documents. [15]

However, not all terms are equally important when characterizing documents and therefore should be disregarded has they have no differentiating power. Preprocessing methods should, then, filter these terms and exclude them from the indexing process as stated below: [72]

- Eliminate commonly used words, known as **stopwords**. The most common words in text documents are articles and prepositions which carry no value and do not describe the documents' content;

- Identify and replace the words by their corresponding root/stem terms, that is, eliminate plurals, conjugations and suffixes. This is known as **stemming**;
- Consider synonyms and phrases, remembering that they should be combined anyhow.

Raw term frequencies are often used to represent the indices and the most common measures are: [72]

- Term frequency (*tf*): measures the real term frequency within a document;
- Inverse document frequency (idf): measures the rareness of a term by comparing the number of documents in the collection with the number of documents containing that term.

Task 3 consists in using the TDM to extract knowledge using common methods such as classification, clustering, association and trend analysis. [15]
The participation of professionals is very important throughout the whole knowledge extraction process whether in the analysis, helping to resolve conflict situations, or indicating alternative routes and complementing information. [55]
To sum up, Text Mining is emerging as a critical component of the next generation of BI tools enabling organizations to compete successfully as companies can now better understand their customers by analyzing their feedbacks. Therefore, it is no surprise that Text Mining applications are in virtually every area of business, including healthcare. [15]

2.6 REpresentational State Transfer

REpresentational State Transfer (REST) is an architecture style for designing networked applications which uses simple Hyper-Text Transfer Protocol (HTTP) rather than using complex mechanisms to connect between machines. REST results then in a coordinated set of architectural constraints to define a uniform connector interface. [73]
The main characteristics of a REST system are defined below: [73–75]

- **Client-Server** – There should be a separation between the server that provides a service, and the client that consumes it;

- **Stateless** — Each request from client to server must contain all the information needed to understand the request and cannot take advantage of any stored context on the server meaning that session state is uniquely kept on the client;
- **Cacheable** — The server must indicate to the client if requests can be cached or not;
- **Uniform Interface** — The communication method between a client and a server must be uniform;
- **Layered System** — Communication between a client and a server should be standardized in such a way that allows intermediaries to answer to requests instead of the end server, without the client having to do anything different;
- **Code on demand** — Servers can provide executable code or scripts for clients to execute in their context. This constraint is optional.

These constraints improve some properties such as **portability** of the user interface across multiple platforms and simplify the server components. [73, 74]

The properties of **visibility**, **reliability** and **scalability** are also improved. Visibility is improved because the system does not need to look beyond a single request datum in order to determine the full nature of the request. Reliability is improved because it eases the task of recovering from partial failures. Scalability is improved because by not having to store session state between requests, the server can quickly free resources. This further simplifies implementation since the server doesn't have to manage resources usage across requests. [73, 74]

The REST architecture was originally designed to fit the HTTP protocol used in the world wide web. Web services following a REST architecture are usually called RESTful web services or RESTful web Application Programming Interfaces (APIs) (or simply REST APIs). In such systems, the resources are represented by Uniform Resource Identifiers (URIs). Clients send requests to these URIs using the methods defined by the HTTP protocol and, as a result, an action is performed. [73–75]

The HTTP request methods are usually designed to affect a certain resource in standard ways, explained in Table 2.1. Together, these describe the Create, Read, Update and Delete (CRUD) operations defined on the resources. [75]

Table 2.1: HTTP request methods.

HTTP method	**Action**
POST	Create a new resource
GET	Read a resource
PUT	Update a resource
DELETE	Delete a resource

The REST design architecture does not require a specific format for the data provided with the requests and/or responses. However, the most common data format is JavaScript Object Notation (JSON).

2.7 The Model-View-Controller design pattern

The Model-View-Controller (MVC) design pattern describes a way of structuring an application into three different component parts, so as to separate internal representations of information from the ways that information is presented to the user, allowing the creation of complex solutions through easily organized and maintainable code. [76] Typically, a certain content is requested by entering a URI. The **Controller** receives that request and the **Models** retrieve all the necessary data from the databases. The **Controller** then organizes it and sends it off to the **View** which makes that data available on a specific format for further access, in the front-end layer.

On a more technical level, whenever someone types in a URI in a browser to access a web application, a request is being made to view a certain page within the application.

When building a web application, each URI pattern is associated with different pages and content in the form of **Routes** so when someone enters a URI the application tries to match that URI to one of the predefined routes. Each route is associated with a **Controller** - more specifically, a certain function within a controller, also known as the controller action. As so, when entering a URI, the application attempts to find a matching route and if it finds one, it calls the associated controller action.

2.8 History of Organite

Organite started in 2014 as MSc degree dissertation in Biomedical Engineering at University of Minho. [23] This dissertation resulted in a functional prototype of a decision support platform providing a great contribute in the transplantation field at CHP.

This first version of **Organite** featured the construction of the DW and all the mechanisms needed for feeding it in real-time according to some details provided by the Gabinete Coordenador de Colheita e Transplantação (GCCT) team at CHP which played an essential role in the design and deployment phases in order to develop a functional solution.

Organite's first user interface had a list of all the identified diagnostics and a statistical area with some charts providing simple, yet useful insights on the collected data. It was not easy to understand which patients were available since no information was delivered on the list about the patients' medical condition. The user needed to click on each patient and then understand if the patient was available or not by exploring the patient's details. This was time-consuming and lowered the user experience.

It was also known that the mechanisms responsible for feeding the DW in real-time introduced some erroneous situations and needed to be corrected and restructured.

In this sense, the author proposes to continue **Organite**'s development and improvement by developing a second version of the prototype.

Chapter 3

Donation and Transplantation Activity

3.1 Current Practice in Organ Donation and Transplantation

Transplantation is the treatment of choice for severe organ failure. Despite this, the number of patients waiting for one or more organs is still increasing in Europe and as the shortage of organs has become more critical, new proposals have been developed to increase the organ donation potential. [3, 77]

In Portugal, as in most countries, many systematic approaches and programs to increase organ donor rates and also to develop deceased donation to its maximum potential have been implemented, such as the initiation of the presumed consent legislation. [1, 2, 5, 6]

In fact, according to annual reports from the National Transplant Coordination, the number of Brain Dead (BD) donors in Portugal has been increasing since 2012. Following this trend, there has been a change in how deceased donation is performed and most countries increased their deceased donor rates, in the past years. [1,22] However, to perform a qualitative comparative analysis between different hospitals or countries, uniform criteria for potential organ donor is needed yet difficult to achieve. [78]

Even though there is no consensus on uniform criteria for defining a potential organ donor, a comprehensive review of organ donation considers BD patients the ideal multiorgan donors. As a result, the major source of solid organs for transplantation is provided by BD patients. [1, 5] The work of the medical team at Gabinete Coordenador de Colheita e Transplantação (GCCT) focuses on determining precursors of brain death that can provide significant results in creating a reliable pool of potential organ donors. As so, potential organ donors should be identified as soon as possible and should never be missed or delayed and it is desirable to identify

a potential organ donor as early as possible to achieve a donor conversion rate as high as possible. [3]
However, detection and management of potential organ donors is a complex process which takes several elements in consideration. [77]
Particularly focusing on BD donation, the main objective is to monitor deceased organ donation potential, evaluate performance and identify key areas for improvement. [6]

3.2 Features and Limitations

If a patient meets the criteria defined of a potential organ donor does not become a donor, it is de facto because of medical or social factors. [3, 78]
Parameters like age and medical condition play an essential role in excluding a patient from the pool of potential organ donors. The same holds for some medical reasons such as severe viral, bacterial or fungal infections and malignant neoplasm. As so, the actual pool of potential organ donors is obtained after excluding such patients that evidence the previous conditions. [3]
The social factors for non-procurement include family refusal or prior patient refusal, as stated by the law. Targeting education campaigns of the general population to increase awareness and understanding of brain death is one way to overcome this. [3, 78] Also, ethical and moral issues have been raised as the process of transplantation must respect and consider the patients' best interests.
After deciding on a patient's donation, there is a need to ensure that all developments have a firm basis in legal and ethical practice with equity, quality, and safety at their core. [1] In Portugal, this approach considers two fundamental aspects: regional aspects and clinical aspects.
The regional aspects determine that the organs from donors of a given region are transplanted in the same region, to decrease the warm ischemia time (maximum period of time that can elapse between the harvest of an organ and its transplant).
The clinical aspects determine the donor/receiver compatibility and the patient's needs. There is one clinical criterion above the regional ones that consists on the patient's emergency of the transplant. An emergent transplant request has absolute priority in all national territory. The transplant team decides, consulting the waiting list, who is the most suitable patient to receive an organ, following the clinical criteria.
In short, one can see that the request for organ donation is a very complex process that requires knowledge

of medical facts and ethical and legal principles, as well as an understanding of the myriad of psychological and social factors that influence human decision making in such a critical field. [79]

Part III

Defining the Potential Organ Donor Criteria

Chapter 4

Knowledge Discovery in Transplantation

4.1 Introduction

Clinical medical records hold a wealth of information that may prove to be valuable if used to full advantage. These increasingly electronic clinical records include the patient's medical history, prescriptions, diagnosis, physician notes, etc. collected from the various Health Information Systems (HIS). [7–9, 71]

Accordingly, it is not a surprise that documents in the scientific literature play an important role in life by supporting as a potential source for underlying knowledge discovery. [80]

However, besides their potential to improve healthcare quality, documents that comprise the health records are wide-ranging in complexity, length and use of technical vocabulary, making knowledge discovery a paradigm. [8]

As stated previously in this dissertation, Brain Dead (BD) donors constitute the primary source of deceased donations and are often a result of patients who are victims of stroke or other severe brain lesions. A predictive risk modeling for early detection of patients with the highest probability to evolve to brain death can have a significant impact on BD donation. [1, 4]

4.2 Methods for Defining a Potential Organ Donor Criteria

Detecting critical factors in the process of BD isn't always easy. It requires the construction of specific indicators that, altogether, contribute to the optimization of the donation potential.

In the most common, yet not unique, scenario, a potential donor is hospitalized in an Intensive Care Unit

(ICU) or Emergency Room (ER) whose clinical condition is suspected to fulfill brain death criteria. However, there are several factors that may indicate a patient is evolving to brain death and it is on predicting that evolution this modeling was built.

4.2.1 Using International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes to identify a potential organ donor

At a preliminary stage, a crucial criterion for identifying a death that may yield a potential organ donor consists on filtering medical record reviews indicating primary brain death.

This definition of potential organ donor returns patients' records clinically consistent with organ donor potential based on ICD-9-CM codes.

In that sense, a list of ICD-9-CM codes was developed by the Gabinete Coordenador de Colheita e Transplantação (GCCT) team at Centro Hospitalar do Porto, Hospital de Santo António, EPE (CHP) based on retrospective data analysis from recent years and includes a set of diagnostics that may yield a potential BD donor. For example, the list contains ICD-9-CM codes for subarachnoid hemorrhage (SAH), traumatic brain injury (TBI) and intracerebral hemorrhage (ICH) which are known to precede brain death in over 80% of the cases, according to the literature. [81]

The complete list of ICD-9-CM codes used for this purpose is represented in Table 4.1.

Even though the creation and management of the procedures to retrieve and store the results of this method are out of the scope of this dissertation, it mainly searches for patients' records on the diverse hospital databases and retrieves the ones that are consistent with any of the codes listed in Table 4.1.

4.2.2 A Text Mining Approach for Potential Organ Donors Detection

Another expressive way of detecting patients that may lead to BD donors rests on the evaluation of the patient's Cranio-Encephalic Computerized Tomography (CE CT) scan reports looking for medical evidence of devastating neurological events. [82]

To best classify and identify potential donors, a systematic approach for detection was adopted. This approach is designed to retrieve a set of most common keywords and expressions in potential organ donors' CE CT scan reports so it can be applied to the software-based screening system.

In order to get clinically relevant and accurate results, the context of this approach was narrowed to the CE

Table 4.1: ICD-9-CM Codes for Inclusion of Potential Organ Donors

ICD-9-CM Code	Description
191	Malignant neoplasm of brain
192	Malignant neoplasm of other and unspecified parts of nervous system
225	Benign neoplasm of brain and other parts of nervous system
320	Bacterial meningitis
324	Intracranial and intraspinal abscess
348.1	Anoxic brain damage
348.4	Compression of brain
348.5	Cerebral edema
348.8	Other conditions of brain
348.9	Unspecified condition of brain
430	Subarachnoid hemorrhage
431	Intracerebral hemorrhage
432	Other and unspecified intracranial hemorrhage
433	Occlusion and stenosis of precerebral arteries
434	Occlusion of cerebral arteries
435	Transient cerebral ischemia
436	Acute, but ill-defined, cerebrovascular disease
800	Fracture of vault of skull
801	Fracture of base of skull
803	Other and unqualified skull fractures
804	Multiple fractures involving skull or face with other bones
850	Concussion
851	Cerebral laceration and contusion
852	Subarachnoid, subdural, and extradural hemorrhage, following injury
853	Other and unspecified intracranial hemorrhage following injury
854	Intracranial injury of other and unspecified nature

CT scan reports from those patients who became actual organ donors from the last five years.

CE CT Scan Reports

A CE CT scan report is commonly expressed by references to disorders and diseases, i.e. abnormalities of the brain, which describe the radiologist's observations of the patient's medical conditions in the associated medical images. Altogether, these disorders and diseases registered may become a significant indicator for future early organ donor detection. [7]

However, CE CT scan reports are in free text format and usually unprocessed, which becomes a great barrier between the scan reports and the medical professionals who need them, making the information and knowledge retrieval difficult. [7]

The problem here lies in the inherent difficulty in manually analyzing a large dataset of scan reports. Thus, text mining offers Natural Language Processing (NLP) techniques to process the text and the rules in frequent medical finding pattern database. [7]

System Architecture

The main objective of this system is to **(a) extract critical medical findings and related information in the free text scan reports** and **(b) deploy that structured knowledge for future events detection**.

(a) Keywords/Expressions Extraction Module The process began with a collection of documents to be analyzed, that is, a collection of CE CT scan reports from actual organ donors from 2010 to 2015. Data extraction and cleaning measures are the first processes and require meticulous and precise attention to ensure that the data is valid and the information is complete. Domain knowledge plays a fundamental role at this stage.

For a document, d, a weight can be assigned as a reference to the number of occurrences of the term, t, in that document. This is referred to as term frequency and, as not all words are equally important in a document, a list of stopwords to be ignored in the document while performing the analysis was adopted. This stoplist contains commonly occurring terms that do not contribute in any way to the retrieval and scoring process, providing more reliable results. [83]

Term frequency considers that all terms are equally important when it comes to assessing relevancy on a query. In fact, certain terms have little or no discriminating power in determining relevance. To correctly assess a meaningful relevance determination, the concept of collection frequency, *cf* defines the total number of occurrences of a term in the collection. However, it is more commonplace to use the document frequency, df, to define the number of documents in the collection c that contain a term t. By trying to discriminate between documents for the purpose of scoring, the df proves to be more effective as it is a document-level statistic rather than a collection-level one, for the term. [83]

With this in mind, an algorithm run for n-gram frequency extraction was performed resulting in a table of all the n-grams set along with their corresponding frequency in the collection of documents. This step required a general understanding of algorithms capable of converting raw textual data into useful numeric information that users can comprehend. This step was achieved using the R programming language, the Integrated Development Environment (IDE) R Studio and the tm package, specific for text mining. [84, 85]

A total of 134 CE CT scan reports performed by 62 actual donors from 2010 to 2015 were analyzed. To sum up the results of this algorithm run, Table 4.2 presents the most frequent terms, along with their frequency measures within the collection of CE CT scan reports performed by actual organ donors. A minimum frequency limit of 20 occurrences was set in order to determine a certain degree of reliability.

Due to the essential participation of the medical team in this type of solutions, the final set of keywords and expressions was discussed with responsible medical professional working in the GCCT team at CHP. This way it was possible to integrate their field expertise in deciding the criticality of a word occurrence, according to the defaults of medical findings related to devastating neurological events.This permitted the removal of some expressions that wouldn't produce any added value to the system since they represent common terms in any CE CT scan report, often names of brain structures such as parenchyma and cerebellar tonsils, just to name a few.

Also, some synonyms were specified in order to increase the potential of this tool, and additional keywords were introduced. The final set of clinical conditions represented in the form of keywords and expressions is displayed in Table 4.3.

This concludes a two-step process for keywords and expressions extraction: preprocessing the large volume of textual data and developing useful predictions based on that data.

To draw up a schematic view on this, Figure 4.1 represents the extraction module architecture.

(b) CE CT Scan Reports Classification Module Next, a retrieval module was constructed and is responsible for building a connection between the set of keywords previously defined and the scan reports database.

Anytime this module finds a CE CT scan report, it applies a natural language algorithm in order to identify the affirmation of any of the expressions defined previously in Table 4.3, using a simple negation module to

Table 4.2: Most frequent keywords and expressions and their collection frequency (cf_t), document frequency (df_t) and number of patients retrieved in the collection of CE CT scan reports performed by actual donors from 2010 to 2015.

Keywords or Expressions	**cf**$_t$	**df**$_t$	**patients**$_t$
contusion	21	18	18
effacement of the sulci	23	22	17
craniectomy	25	14	7
hemorrhagic	30	25	18
subarachnoid hemorrhage	33	32	25
space conflict	37	36	24
hernia	37	26	25
ischemic	37	30	19
midline shift	40	39	29
cerebrospinal fluid	44	40	29
aneurism	46	28	13
parenchyma	46	40	28
hydrocephalus	49	46	30
edema	52	46	34
injuries	52	40	26
hemorrhage	56	40	34
mass effect	62	54	32
hematoma	63	39	29
cerebellar tonsils	66	62	42
heart attack	69	42	24
hypodensity	74	56	36
foramen magnum	76	69	46

disregard negated findings due to the underlying characteristics of medical reporting.

This approach is based on the ConText algorithm, which is, respectively, based on a negation algorithm called NegEx. [71, 86] This is a regular-expression-based algorithm for determining the status of the contextual features in medical reports and has proven to perform very well at identifying negated conditions. [86]

This adapted version of the algorithm takes as input CE CT scan reports and a set of clinical conditions in

Table 4.3: Final set of keywords and expressions.

Keywords or Expressions	
aneurism	injury
contusion	ischemic (or ischemia)
craniectomy	mass effect
edema	midline shift
effacement of the sulci	space conflict
heart attack	space occupying lesion
hematoma	stroke
hemorrhage (or hemorrhagic)	subarachnoid hemorrhage
hernia	trauma
hypodensity	thrombosis
hydrocephalus	ventricular dilatation

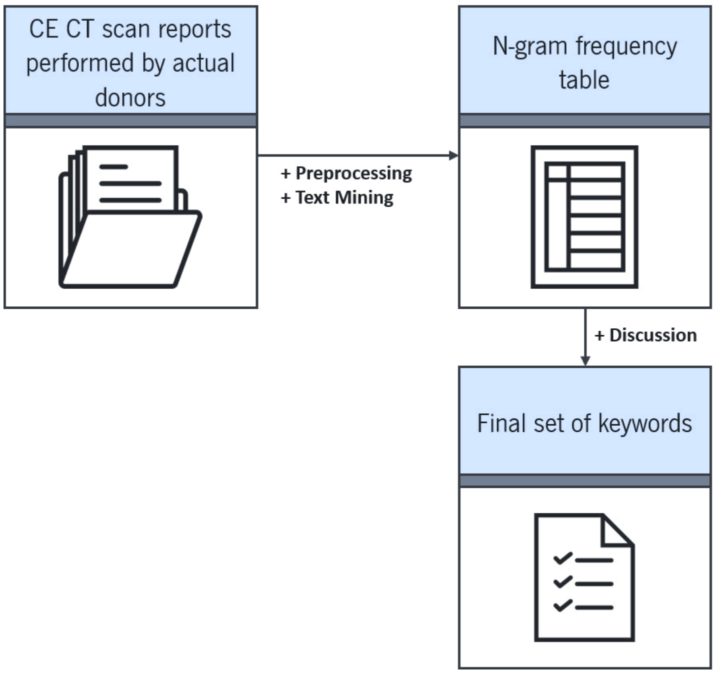


Figure 4.1: Keywords/Expressions extraction module architecture.

the form of keywords and/or expressions as previously defined, and outputs, for each condition, the value for contextual features or modifiers. This adaptation of ConText only determines values for the Negation modifier - affirmed or negated. [86]

By identifying the existence of trigger terms such as “no sign of” and “without”, the algorithm searches for the keywords/expressions that fall within the scope of that trigger term and determines if those clinical conditions should be negated. [86] If the algorithm classifies the match as an affirmation of any of the keywords, it adds the patient’s sequential number and the exam ID to a special table referencing all the detected cases.
Figure 4.2 represents the classification module architecture.

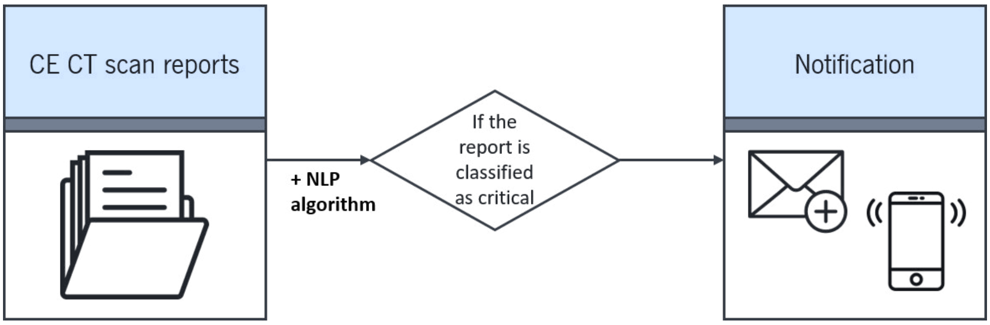


Figure 4.2: Classification module architecture.

This module is particularly notable for its application within the Decision Support System (DSS) described later in this dissertation. By signaling patients with the highest potential of becoming BD donors, an SMS and email alert notification is then forwarded to the GCCT team providing an added value to the system, improving their awareness of patients’ evolution.

4.2.3 Description of the critical pathway

All things considered, it is now possible to portray an algorithm to best describe the potential organ donor profile regarding both the list of ICD-9-CM codes and the set of keywords present in CE CT scan reports.
Figure 4.3 portrays the critical pathway for deceased organ donation suggested using the information from the previous steps.
In an initial phase of **identification**, any patient with a clinical diagnosis consistent with any of the defined in Table 4.1 representing severe brain damage is immediately added to the pool of potential organ donors.
In a subsequent **monitoring** phase, if any of these patients in the actual pool of potential organ donors perform a CE CT scan whose results indicate the affirmation of at least one of the keywords defined in Table 4.3 a notification alert is forwarded to the GCCT team.
Using these criteria, the responsible team at GCCT has the ability to monitor the identified potential organ

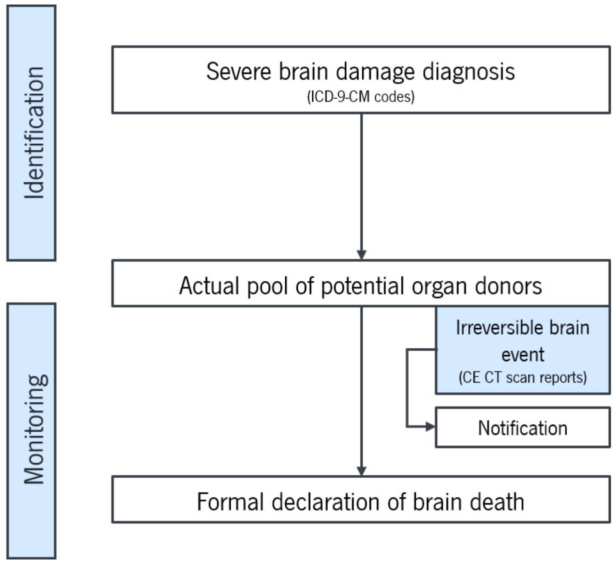


Figure 4.3: The critical pathway for deceased organ donation.

donors more accurately and timely until a formal declaration of brain death occurs. Once a patient is declared BD, the GCCT team must take fast actions to achieve the best results.

4.3 Conclusions

The approach presented above represents a solution which may help characterize the type of patients with a greater probability of evolving into brain death, which can be seen as a precursor for organ donation.

Both the list of ICD-9-CM codes and the set of keywords present in CE CT scan reports identified by the extraction module aim at defining the potential organ donor profile. In the final analysis, both proved to deliver considerable results as they corroborate the work done by the GCCT team.

Event detection is also another powerful feature of this system. By signaling the CE CT scan reports that express any of the predefined clinical conditions, the classification module bridges the gap between clinicians and the reports database. The notification system provides an added value to the system, improving the GCCT team members' awareness of patients' evolution, signaling patients with the highest potential of becoming BD donors, that is, those who need more attention.

Following this systematic approach, it is now possible to automatically acquire, process and deliver useful data to end-users reducing drastically the time between the period of possible brain death recognition and a formal brain death diagnosis, contributing to a valuable method to increase the number of actual donors at

CHP.

The main gains of this approach can, then, be summarized as:

- The definition of criteria for potential organ donors based on ICD-9-CM codes and a set of keywords present in CE CT scan reports;

- The identification of the occurrence of critical events on potential donors' CE CT scan reports and referral of the GCCT team at CHP for closer monitoring, using the criteria defined above.

Regardless of that, selection or rejection of potential donors should not be decided exclusively based on the system's results.

In the future, further knowledge on potential donors' pattern recognition must be assessed in order to serve the main purpose of continuous optimization on their detection. Data mining models concerning new pathways must be addressed in order to explore other targets.

Supplementary testing and evaluation concerning reliability issues and statistical measures like sensitivity and specificity studies must be addressed in order to evaluate the classification module's performance.

Part IV

Design and Implementation

Chapter 5

System Description

5.1 System Overview

The system developed in this project, named **Organite**, attempts to build a Clinical Decision Support System (CDSS) solution to identify and follow the evolution of potential organ donors providing the qualified staff all the tools they need in the decision-making process. Its origin lies in a previously defined system, also called **Organite** with the same purpose, first developed in 2014 also by a MSc in Biomedical Engineering student. To meet the continuous improvement principle, minor errors and limitations found in that previous version will be fixed, and a new set of functionalities and technologies will be deployed.

Such a system proves to be very efficient since the manual process of identification and monitoring of a potential donor is very complex and time-consuming. As data comes from various Health Information Systems (HIS), it may happen that some cases might go unnoticed in the large amount of data.

Through **Organite**, the Gabinete Coordenador de Colheita e Transplantação (GCCT) team is able to access and interact with data from diverse databases and services in the form of a web-based platform using a standard web browser, locally at Centro Hospitalar do Porto, Hospital de Santo António, EPE (CHP).

5.2 System Architecture

Organite, as it is now constructed, follows a REpresentational State Transfer (REST) architecture. This is a client-server web architecture for network applications that partitions tasks offered by the providers of a resource or service - **servers** - and content or service requesters - **clients**. These two communicate over a computer network on separate machines using simple Hyper-Text Transfer Protocol (HTTP) requests initiated by the client and waiting for a response from the server as described in Section 2.6.

The REST design architecture does not require a specific format for the data provided with the requests and/or responses. However, the most common data format used in this dissertation is JavaScript Object Notation (JSON) for both request and response. This data format is defined in the Content-Type header field.

Figure 5.1 depicts the REST architecture followed in this dissertation.

Figure 5.1: The REST architecture.

5.3 System Components

The system's components can be divided into two parts: The back-end and the front-end. These two distinguish the separation of concerns between the data access layer - **back-end** - and the presentation layer - **front-end**.

5.3.1 The Back-end

The back-end usually consists of the server side and, accordingly, contemplates a server, an application and the databases needed to perform actions and pull content and must never be visible to users. In simpler words, the back-end interacts with a server to return user ready results.

Operations on the back-end are complex and consider the development of an application using server-side code which connects with a database to look up, save and/or change data and return it back to the users. The back-end is also concerned with questions of security, structure, and content management.

There are a lot of libraries and frameworks available to make the coding of back-end simpler and improve the system's overall performance. The framework used in this dissertation is Flask, a micro framework for Python.

The back-end follows a Model-View-Controller (MVC) design pattern which divides a web application into three parts as described in Section 2.7.

In this specific scenario, after entering a Uniform Resource Identifier (URI) and calling the associated controller action typically two operations occur: the **Models** (or directly SQLALchemy) are used to retrieve the necessary data from databases, and that data is passed to a **View** which returns the requested data. This retrieved data is generally added to a data structure or a specific data format, like JSON and is accessible within the front-end layer.

Figure 5.2 describes the back-end MVC design pattern.

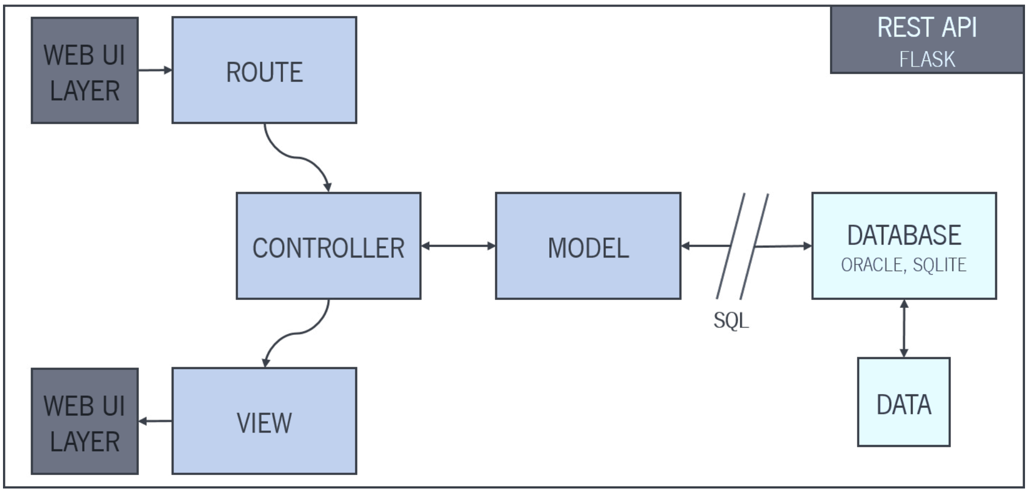


Figure 5.2: The back-end MVC design pattern.

Database technologies

As specified above, the back-end needs databases to get data in and out. As for this dissertation, the database technologies used were Oracle and SQLite. On the one hand, and since the HIS at CHP use Oracle databases to automate their actions, its use was implicit. On the other hand, to store data about the authenticated users and roles, a SQLite database was implemented due to its simplicity and to abstract the user management data from the content data.

On the one hand, Oracle is one of the most common databases technologies in the business world since it provides a wide set of support tools allowing the utilization of not only Structured Query Language (SQL) but also Procedural Language/Structured Query Language (PL/SQL), which is Oracle Corporation's procedural extension for SQL, and has the ability to communicate with diverse systems. [87]

SQL Developer, the Oracle Database Integrated Development Environment (IDE), was used to easily manage the database's objects. It provides powerful editors for working with SQL, PL/SQL, Stored Java Procedures, and eXtensible Markup Language (XML). It is also possible to run queries, execute, debug, etc. [88]

On the other hand, SQLite can not be directly compared to client-server SQL database engines that strive to implement a shared repository of enterprise data, like Oracle does. SQLite aims to provide local storage for individual applications and devices emphasizing economy, efficiency, reliability, independence, and simplicity. Even implemented on a client-server architecture, like this one, SQLite proves to be very efficient as a server-side database since the application content resides on the same machine as the SQLite database. [89]

Programming in Java

The search, filtering (using an inference mechanism), consolidation and Extract, Transform and Loading (ETL) procedures represent one main parcel of the whole system. Using Java, this component is responsible for readying the data to be stored in the potential organ donors repository, using the criteria defined in Section 4.2.

The Java programming language can be described as a concurrent, object-oriented and relatively high-level language which is platform independent meaning that the developed programs can run on any operating system with Java installed.

An IDE was used in the development phase, named Eclipse which combines language support for many languages like Java. Eclipse is open source, both flexible and extensible and has a large community of support helping out solving minor issues.

Even though this component is out of the scope of this dissertation, some progresses and expansions were implemented in order to "keep the engine running", mostly minor corrections on the previous version and the creation of automatic procedures to fulfill the new requirements. [90]

Web Framework: Flask

Choosing a framework for web development using Python can be tricky due to the number of existing technologies available. A web framework is used to ease the developer's life when building reliable, scalable and maintainable web applications helping structure projects so it becomes easier to understand. With this in mind, a web framework provides a set of resources and tools so software developers can build and manage web applications, web services and websites.

Flask proved to be a nimble micro framework that strives to be simple and small. While it may be true, Flask provides extensions to add as many functionalities as the user can think of, such as Object-Relational Mapping (ORM) with SQLAlchemy. SQLAlchemy was used for all interactions with the databases and was chosen due to how simple yet powerfully it is able to carry out complex queries extremely efficiently.

Flask can also be elaborated over a MVC pattern making it easy to develop a complete web application.

As an extensible framework, allowing one to pick packages to extend its functionalities, Flask is not limited to the term «micro». [91]

Flask was chosen to replace Django due to the lightweight nature of the framework. Django has a lot of features out-of-the-box, but everything is built in, contributing to a significant amount of unnecessary content. At the same time, Django is difficult to change and has a steep learning curve while Flask has the ability to add complexity as necessary and users can learn as they go.

Several other frameworks using other languages, like JavaScript (Express.js), were evaluated but Python with Flask proved to be superior in most aspects.

5.3.2 Front-end

The front-end is the part of the code that is usually visible to the users as an interface inviting them to interact with the application. Its main purpose is to interact with the user and present the data in a well-defined style and matter.

The front-end is directly accessed by the user to receive or utilize back-end capabilities, enabling users to access and request the features and services of the underlying information system.

Everything on the front-end is mainly written in HyperText Markup Language (HTML), Cascading Style Sheets (CSS) and JavaScript. In front-end development, there are many libraries and frameworks available to make the coding easier like AngularJS (a JavaScript framework), D3.js (a JavaScript library) and Bootstrap (a front-end framework). In this dissertation, the framework used specifically to describe the front-end behavior is AngularJS and was chosen due to its simplicity and easiness of implementation since AngularJS doesn't require additional plugins or frameworks to build a data-driven web application.

All the points up till now mean that you get to write less code. You don't have to write your own MVC pipeline. The view is defined using HTML, which is more concise. Data models are simpler to write without getters/setters. Data-binding means you don't have to put data into the view manually. Since directives are separate from app code, they can be written by another team in parallel with minimal integration issues. Filters allow you to manipulate data on the view level without changing your controllers. Yes, this is sort of a summary bullet point, but writing less code is a big deal!

Similarly to the back-end, the front-end also assumes a MVC design pattern. When users input a certain URI, AngularJS deals directly with the request and using its routing provider it wires together controllers, view templates and the current URI location in the browser. As the client-side is abstracted from the server-side, AngularJS also takes responsibility for retrieving the contents from the REST Application Programming Interfaces (APIs) that deliver content.

Once the AngularJS application is launched, the model, view, controller and all HTML files are loaded on the user's device and run entirely on the user's hardware. Calls are then made to the back-end REST services, where all business logic and processes are located.

Figure 5.3 describes the front-end MVC design pattern.

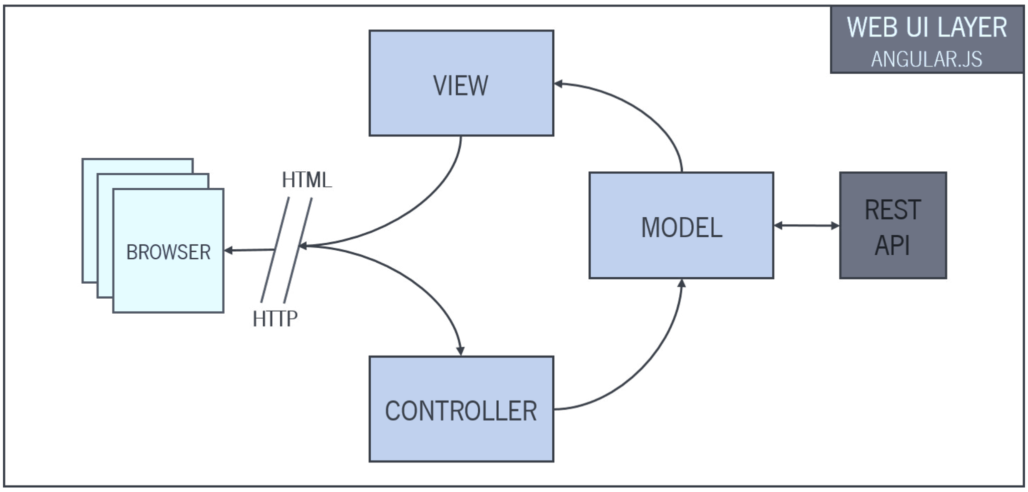


Figure 5.3: The front-end MVC design pattern.

AngularJS

AngularJS is a JavaScript structural framework maintained by Google for dynamic web pages. AngularJS's data binding and dependency injection eliminate much of the code needed to write with its ability to use and extend HTML's language and syntax to express the application's components clearly and succinctly. [92]

Alone, AngularJS provides a complete client-side solution by handling the pages' Document Object Model

(DOM) directly and building the bindings based on Angular-specific element attributes, decoupling it from the application logic, meeting the requirements for decoupling the client-side from the server-side. [92]

Bootstrap

Bootstrap is designed to make front-end faster and easier to almost every type of website, application or device with a single code base. It is made for web development of dynamic and flexible web pages offering a large set of tools to customize how content is displayed to deliver the best user-experience possible. [93]

Its adaptability to different devices makes Bootstrap a unique framework due to the increasing trend to use mobile devices instead of desktop ones, inclusive at hospitals. This particularity avoids having to create different environments for different devices and users can have the same navigational experience whether on the desktop or on their mobile devices.

D3.js

D3.js (Data-Driven Documents) is a JavaScript library used to manipulate documents based on data. With D3.js, it is possible to bind arbitrary data to a DOM and generate HTML tables or, using the same data, create interactive Scalable Vector Graphics (SVG) bar charts. [94]

NVD3 is a light collection of reusable graphic components based on D3.js simplifying its implementation without losing its potential. NVD3's integration is simpler, more intuitive and proves to be quite efficient when customizing the reporting and data visualizations, especially in the form of dashboards. [95] At this stage, the need to adopt a robust Business Intelligence (BI) tool to provide better functionalities for reporting and Online Analytical Processing (OLAP) was not felt, however, when proceeding into a more complex stage, this will probably make sense.

To integrate the chart components of NVD3 into AngularJS's directives a module named Angular-nvD3 was used.

Chapter 6

Implementation

The implementation of web applications to integrate information and other resources from the various data sources and to make them available through a unique, user-friendly interface is more than accepted in organizations. [18]

Together, these web applications can support the responsible teams in their daily work, especially supporting their decision-making processes, applying the concept of Business Intelligence (BI). That said, the aim of this web application is to represent data in a way the Gabinete Coordenador de Colheita e Transplantação (GCCT) team can easily identify key factors and, ultimately, perform more targeted actions.

The result must be a high-level solution ready to answer the delineated objectives and provide a fully operational decision support tool in the transplantation field.

The BI's support infrastructure architecture follows the approach proposed by Han and Kamber [52] and Eckerson [53] composed of three levels: The Data Sources level, the Data Warehouse level, and the Business Analytics level, as described in Section 2.3. A brief explanation of the main components will be provided in the next sections as a complete description is already available. [23]

6.1 Data Sources

The Data Sources represent all the operational environment of the application. These Data Sources are represented as databases including the data from different operational systems, responsible for feeding the Data Warehouse.

After studying the available databases ate Centro Hospitalar do Porto, Hospital de Santo António, EPE (CHP), one central data source proved to be the main data source, Processo Clínico Eletrónico (PCE), and two additional, SClínico and Emergency System, as auxiliary data sources.

The inclusion of the Emergency System is more than justified by the fact that a high number of patients with devastating injuries enter the hospital through the Emergency Room (ER) and only later on they are effectively registered in the PCE, in case they are admitted.

The SClínico system comprises the former Sistema de Apoio Médico (SAM) and Sistema de Apoio à Prática de Enfermagem (SAPE) which contain data from both the medical clinical and nursing areas, providing an essential source of diagnostics.

PCE works as an interoperability key between all the data sources and, analogous to the previous, works as an essential source of diagnostics relative to admissions. Likewise, PCE also aggregates data about the patient, including unique association identifiers in order to connect with the remaining systems enabling real-time monitoring of patients.

Data from these databases are loaded into the Data Warehouse through Application Programming Interfaces (APIs). This system in particular utilizes Open Database Connectivity (ODBC) to enable data access from data sources and its posterior download.

6.2 Data Warehouse

The second level of the BI architecture defines both the Data Staging Area and the Data Warehouse.

The first plays an essential role in validating, cleaning and working over the data previously extracted from the Data Sources, also known as Extract, Transform and Loading (ETL) procedures. [18] In this dissertation, a Java application is responsible for this step.

Above the Data Staging Area, the Data Warehouse represents the analytical repository where the prepared data from the previous steps is stored. It is atop this repository that all the services with the purpose to deliver information to end users are implemented. [18]

As the creation and management of the repository where the data is stored are out of the scope of this dissertation, only its architecture will be described for the sake of clarity and a proper understanding of the system's behavior:

- **ID**: Unique incremental identifier for every entry on the table;
- **Episode ID**: Unique identifier for every clinical episode. A clinical episode is defined as a set of events registered on a certain clinical context since a patient is admitted until the patient's medical discharge. Multiple episodes can share the same ID, whereas belonging to different modules;
- **International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) Code**: ICD-9-CM code of the registered diagnosis. During an episode, different diagnosis can be assigned, each representing a new table entry;
- **Notes**: Diagnosis description, medical notes, nursing notes, commentaries and descriptions of the patient's state, usually recorded when a new ICD-9-CM code is registered;
- **Date of diagnosis**: Date when the diagnosis was assigned;
- **Date found**: Date when the diagnosis was found by the system and recorded into the repository;
- **Professional ID**: Sequential number that unequivocally identifies the medical professional who registered the diagnosis;
- **Module code**: Clinical module where the the episode took place;
- **Patient's Sequential Number**: Along with the **Process Number**, unequivocally identifies each patient and his/her process number;
- **Specialty code**: Code of the medical specialty where the episode took place. A column representing the **Specialty description** (full name) is also available;
- **Date of birth**: Patient's date of birth;
- **Gender**: Patient's gender;
- **Status**: Medical condition of the patient at present time that can be **Deceased**, **Admitted**, **Admitted in the ER**, **Medical Discharge**, **Unknown** or **Entry**;

- **Recent diagnosis**: Working as a boolean parameter, identifies the most recent diagnosis for each patient as 1, and 0 otherwise.

It should be noted that the **Status** and **Recent diagnosis** parameters represent a novelty introduced to ease some of the procedures featured.

6.3 Business Analytics

The third and last level represents the technologies that support real-time analytics, enabling end-users to interact and explore the data from the repository previously created. Its objective lies on providing results to end-users through a set of visualization tools from reporting to decision support web applications. [18]
In this dissertation, and due to how this type of data must be analyzed, a set of tables and graphical elements for statistical means was developed.

6.3.1 Representing the potential donors repository

This feature is particularly critical since it provides a way of displaying a list of all patients identified by the system as possible organ donors. This section displays every patient only once, with data from the patient's most recent diagnosis. This endows the platform with real-time monitoring capabilities providing the responsible members a quick-view on the patients' condition.
In order to meet the required specifications, two sections were built:

- **A history of potential donors**: This section contains all the potential donors identified by the platform since its implementation;

- **A selection of current potential donors**: This section filters the results to show only the applicable identified cases.

History of potential donors

This section presents all the potential donors identified by the platform since its implementation in the form of a table with relevant data to accurately identify and access the patients' details.
The set of attributes available in this first table includes the patients' Sequential Number (Número Sequencial), his/her most recent diagnosis' Clinical Episode ID (ID Episódio), Notes (Anotações), Specialty (Especialidade)

and Date (Data do Diagnóstico). A column showing the patients' Medical Condition (Estado) is also available and can be:

- **Deceased (Faleceu)**: If the patient is dead;
- **Admitted (Internado)**: If the patient is currently admitted in the hospital;
- **Admitted in the ER (Urgências)**: If the patient is currently admitted in the ER;
- **Medical Discharge (Alta)**: If the patient has already had medical discharge;
- **Unknown (Desconhecido)**: If the patient's condition is none of the above and can't be correctly determined;
- **Entry (Entrada)**: Before the automatic procedure determine the patient's medical condition from any of the above, this parameter is identified as an **Entry**.

On the table, each medical condition has a color to ease the identification process.

To provide this section with some input options from the user, it is also possible to order the results based on the Clinical Episode ID, Specialty, Date of Diagnosis or Medical Condition's column and filter the results shown based on the Clinical Episode ID, Date of Diagnosis and patients' Sequential Number input data from users.

Figure 6.1 shows an overview of this page.

This page is purposed to provide the medical team all the means on the background of the system for statistical and learning purposes.

A fully detailed view of each patient can also be accessed from this page by clicking over the patient's Sequential Number.

Current potential donors

This section presents all the applicable identified potential donors, also in the form of a table with the same set of attributes and features.

This page concatenates the actual cases that the transplant team must considerate, that is, it excludes the patients that had deceased or have had medical discharge.

Figure 6.1: History of potential donors' page.

This page plays an essential role in the whole platform since it includes the most probable cases of becoming an actual organ donor.

Figure 6.2 shows an overview of this page

Figure 6.2: Current potential donors' page.

The detailed view of each patient can also be accessed from this page by clicking over the patient's Sequential Number.

Detailed information on potential donors

After analyzing the list of potential donors, there comes the need to access detailed aspects on the patient's condition. Some necessary parameters like the patient's medical and clinical tests are very relevant to determine the patient's potential of becoming an actual donor.

To meet these requirements, a detailed view of each patient was created to answer some unanswered questions.

The list of fields about the patient that this view offers includes:

- Sequential Number (Número Sequencial);
- Process Number (Número de Processo);
- Gender (Género);
- Date of Birth (Data de Nascimento);
- Medical Condition (Estado);
- Most Recent Diagnosis (Diagnóstico mais recente);
- PCE;

The Medical Condition field holds some additional information for each situation. In case the patient had **Deceased**, the date of death, medical specialty and module where it occurred are displayed.

In case the patient has had **Medical Discharge**, the discharge date is displayed.

In case the patient is **Admitted** or **Admitted in the ER**, the admission date, medical specialty, room and bed where the patient is admitted are displayed.

In case the patient's medical condition is **Unknown** or **Entry**, no further details are provided.

The PCE field represents a link to the CHP's electronic health record, providing a full detailed view of the patient's details along with his/her course in the hospital.

Three tabs displaying all the identified diagnosis, medical tests and clinical tests of the referred patient are also available allowing the GCCT team to access this type of information without needing to resort to other systems. A link to the medical and clinical test reports is also available.

Figure 6.3 shows an overview of this page.

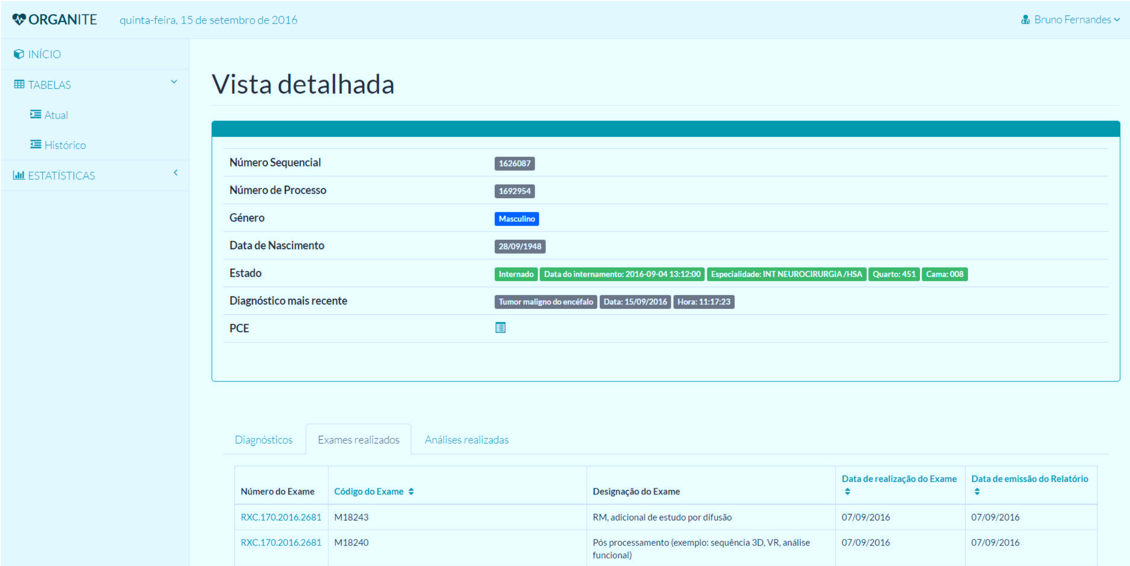


Figure 6.3: Detailed view on a patient.

The full representation of the elements of this page can be found on Appendix A.

6.3.2 Representing the BI indicators

To provide an overview of the data collected by the system, a set of graphical elements was developed in order to facilitate data analysis by the GCCT team.

Assuming the form of dashboards, these elements contain data aggregated into different perspectives revealing the most important aspects to be analyzed.

At the same time, this section offers the possibility to generate simple, downloadable reports in the form of text files that the GCCT team can utilize for further offline analysis.

Statistical Graphics

This section offers the following set of dashboards:

- Counter of all the identified cases in the last 24 hours;
- Bar chart representing the total number of entries per clinical module within the last week;

- Interactive radial chart representing the total number of entries per clinical module, specialty and ICD-9-CM code;
- Donut chart representing the total relative distribution of entries per medical specialty;
- Stacked area chart representing the total number of entries per clinical module, within the last year, grouped by month. This graphic offers the possibility to alternate between the stacked, stream and expanded views.

If correctly analyzed, the data provided by these elements can help develop new methodologies and techniques on how to detect and monitor potential organ donors, since it is possible to identify, for example, the clinical areas with most identified cases. This feature can be essential in developing an ever-improving potential organ donor identification criteria.

Figure 6.4 shows an overview of this page as the complete set of dashboards can be found in Appendix B.

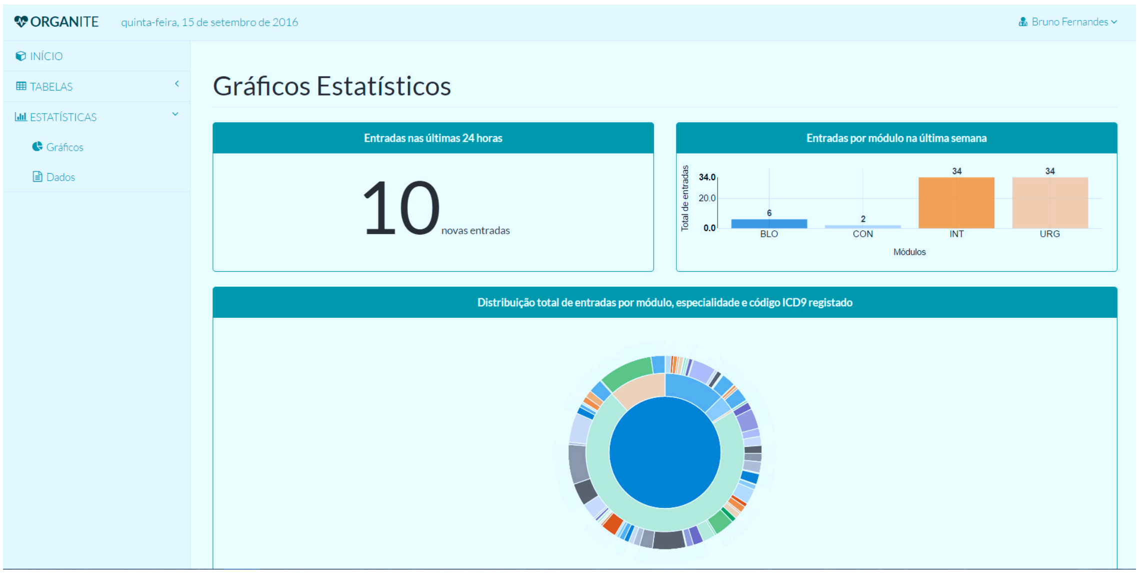


Figure 6.4: Overview of the statistical graphics page.

The counter of all the identified cases in the last 24 hours serves the purpose of counting the total number of patients identified in the last day as to easily highlight those who were recently added to the system.

The bar chart representing the total number of entries per clinical module within the last week enables a fast understanding on which clinical modules are responsible for the largest number of identified cases within the last week. As this value can change from week to week it was decided it would be convenient to display this information in this way.

Another way of displaying information takes the form of a radial chart to represent the total number of entries per clinical module, specialty and ICD-9-CM code which adds interactivity to the way this information is displayed by enabling users to progress in the medical specialties and ICD-9-CM codes by clicking on them. Users can, then, get complete insights on which clinical areas are more contributive to the number of potential organ donors.

Following the same purpose, the donut chart representing the total relative distribution of entries per medical specialty adds similar knowledge to the whole system.

For means of getting a monthly overview on how many patients were identified in the last year per clinical module, the area chart represents these values in such a way that it is possible to easily acknowledge the months of the year where there was a greater number of identified cases.

Statistical Data

Another feature also available in **Organite** provides the GCCT team with specific data organized into yearly patient-centered categories. This component also complements the system with the ability to generate downloadable, text-based reports, since there is an option to download a text file with the required data.

The data represented in this section is organized on a table with drop-down buttons to ease data visualization and understanding. Considering that the data relates to the year the user must preselect, the complete set of parameters represented is:

- Total number of entries (Total de entradas);
- Average number of entries per day (Média de entradas por dia);
- Total number of entries per month (Total de entradas por mês);
- Total number of entries per gender (Total de entradas por género);
- Total number of entries per age group (Total de entradas por faixa etária). This parameter is divided into four age groups: 0-18; 19-35; 36-50 and 50+.

Figure 6.5 represents a view on this page.

Figure 6.5: Statistical data on identified cases, per year.

This set of parameters complements the information provided in the Statistical Graphics section. Users can now get information on the total number of entries verified and compare the values among the different years and months within years to understand the evolution in the number of identified cases.

It is now also possible to recognize which age group has a greater contribute as well as which gender dominates the number of identified cases.

In general, users can now make sense of a large amount of data and, through this indicators, get useful insights and define better strategies to attain the best results.

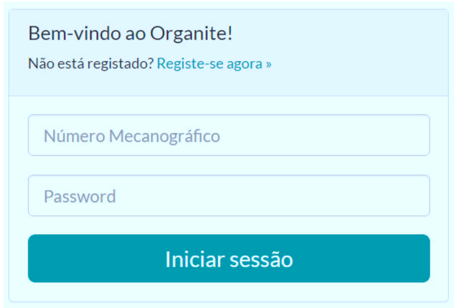
6.3.3 Users management

Organite also offers the ability to register and login users into the system, meeting the authentication requirements for such a critical platform.

At the registry, users must fill a form providing their Mechanographical Number (Número Mecanográfico), Name (Nome), Email Address, Phone Number (Telemóvel) and Password. The Mechanographical Number and Password serve for login purposes.

Figure 6.6 represents both the Registry and Login forms.

(a) Registry form.



(b) Login form.

Figure 6.6: (a) Registry and (b) Login forms for users authentication.

To provide a certain level of authorization and permissions definition, two authorization roles were defined: **Administrator** and **User**.

Administrators are given full use of the platform access methods and advanced features. This includes the ability to delegate other users' participation and permissions on the platform. After registering, Users must wait until an administrator confirms their registry and assign their role. Administrators can also delete and edit accounts as necessary.

Figure 6.7 represents a view on the Users Management page.

Administrators can also view the full list of ICD-9-CM codes that the system is using to collect potential organ donors.

Users, on the other hand, are given access to the most basic features offered by the platform. Their participation is limited and always monitored and defined by administrators of the platform. Also, Users must be authenticated by an Administrator in order to be able to login. They are not granted access to any of the patients lists nor the Statistical Data page and can only access the home screen and Statistical Graphics pages.

Figure 6.7: Users Management page.

6.4 Servers and Hosting

The web platform is hosted on a hospital server and can be accessed through the following address at the CHP's intranet: http://172.21.201.151/organite. This server is running Ubuntu 14.04.5 LTS (Trusty Tahr). The Java procedures responsible for inserting and updating data in the repository are running on a different server - hsa-siima - as well as the procedures responsible for detecting the occurrence of critical events for notification. This server is running Windows 7.

Chapter 7

Proof of Concept

Every change to an Information Technology (IT) project should be tested before delivered to end-users. It is important to verify that certain concepts, technologies, and methodologies have the potential for real-world application and in IT, the term is usually associated with the development of a prototype as a tool to prove feasibility of an IT infrastructure. [96, 97]

As for the results of a proof of concept it is possible to evaluate possible structural changes and project specifications in order to correct eventual problems before the final version of the product. [96, 97]

To meet these requirements, a SWOT analysis was performed and a final conclusion was drawn. The SWOT analysis works as a structured planning method that evaluates **S**trengths, **W**eaknesses, **O**pportunities and **T**hreats of a project. In this particular case, the SWOT analysis consists of a monitoring process of the platform developed and all of its features in order to identify internal and external factors that are favorable and unfavorable to achieve its purposes.

Table 7.1 exposes the SWOT analysis performed on the platform.

In fact, the platform is capable of increasing the potential of organ donors with the help of knowledgeable medical professionals who can manipulate its features and take actions.

Organite offers a light environment without any major slow loadings integrated on a clean, responsible and intuitive layout. When it comes to web platforms, bigger is not always better. Simplicity is a valued quality, given that the side panel must serve a practical purpose.

Data dispersed over several systems is now synthesized on a single place enabling users to gain a merged

Table 7.1: SWOT analysis on the platform.

Strengths	**Weaknesses**
+ Shared resources amongst different platforms; + Interactive and intuitive navigation and search; + Useful, relevant and goal-oriented content; + Interoperation of data; + Easy maintenance. + Security and confidentiality provided by reliable authentication levels.	+ Intranet connection required; + Overloaded servers when there are frequent simultaneous or complex client requests; + Since the platform features a centralized architecture if a critical server fails, client requests are not accomplished.
Opportunities	**Threats**
+ Implementation of new functionalities; + New technologies to improve user experience; + Continued use may leverage the increase in the number of actual transplants.	+ Changes on the database organizational schema may result in an unpredictable behavior of the platform; + Difficult acceptance and adaptability by healthcare professionals to new, technological platforms.

view of the relevant data needed for decision-making.

It's also easy to replace, repair, upgrade and relocate a server while clients remain unaffected since the system represents a distributed model with dispersed responsibilities among independent computers integrated across a network. This unawareness of change is called encapsulation.

Despite that, there is still room for improvements by implementing new functionalities and technologies to excel user experience. In the end, the number of actual transplants may, at best, increase.

However, the system still displays some weaknesses such as slow loading speeds when frequent simultaneous or complex client requests are made, or the lack of distributed nodes to assure the system still runs when a critical server fails.

Also, and as expected, database organizational schema changes may alter the system's behavior. On the other hand, and ultimately, it is of extreme difficulty to educate and persuade healthcare professionals to use technological platforms which becomes a barrier to any IT solution.

Part V

Conclusions

Chapter 8

Conclusions

In this concluding chapter the project's results are discussed in Section 8.1. The contributions of this dissertation and their possible impact are summarized in Section 8.2 and the important directions for further research are discussed in Section 8.3.

8.1 Discussion

The central problem addressed in this dissertation is the low transplantation rates verified at Centro Hospitalar do Porto, Hospital de Santo António, EPE (CHP). As so, this dissertation's main purpose was to develop a prototype of a decision support platform in the transplantation field in order to provide reliable and timely information to complement the decision-making process.

This project enabled the exploration of concepts like Business Intelligence (BI) and Knowledge Discovery to introduce a valuable solution that may contribute to the continuous optimization of such a crucial and sensitive field as the organ transplantation is. As so, it is a solution in constant development and innovation that should meet the field demands. According to the Design Science Research (DSR) methodology adopted in this project, Table 8.1 summarizes the work at the different steps.

Following this research methodology and after stating the research problem and questions, it was possible to define multiple objectives in order to answer the questions raised in Chapter 1 aiming at acquiring knowledge on how to design and develop a high-level solution to improve the Gabinete Coordenador de Colheita e Transplantação (GCCT) team results at CHP.

Table 8.1: Summary of the work at the different DSR steps.

DSR steps	Activities
1. Problem Identification and Motivation	+ Problem definition; + Classification and description of the needs observed in the transplantation field; + Review of the most consistent and suitable methods and technologies to solve the defined aims.
2. Objectives of the Solution	+ Definition of the Text Mining approaches to assess the potential organ donor profile; + Implementation of minor corrections on the Extract, Transform and Loading (ETL) process; + Construction of a Decision Support System (DSS) solution based on BI to interpret and demonstrate the data needed for decision-making.
3. Design and Development	+ Implementation and interpretation of the Text Mining techniques; + Platform design and development.
4. Demonstration and Evaluation	+ Definition of effective potential organ donor criteria; + Creation of the platform prototype.
5. Communication	+ Academic publishings; + Evidence of the work developed; + Delivery of the solution to end-users.

An artifact was then designed and implemented, in this particular case, the functional prototype of the system, and further evaluated according to criteria implicit in the proposal. This includes both the definition of effective potential organ donor criteria and the web-based platform.

The definition of effective potential organ donor criteria demonstrates the power of analyzing real data from actual organ donors by applying Text Mining to quickly identify keywords from a large quantity of textual data. This helps to answer the question of the eligibility criteria of a potential organ donor.

In what concerns the repository of potential organ donors, the implementation of correction measures and new filters proved to meet the objectives of continuous improvement and refinement of the results obtained. The web-based platform was a natural evolution to the existing one by adopting new, simpler and more efficient technologies to make it lighter and cleaner with AngularJS and Flask which demonstrated to respond better and faster than Django.

Lastly in the DSR methodology, the results achieved were communicated through academic publishings and the final solution was delivered to end-users. The resulting strategy utilizes these outcomes and provides an actionable plan to increase productivity and provide a greater service meeting the objectives stated at first and answering the raised research problem.

8.2 Contributions

The system developed represents an innovative way of supporting decision making in the transplants environment. Following this systematic approach, it is now possible to automatically acquire, process and present useful data to end-users reducing drastically the time between the period of possible brain death recognition and a formal brain death diagnosis, contributing to a valuable method to increase the number of actual donors.

The Knowledge Discovery in Databases (KDD) component of this dissertation proved to be a very reasonable method for defining the potential organ donor criteria based on Cranio-Encephalic Computerized Tomography (CE CT) scan reports. Taking real clinical information from actual donations from the past five years as input and applying more or less complex text mining procedures, the output is likely to produce more realistic predictive models.

Unlike other similar tools, this type of screening system complements itself with the two-step mechanism. It takes all patients diagnosed with any of the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) codes defined, centralizes that information and displays it in a graphical interface specially designed to allow a fast, interactive understanding by end-users.

At the same time, end-users can now follow-up potential organ donors along with their course in the hospital having all the data they need in a single site, accessible anytime and anywhere inside the hospital's intranet. The notification system improves the GCCT team members' awareness of patients' evolution, signaling patients with the highest potential of becoming Brain Dead (BD) donors, that is, those who need more attention and may, therefore, result in actual organ donors.

The clinical indicators in the form of graphics constitute an effective method that can contribute, through evidence, for an improved clinical decision-making.

The main contributions of this dissertation can, then, be summarized as:

- The definition of more accurate potential organ donor selection criteria;
- The construction, using the selection criteria model, of a potential donors pool;
- The follow-up of potential donors through a structured interface anytime and anywhere inside the hospital's intranet;
- The notification of the occurrence of devastating neurological events on potential donors CE CT scan reports.

Through the presentation of **Organite** together with the GCCT team at CHP it has been proved that the adoption of a technological tool to complement the medical team's traditional methods may be more satisfactorily accepted if end-users acknowledge the project's objectives and the gains it may bring for their daily work. The fact that some members might not be inclined to change their professional routines may constitute a limitation for the implementation of such solutions. If these issues are not resolved the systems run the risk of being rejected and fall into disuse.

8.3 Future Work

The innovative feature of this project reflects its ability to evolve in virtually every direction as there will always be space for improvements besides the required continuous maintenance and the manifestation of new clinical circumstances.

Further knowledge on predictive indicators that evidence a patient's potential for organ donation must be assessed in order to upgrade the potential organ donor criteria. Also, Data Mining (DM) models concerning new pathways must be addressed in order to explore other targets.

It would be interesting to explore more clinical records to assess the patient's probability of evolving into an actual organ donor such as the Glasgow Coma Scale (GCS), pupil reactivity, amongst others. The collection of such data would be of great help if combined with intelligence algorithms to identify patterns and trends in the patient's records.

It would also be relevant to implement a logistics module responsible for implementing either an exclusion criteria model and the donor/recipient pair selection as stated by the Portuguese law and the hospital's principles. This would include contraindications for donation such as risk behaviors and the presence of a potentially transmissible infectious disease. Also, if the potential donor has registered in life in Registo Nacional de Não Dadores (RENNDA), his/her contribution must be rejected.

Furthermore, it may also have considerable impact in donation if the platform offered any kind of donor/recipient pair selection tools that would allow the input of some data concerning the recipient's clinical tests (blood type, HLA typing and crossmatching) and it would output the best matched potential organ donors organized in some way.

Additionally, a number of relevant trials should be addressed together with the GCCT team at CHP in order to adapt the platform to their needs and correct any eventual limitations concerning the quality of information, time in getting information, time in decision-making, multi-query ability, graphical aspect of the platform and system usability. This type of trials could easily fit in the form of questionnaires.

Bibliography

[1] C Rudge, R Matesanz, FL Delmonico, and J Chapman. International practices of organ donation. *British journal of anaesthesia*, 108(suppl 1):i48–i55, 2012.

[2] Rafael Matesanz. A decade of continuous improvement in cadaveric organ donation: the spanish model. *Nefrologia*, 21(Suppl 5):59–67, 2001.

[3] Yorick J de Groot, Nichon E Jansen, Jan Bakker, Michael A Kuiper, Stan Aerdts, Andrew IR Maas, Eelco FM Wijdicks, Hendrik A van Leiden, Andries J Hoitsma, Berry HPH Kremer, et al. Imminent brain death: point of departure for potential heart-beating organ donor recognition. *Intensive care medicine*, 36(9):1488–1494, 2010.

[4] AP Fernandes, A Gomes, J Veiga, D Ermida, and T Vardasca. Imaging screening of catastrophic neurological events using a software tool: Preliminary results. In *Transplantation proceedings*, volume 47, pages 1001–1004. Elsevier, 2015.

[5] J Erwin, O Kompanje, Nichon E Jansen, and Yorick J de Groot. 'in plain language': uniform criteria for organ donor recognition. *Intensive care medicine*, 39(8):1492, 2013.

[6] Rafael Matesanz, Beatriz Domínguez-Gil, Elisabeth Coll, Gloria de la Rosa, and Rosario Marazuela. Spanish experience as a leading country: what kind of measures were taken? *Transplant International*, 24(4):333–343, 2011.

[7] Tianxia Gong, Chew Lim Tan, Tze Yun Leong, Cheng Kiang Lee, Boon Chuan Pang, CC Tchoyoson Lim, Qi Tian, Suisheng Tang, and Zhuo Zhang. Text mining in radiology reports. In *2008 Eighth IEEE International Conference on Data Mining*, pages 815–820. IEEE, 2008.

[8] Uzma Raja, Tara Mitchell, Timothy Day, and J Michael Hardin. Text mining in healthcare. applications and opportunities. *J Healthc Inf Manag*, 22(3):52–56, 2008.

[9] Xiaohua Zhou, Hyoil Han, Isaac Chankai, Ann Prestrud, and Ari Brooks. Approaches to text mining for clinical medical records. In *Proceedings of the 2006 ACM symposium on Applied computing*, pages 235–239. ACM, 2006.

[10] José Machado, Victor Alves, António Abelha, and José Neves. Ambient intelligence via multiagent systems in the medical arena. *Engineering intelligent systems for electrical engineering and communications*, 15(3):151–157, 2007.

[11] Luciana Cardoso, Fernando Marins, César Quintas, Filipe Portela, Manuel Santos, António Abelha, and José Machado. Interoperability in healthcare. *Cloud Computing Applications for Quality Health Care Delivery*, pages 78–101, 2014.

[12] Miguel Miranda, Júlio Duarte, António Abelha, José Manuel Machado, José Neves, and João Neves. Interoperability in healthcare. 2010.

[13] Hugo Peixoto, Manuel Santos, António Abelha, and José Machado. Intelligence in interoperability with aida. In *International Symposium on Methodologies for Intelligent Systems*, pages 264–273. Springer, 2012.

[14] Maribel Yasmina Santos and Isabel Ramos. *Business Intelligence: Tecnologias da informação na gestão de conhecimento*. FCA-Editora de Informática, 2006.

[15] Efraim Turban, Ramesh Sharda, and Dursun Delen. *Decision support and business intelligence systems*. Pearson Education India, 2011.

[16] Jungpil Hahn and Mani R Subramani. A framework of knowledge management systems: issues and challenges for theory and practice. In *Proceedings of the twenty first international conference on Information systems*, pages 302–312. Association for Information Systems, 2000.

[17] Maryam Alavi and Dorothy E Leidner. Review: knowledge management and knowledge management systems: conceptual foundations and research issues. *Knowledge Management: Critical Perspectives on Business and Management*, 3(1):163, 2005.

[18] Bruno Cortes. Sistemas de suporte à decisão. *FCA-Editora Informática, Lisboa R Modelo de Gestão da Rota do Românico do Vale do Sousa | Rosário Correia Machado*, 2005.

[19] E Turban, RK Rainer, and RE Potter. Introduction to information technology, 2005.

[20] Dinu Airinei and Daniel Homocianu. Dss vs. business intelligence. *Revista Economica*, 2009.

[21] Mihaela Ivan and Manole Velicanu. Healthcare industry improvement with business intelligence. *Informatica Economica*, 19(2):81, 2015.

[22] Coordenação Nacional da Transplantação. Resumo da atividade nacional, janeiro-junho 2016. Technical report, Instituto Português do Sangue e Transplantação (IPST, IP), 7 2016.

[23] Luís Torres. Plataforma de suporte à decisão médica para transplante de órgãos baseada em business intelligence. Master's thesis, Universidade do Minho, Outubro 2015.

[24] Ashis Jalote-Parmar. *Workflow Driven Decision Support Systems: A case of an intra-operative visualization system for surgeons*. TU Delft, Delft University of Technology, 2009.

[25] Luca Palazzo, Paolo Sernani, Andrea Claudi, Gianluca Dolcini, Gianluigi Biancucci, and Aldo Franco Dragoni. A multi-agent architecture for health information systems. In *International Workshop on Artificial Intelligence and NetMedicine*, page 41, 2013.

[26] Luciana Cardoso, Fernando Marins, Filipe Portela, Manuel Santos, António Abelha, and José Machado. The next generation of interoperability agents in healthcare. *International journal of environmental research and public health*, 11(5):5349–5371, 2014.

[27] Soudabeh Khodambashi. Business process re-engineering application in healthcare in a relation to health information systems. *Procedia Technology*, 9:949–957, 2013.

[28] Reinhold Haux, Alfred Winter, Elske Ammenwerth, and Birgit Brigl. *Strategic information management in hospitals: an introduction to hospital information systems*. Springer Science & Business Media, 2013.

[29] Júlio Duarte, Carlos Filipe Portela, António Abelha, José Machado, and Manuel Filipe Santos. Electronic health record in dermatology service. In *International Conference on ENTERprise Information Systems*, pages 156–164. Springer, 2011.

[30] Ricardo Costa, Paulo Novais, José Machado, Carlos Alberto, and José Neves. Inter-organization cooperation for care of the elderly. In *Integration and Innovation Orient to E-Society Volume 2*, pages 200–208. Springer, 2007.

[31] Luciana Cardoso, Fernando Marins, Filipe Portela, António Abelha, and José Machado. Healthcare interoperability through intelligent agent technology. *Procedia Technology*, 16:1334–1341, 2014.

[32] Basit Chaudhry, Jerome Wang, Shinyi Wu, Margaret Maglione, Walter Mojica, Elizabeth Roth, Sally C Morton, and Paul G Shekelle. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. *Annals of internal medicine*, 144(10):742–752, 2006.

[33] Miguel Miranda, José Machado, António Abelha, and José Neves. Healthcare interoperability through a jade based multi-agent platform. In *Intelligent Distributed Computing VI*, pages 83–88. Springer, 2013.

[34] Filipe Portela, Alexandra Cabral, António Abelha, Maria Salazar, César Quintas, José Machado, and MF Santos. Knowledge acquisition process for intelligent decision support in critical health care. *Healthcare Administration: Concepts, Methodologies, Tools, and Applications: Concepts, Methodologies, Tools, and Applications*, 270, 2014.

[35] Miguel Miranda, Júlio Duarte, António Abelha, José Manuel Machado, José Neves, and João Neves. Interoperability in healthcare. 2010.

[36] António Abelha, Eliana Pereira, Andreia Brandão, Filipe Portela, Manuel Filipe Santos, José Machado, and Jorge Braga. Improving quality of services in maternity care triage system. *International Journal of E-Health and Medical Communications (IJEHMC)*, 6(2):10–26, 2015.

[37] Júlio Duarte, Carlos Filipe Portela, António Abelha, José Machado, and Manuel Filipe Santos. Electronic health record in dermatology service. In *International Conference on ENTERprise Information Systems*, pages 156–164. Springer, 2011.

[38] José Manuel Machado, António Abelha, Manuel Santos, and José Neves. Multiagent based problem solving in medical decision support systems. In *2nd International Conference on Knowledge Engineering and Decision Support, Knowledge and Decision Technologies*, 2006.

[39] Jose Machado, António Abelha, José Neves, and Manuel Santos. Ambient intelligence in medicine. In *2006 IEEE Biomedical Circuits and Systems Conference*, pages 94–97. IEEE, 2006.

[40] Markéta Škrobáčková and Hana Kopáčková. Decision support systems or business intelligence: what can help in decision making? *Scientific papers of the University of Pardubice. Series D, Faculty of Economics and Administration. 10 (2006)*, 2006.

[41] Öykü Işık, Mary C Jones, and Anna Sidorova. Business intelligence success: The roles of bi capabilities and decision environments. *Information & Management*, 50(1):13–23, 2013.

[42] Osama T Ali, Ali Bou Nassif, and Luiz Fernando Capretz. Business intelligence solutions in healthcare a case study: Transforming oltp system to bi solution. In *Communications and Information Technology (ICCIT), 2013 Third International Conference on*, pages 209–214. IEEE, 2013.

[43] Kensaku Kawamoto, Caitlin A Houlihan, E Andrew Balas, and David F Lobach. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *Bmj*, 330(7494):765, 2005.

[44] Tiffani J Bright, Anthony Wong, Ravi Dhurjati, Erin Bristow, Lori Bastian, Remy R Coeytaux, Gregory Samsa, Vic Hasselblad, John W Williams, Michael D Musty, et al. Effect of clinical decision-support systems: a systematic review. *Annals of internal medicine*, 157(1):29–43, 2012.

[45] Steven Shea, William DuMouchel, and Lisa Bahamonde. A meta-analysis of 16 randomized controlled trials to evaluate computer-based clinical reminder systems for preventive care in the ambulatory setting. *Journal of the American Medical Informatics Association*, 3(6):399–409, 1996.

[46] E Andrew Balas, Scott Weingarten, Candace T Garb, David Blumenthal, Suzanne Austin Boren, and Gordon D Brown. Improving preventive care by prompting physicians. *Archives of Internal Medicine*, 160(3):301–308, 2000.

[47] Dereck L Hunt, R Brian Haynes, Steven E Hanna, and Kristina Smith. Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review. *Jama*, 280(15):1339–1346, 1998.

[48] Eta S Berner. Clinical decision support systems: state of the art. *AHRQ publication*, 90069, 2009.

[49] William F Cody, Jeffrey T Kreulen, Vikas Krishna, and W Scott Spangler. The integration of business intelligence and knowledge management. *IBM systems journal*, 41(4):697–713, 2002.

[50] Celina Olszak, Kornelia Batko, et al. Business intelligence systems. new chances and possibilities for healthcare organizations. *Informatyka Ekonomiczna*, (25):123–138, 2012.

[51] Adela Bâra and Ion Lungu. *Improving Decision Support Systems with Data Mining Techniques.* INTECH Open Access Publisher, 2012.

[52] Han Jiawei and Micheline Kamber. Data mining: concepts and techniques. *San Francisco, CA, itd: Morgan Kaufmann*, 5, 2001.

[53] W Eckerson. Smart companies in the 21st century: The secrets of creating successful business intelligence solutions. tdwi the data warehousing institute report series, 1-35, 2003.

[54] Stephen R Gardner. Building the data warehouse. *Communications of the ACM*, 41(9):53, 1998.

[55] Solange Oliveira Rezende. *Sistemas inteligentes: fundamentos e aplicações*. Editora Manole Ltda, 2003.

[56] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth. From data mining to knowledge discovery in databases. *AI magazine*, 17(3):37, 1996.

[57] Robert Groth. *Data mining: building competitive advantage*. prentice Hall PTR, 2000.

[58] Naohiko Uramoto, Hirofumi Matsuzawa, Tohru Nagano, Akiko Murakami, Hironori Takeuchi, and Koichi Takeda. A text-mining system for knowledge discovery from biomedical documents. *IBM Systems Journal*, 43(3):516–533, 2004.

[59] Marti A Hearst. Untangling text data mining. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, pages 3–10. Association for Computational Linguistics, 1999.

[60] Vladimir Brusic and John Zeleznikow. Knowledge discovery and data mining in biological databases. *The Knowledge Engineering Review*, 14(03):257–277, 1999.

[61] Don R Swanson and Neil R Smalheiser. An interactive system for finding complementary literatures: a stimulus to scientific discovery. *Artificial intelligence*, 91(2):183–203, 1997.

[62] Rakesh Agrawal, Ramakrishnan Srikant, et al. Fast algorithms for mining association rules. In *Proc. 20th int. conf. very large data bases, VLDB*, volume 1215, pages 487–499, 1994.

[63] Hsinchun Chen, Sherrilynne S Fuller, Carol Friedman, and William Hersh. *Medical informatics: knowledge management and data mining in biomedicine*, volume 8. Springer Science & Business Media, 2006.

[64] Efraim Turban, J Aronson, and Ting-Peng Liang. *Decision Support Systems and Intelligent Systems 7 "" Edition*. Pearson Prentice Hall, 2005.

[65] Ronen Feldman and James Sanger. *The text mining handbook: advanced approaches in analyzing unstructured data*. Cambridge University Press, 2007.

[66] Hsinchun Chen, Sherrilynne S Fuller, Carol Friedman, and William Hersh. *Medical informatics: knowledge management and data mining in biomedicine*, volume 8. Springer Science & Business Media, 2006.

[67] Denis Cedeno Moreno and Miguel Vargas-Lombardo. Towards a model of knowledge extraction of text mining for palliative care patients in panama.

[68] Efraim Turban, Ramesh Sharda, Jay E Aronson, and David King. *Business intelligence: A managerial approach*. Pearson Prentice Hall Upper Saddle River, NJ, 2008.

[69] Martin Ellingsworth and Dan Sullivan. Text mining improves business intelligence and predictive modeling in insurance. *Information Management*, 13(7):42, 2003.

[70] Lynette Hirschman, Jong C Park, Junichi Tsujii, Limsoon Wong, and Cathy H Wu. Accomplishments and challenges in literature data mining for biology. *Bioinformatics*, 18(12):1553–1561, 2002.

[71] Wendy W Chapman, Will Bridewell, Paul Hanbury, Gregory F Cooper, and Bruce G Buchanan. A simple algorithm for identifying negated findings and diseases in discharge summaries. *Journal of biomedical informatics*, 34(5):301–310, 2001.

[72] S Vijayarani, Ms J Ilamathi, and Ms Nithya. Preprocessing techniques for text mining-an overview. *vol*, 5:7–16.

[73] Roy Thomas Fielding. *Architectural styles and the design of network-based software architectures*. PhD thesis, University of California, Irvine, 2000.

[74] Roy T Fielding and Richard N Taylor. Principled design of the modern web architecture. *ACM Transactions on Internet Technology (TOIT)*, 2(2):115–150, 2002.

[75] Alex Rodriguez. Restful web services: The basics. *IBM developerWorks*, 2008.

[76] Jeffrey D Walker and Steven C Chapra. A client-side web application for interactive environmental simulation modeling. *Environmental Modelling & Software*, 55:49–60, 2014.

[77] Michelle J Irving, Allison Tong, Stephen Jan, Alan Cass, John Rose, Steven Chadban, Richard D Allen, Jonathan C Craig, Germaine Wong, and Kirsten Howard. Factors that influence the decision to be an organ donor: a systematic review of the qualitative literature. *Nephrology Dialysis Transplantation*, 27(6):2526–2533, 2012.

[78] Yorick J de Groot, Eelco FM Wijdicks, Mathieu van der Jagt, Jan Bakker, Hester F Lingsma, Jan NM IJzermans, and Erwin JO Kompanje. Donor conversion rates depend on the assessment tools used in the evaluation of potential organ donors. *Intensive care medicine*, 37(4):665–670, 2011.

[79] Barbara J Daly. End-of-life decision making, organ donation, and critical care nurses. *Critical Care Nurse*, 26(2):78–86, 2006.

[80] Naohiko Uramoto, Hirofumi Matsuzawa, Tohru Nagano, Akiko Murakami, Hironori Takeuchi, and Koichi Takeda. A text-mining system for knowledge discovery from biomedical documents. *IBM Systems Journal*, 43(3):516–533, 2004.

[81] CE Lovelock, GJE Rinkel, and PM Rothwell. Time trends in outcome of subarachnoid hemorrhage population-based study and systematic review. *Neurology*, 74(19):1494–1501, 2010.

[82] A.P. Fernandes. Early imagiology criteria for the screening of possible brain-death donors at the emergency room. Master's thesis, University of Barcelona, 2013.

[83] CD MANNING, P Raghavan, and H Schütze. Introduction to information retrieval: Cambridge: 2008.

[84] R Studio. Rstudio: integrated development environment for r. *RStudio Inc, Boston, Massachusetts*, 2012.

[85] Ingo Feinerer. Introduction to the tm package text mining in r, 2015.

[86] Wendy W Chapman, David Chu, and John N Dowling. Context: An algorithm for identifying contextual features from clinical text. In *Proceedings of the Workshop on BioNLP 2007: Biological, Translational, and Clinical Language Processing*, pages 81–88. Association for Computational Linguistics, 2007.

[87] Introduction to oracle database, accessed on 29-07-2016. Available online at https://docs.oracle.com/cd/E11882_01/server.112/e40540/intro.htm#CNCPT001.

[88] Sql developer documentation release 4.1, accessed on 29-07-2016. Available online at https://docs.oracle.com/cd/E55747_01/index.htm.

[89] Appropriate Uses For SQLite, accessed on 28-07-2016. Available online at https://sqlite.org/whentouse.html.

[90] James Gosling, Bill Joy, GL Steele Jr, Gilad Bracha, and Alex Buckley. The java r language specification–java se 8 edition. oracle america, 2014.

[91] Miguel Grinberg. *Flask Web Development: Developing Web Applications with Python.* " O'Reilly Media, Inc.", 2014.

[92] Angularjs: Developer guide: Introduction, accessed on 29-07-2016. Available online at https://docs.angularjs.org/guide/introduction.

[93] Bootstrap, accessed on 29-07-2016. Available online at http://getbootstrap.com/.

[94] Data-driven documents, accessed on 29-07-2016. Available online at https://d3js.org/.

[95] Nvd3, accessed on 29-07-2016. Available online at http://nvd3.org/.

[96] Janalta Interactive Inc. Proof of concept (poc), accessed on 13-08-2016. Available online at https://www.techopedia.com/definition/4066/proof-of-concept-poc.

[97] IT Data Center Solutions SystemITS. Prova de conceito, accessed on 13-08-2016. Available online at http://www.systemits.com/pt/solucoes/data-security/12-site-portugues/paginas-pt/institucional/67-prova-de-conceito.

Appendix

Appendix A

Patient's detailed view elements

A.1 Patient's general details

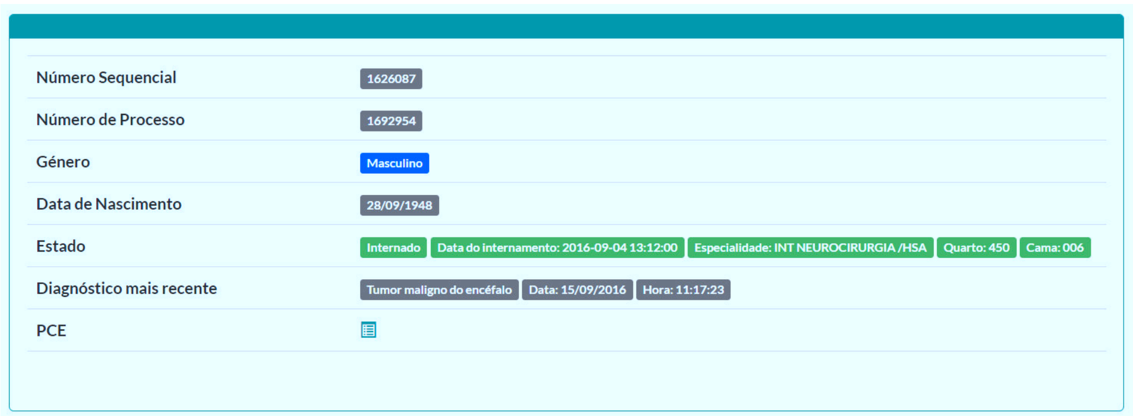


Figure A.1: Patient's general details.

A.2 Patient's clinical details

A.2.1 Diagosis tab

Diagnósticos | Exames realizados | Análises realizadas

ID Episódio	Módulo	Especialidade	Código ICD9	Anotações	Nº. Ord. Profissional	Data do Diagnóstico
16024282	BLO	NEUROCIRURGIA	2252	Tumor benigno do encéfalo e de outras partes do sistema nervoso	49280	12/09/2016
16024661	INT	INT NEUROCIRURGIA /HSA	2252	Tumor benigno do encéfalo e de outras partes do sistema nervoso	49280	12/09/2016

Figure A.2: Patient's diagnosis retrieved by the system.

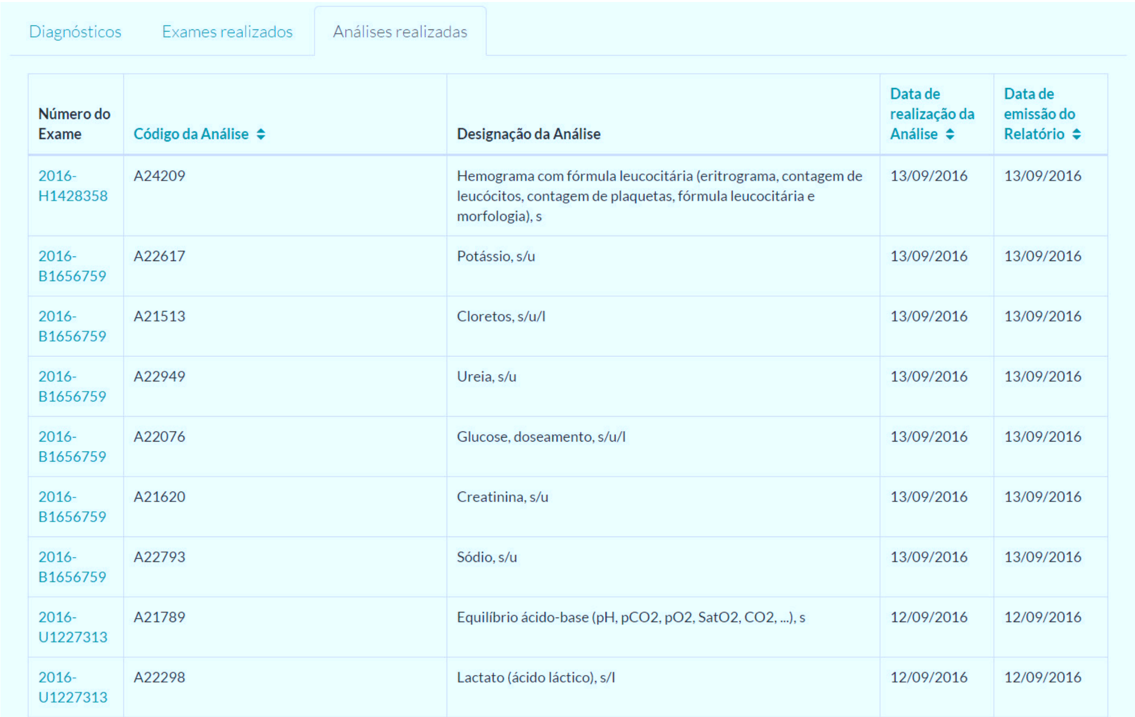
A.2.2 Medical tests tab

Diagnósticos | Exames realizados | Análises realizadas

Número do Exame	Código do Exame	Designação do Exame	Data de realização do Exame	Data de emissão do Relatório
RXC.170.2016.2681	M18243	RM, adicional de estudo por difusão	07/09/2016	07/09/2016
RXC.170.2016.2681	M18240	Pós processamento (exemplo: sequência 3D, VR, análise funcional)	07/09/2016	07/09/2016
RXC.170.2016.2681	M18210	RM, suplemento de contraste	07/09/2016	07/09/2016
RXC.170.2016.2681	M18010	RM do crânio	07/09/2016	07/09/2016

Figure A.3: Patient's medical tests tab.

A.2.3 Clinical tests tab



Número do Exame	Código da Análise	Designação da Análise	Data de realização da Análise	Data de emissão do Relatório
2016-H1428358	A24209	Hemograma com fórmula leucocitária (eritrograma, contagem de leucócitos, contagem de plaquetas, fórmula leucocitária e morfologia), s	13/09/2016	13/09/2016
2016-B1656759	A22617	Potássio, s/u	13/09/2016	13/09/2016
2016-B1656759	A21513	Cloretos, s/u/l	13/09/2016	13/09/2016
2016-B1656759	A22949	Ureia, s/u	13/09/2016	13/09/2016
2016-B1656759	A22076	Glucose, doseamento, s/u/l	13/09/2016	13/09/2016
2016-B1656759	A21620	Creatinina, s/u	13/09/2016	13/09/2016
2016-B1656759	A22793	Sódio, s/u	13/09/2016	13/09/2016
2016-U1227313	A21789	Equilíbrio ácido-base (pH, pCO2, pO2, SatO2, CO2, ...), s	12/09/2016	12/09/2016
2016-U1227313	A22298	Lactato (ácido láctico), s/l	12/09/2016	12/09/2016

Figure A.4: Patient's clinical tests tab.

Appendix B

Dashboards

B.1 Identified cases in the last 24 hours

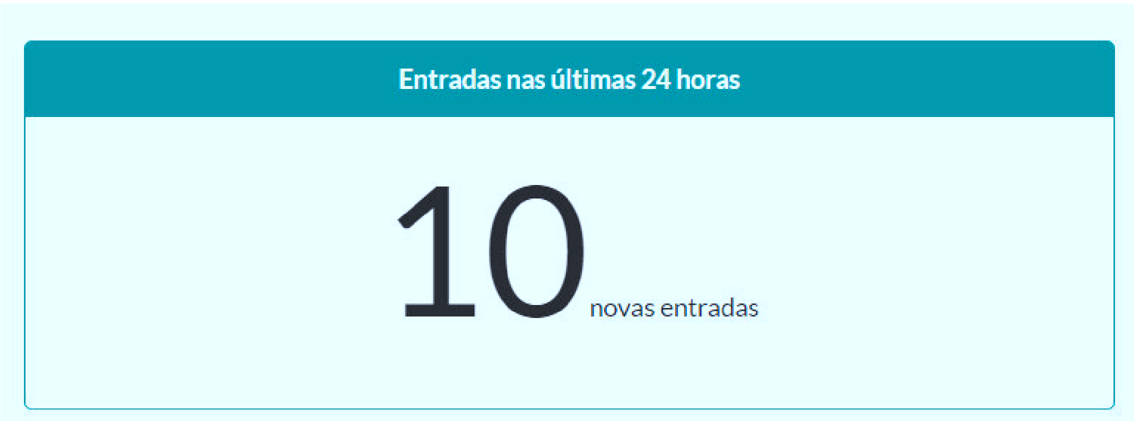


Figure B.1: Counter of all the identified cases in the last 24 hours.

B.2 Total number of entries per clinical module within the last week

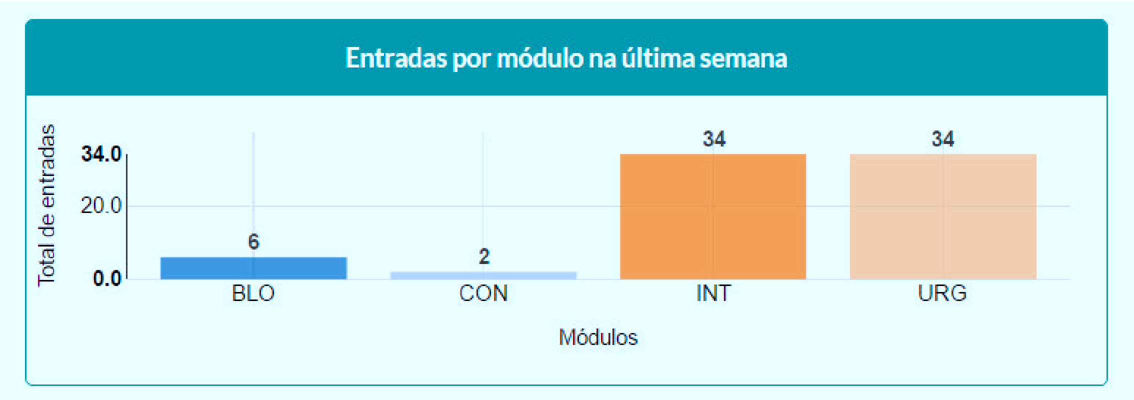


Figure B.2: Bar chart representing the total number of entries per clinical module within the last week.

B.3 Total number of entries per clinical module, specialty and ICD-9-CM code

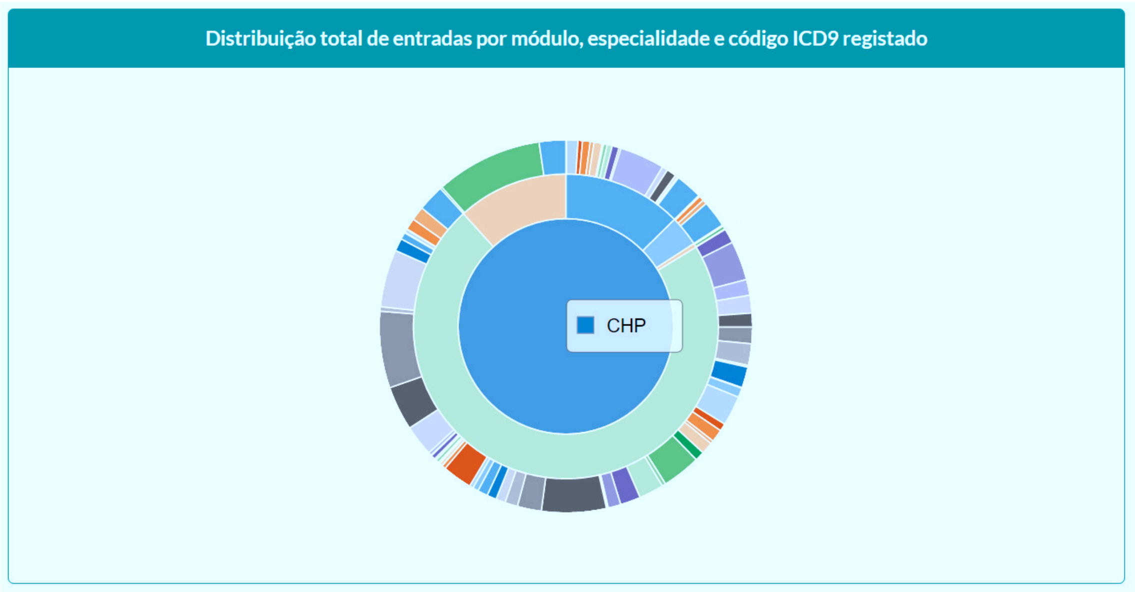


Figure B.3: Interactive radial chart representing the total number of entries per clinical module, specialty and ICD-9-CM code.

B.4 Total relative distribution of entries per medical specialty

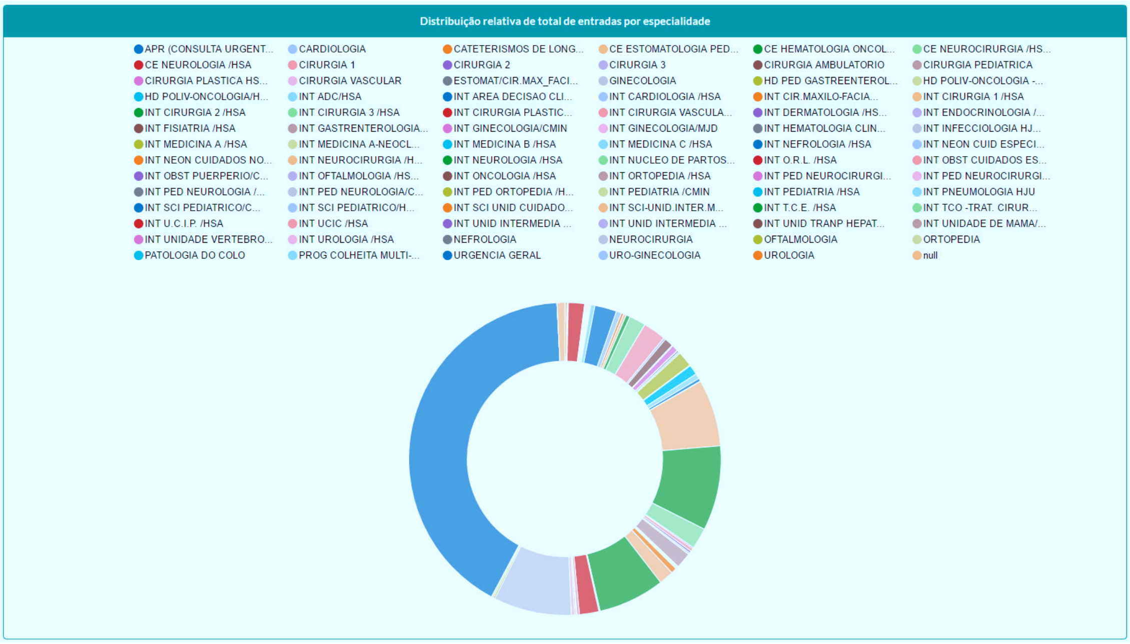
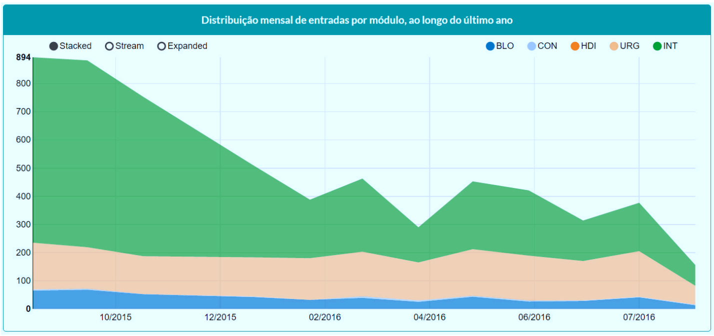
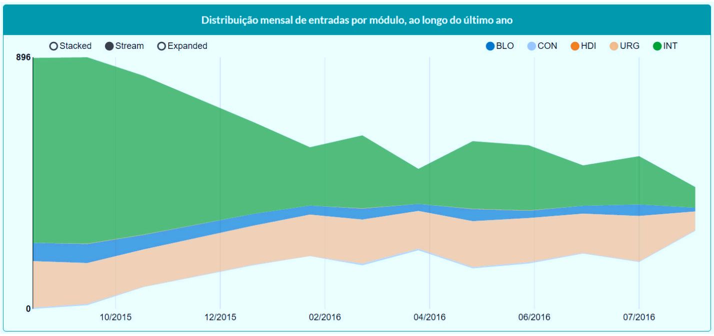


Figure B.4: Donut chart representing the total relative distribution of entries per medical specialty.

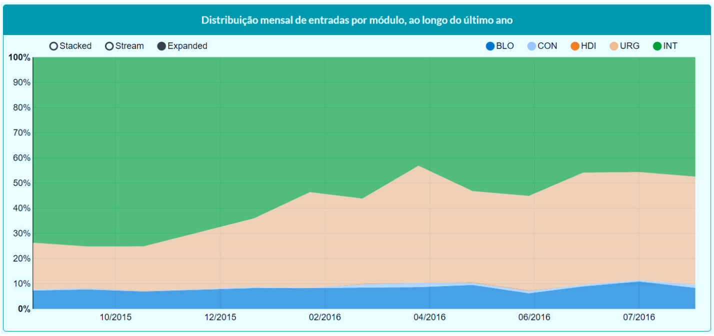
B.5 Total number of entries per clinical module, within the last year, grouped by month



(a) Stacked view.



(b) Stream view.



(c) Expanded view.

Figure B.5: Stacked area chart representing the total number of entries per clinical module, within the last year, grouped by month. (a) Stacked, (b) stream and (c) expanded views.

Appendix C

Published Works

C.1 Steps towards Interoperability in Healthcare Environment

Authors: Hugo Peixoto, Andréa Domingues and Bruno Fernandes

Book: Applying Business Intelligence to Clinical and Healthcare Organizations

Year: 2016

Status: Published

Abstract: Information should be accessible everywhere and at any time to help with clinical decision and be available for clinical studies through data computationally interpretable. This work is based on a set of studies performed at Centro Hospitalar do Tâmega e Sousa. An Electronic Semantic Health Record was formalized and implemented which was delivered through a platform named Agency for the Integration, Diffusion and Archive, which is supported by intelligent agents. Furthermore, to strengthen the relation between the patient and the hospital, an appointment alert system was developed, which allowed the reduction of non-programmed misses and a decrease of costs. Finally to promote user's confidence on Information Systems, an open-source tool was developed that enables the scheduling of preventive actions. These tools allowed continuous improvement of systems and are currently well accepted inside the healthcare unit, proving in real clinical situation the effectiveness and usability of the model.

C.2 Evolutionary Intelligence and Quality-Of-Information: A Specific Case Modelling

Authors: Eliana Pereira, Eva Silva, Bruno Fernandes and José Neves

Book: Applying Business Intelligence to Clinical and Healthcare Organizations

Year: 2016

Status: Published

Abstract: The strategy of making predictions for a specific case or problem, in particular regarding scenarios with incomplete information, should follow a dynamic and formal model. This chapter presents a specific case concerning the employment of professionals for a health institution, as technicians and physicians, to demonstrate a model that requires the Quality-of-Information and the Degree-of-Confidence of the extensions of the predicates that model the universe of discourse. It is also mentioned a virtual intellect, or computational model, in order to maximize the Degree-of-Confidence that is associated with each term in the extensions of the predicates, according to the approximate representation of the universe of discourse. This model is prepared to be adopted by a Business Intelligence platform in order to increase the Quality-of-Information and the Degree-of-Confidence of the extensions in healthcare.

Glossary

brain dead Patient whose brain function is complete and irreversibly lost (including involuntary activity necessary to sustain life) and need machines in order to stay alive. 3, 27, 33, 73

client-server architecture Distributed application structure that partitions tasks or workloads between the providers of a resource or service, called servers, and service requesters, called clients. 48

cranio-encephalic computerized tomography Noninvasive diagnostic imaging procedure that uses a combination of X-Rays and computer technology to produce horizontal, or axial, images (often called slices) of the brain. A CE CT scan may be performed to assess the brain for tumors and other lesions, injuries, intracranial bleeding, structural anomalies such as hydrocephalus, infections, brain function or other conditions. 34, 73

data mining Process of finding correlations or patterns and searching for valuable business information in large relational databases. 13, 74

debug Debugging is the process of locating and fixing or bypassing bugs (errors) in computer program code or the engineering of a hardware device. 48

framework Conceptual structure intended to serve as a support or guide for the building of something that expands the structure into something useful. 46, 48–51

intranet Private network that is contained within an enterprise. It may consist of many interlinked local area networks. Typically, an Intranet uses TCP/IP, HTTP, and other Internet protocols. 65, 73, 74

open source Reference to any program whose source code is made available for use or modification as users or other developers see fit. Open source software is usually developed as a public collaboration and made freely available. 48

text mining The analysis of data contained in natural language text documents. 36, 37