MESG

MESTRADO EM ENGENHARIA DE SERVIÇOS E GESTÃO

Evaluation of a Portuguese computerized cancer registry – a qualitative research

Cátia Andreia Santos Pereira Augusto

Master Thesis

Supervisor at FEUP: Prof. António Brito

Supervisor at HealthySystems: Prof. Ricardo Correia

U.PORTO

FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

2017-09-22

To my son and my husband

for all their unconditional support and constant encouragements

Abstract

A cancer registry is a standardized tool to produce population-based data on cancer incidence and survival. Cancer registries can retrieve and store information on all cancer cases occurring in a defined population. The main sources of data on cancer cases usually include: treatment, diagnostic facilities (oncology centres or hospital departments, pathology laboratories, or imaging facilities etc.) and the official territorial death registry.

The aim of this dissertation thesis is to assess the actual solutions for cancer registries and understanding its needs. This thesis approaches this subject in two approaches: (A) study cancer registry solutions in Europe and (B) study a specific case study, namely a Portuguese cancer registry solution.

To achieve this goal, two studies were made through a qualitative research. The first one (study A) involved a literature review following PRISMA statement and the second one (study B) consisted in an evaluation of the north regional cancer registry (RORENO) of Portugal with the intention to characterize: the main functionalities and its core processes, the team involved, different healthcare institutions in the regional network and an identification of issues and potential improvements.

The results of this research showed that both European and Portuguese reality on this topic share the same concerns and gaps. Cancer registries systems had in general problems due to thr lack of an automatic integration of data from the different sources, difficulty in automatize data quality routines and a lack of harmonization in terms of standards. Most of the tasks are performed manually implying an extra effort from the human resources team that results in a delay in survival and incidence reports production.

In a near future it is crucial to automatize the integration of data linking the different healthcare institutions in the region. However, it is important to think which functionalities this system should give to the institutions in the network to maximize the engagement with these systems. More than a database, these systems should be a source of knowledge available to all the collaborative oncologic network.

Keywords *– cancer registries; qualitative research; BPM, PRISMA, semi-structure interview, observation, KPI.*

Resumo

Um registo de cancro é uma ferramenta padronizada para produzir dados baseados na população sobre incidência e sobrevivência de cancro. Os registos de cancro podem cobrir e armazenar informações sobre todos os casos de cancro de uma população definida. As principais fontes de dados para a descrição de casos de cancro geralmente incluem: dados de tratamento, diagnóstico (centros de oncologia ou departamentos hospitalares, laboratórios de patologia ou instalações de imagem etc.) e o registo oficial de óbito.

O objetivo desta dissertação é avaliar as soluções de registos de cancro implementadas e entender as suas necessidades. Esta dissertação interpela este assunto sob duas abordagens: (A) estudar soluções de registo de cancro na Europa e (B) estudar um estudo de caso específico, nomeadamente uma solução de registro de cancro português.

Para atingir esse objetivo, dois estudos foram realizados através de uma investigação qualitativa. O primeiro (estudo A) envolveu uma revisão da literatura seguindo a *framework* PRISMA e o segundo (estudo B) consistiu numa avaliação do registo regional de cancro do Norte (RORENO) de Portugal com a intenção de caracterizar: as principais funcionalidades e principais processos, a equipa envolvida, diferentes instituições de saúde na rede regional oncológica e uma identificação de problemas e potenciais melhorias.

Os resultados desta investigação mostraram que a realidade europeia e portuguesa sobre esse tema compartilham as mesmas preocupações e lacunas. Os sistemas de registo de cancro revelam problemas devido à falta de uma integração automática de dados das diferentes fontes, dificuldade em automatizar rotinas de qualidade de dados e falta de harmonização em termos de padrões (standards). A maioria das tarefas é realizada manualmente, implicando um esforço extra da equipa de recursos humanos que resulta num atraso na produção de relatórios de sobrevivência e incidência.

Num futuro próximo, é crucial automatizar a integração de dados que ligam as diferentes instituições de saúde da região. No entanto, é importante pensar quais funcionalidades este sistema deve dar às instituições na rede para maximizar o engajamento com esses sistemas. Mais do que uma base de dados, estes sistemas devem ser uma fonte de conhecimento disponível para toda a rede oncológica colaborativa.

Acknowledgments

To my friend and supervisor (for the second time), Professor Ricardo Correia, whose availability and help whenever I requested, expertise, understanding, and friendship were essential in accomplishing this work.

A special gratitude to HealthySystems for the opportunity to develop this work and for all the support.

I would like to thank the RORENO group that actively collaborate in the research process and IPO-Porto for the opportunity given in publish this work.

I would like to my supervisor at FEUP, Professor António Brito, I would like to acknowledge his support and guidance given during the thesis development

To all my master degree colleagues who accompanied me in this wonderful and challenging journey.

Thank to my friends and family who helped and encouraged me to proceed.

Last, but definitely not least, to my husband Alexandre for his unconditional support and encouragement, without him I would never enroll in this master degree and to my son Leonardo for giving all the motivation I needed.

Table of Contents

List of Tables

List of Figures

List of abbreviations

AD	Active directory
ADT	Admission discharge and transfer system
BI	Business Intelligence
BPM	Business Process Management
CNPD	*Comissão Nacional de Proteção de Dados*
CR	Cancer registry
DICOM	Digital Imaging and Communications in Medicine
ENCR	European Network of Cancer Registries
EPR	Electronic Patient Record
EU	European Union
GDPR	General Data Protection Regulation
GTDS	*Das Gießener Tumordokumentationssystem*
HL7	Health Level 7
HIS	Hospital Information System
IPO-Porto	*Instituto Português de Oncologia do Porto Francisco Gentil, EPE*
IACR	International Association of Cancer Registries
IARC	International Agency for Research on Cancer
ICD-O	International Classification of Diseases for Oncology 3rd Revision
ICD-9	International Classification of Diseases 9th Revision
IT	Information Technology
KPI	Key Performance Indicator
LDAP	Lightweight Directory Access Protocol
LIS	Laboratory Information System
MSc	Master Science
ODISSEIA	Oncological Diseases Information System
PIA	Privacy Impact Assessment
PDS	*Plataforma de Dados da Saúde*
PRISMA	Preferred Reporting Items for Systematic reviews and Meta-Analyses
RIS	Radiology Information System
RNU	*Registo Nacional de Utentes*
RO	*Registo Oncológico*
RON	*Registo Oncológico Nacional*

ROR	*Registo Oncológico Regional*
RORENO	*Registo Oncológico Regional do Norte*
ROR-Centro	*Registo Oncológico Regional do Centro*
ROR-Sul	*Registo Oncológico Regional do Sul*
SICO	*Sistema de Informação dos Certificados de Óbito*
UML	Unified Modeling Language
WHO	World Health Organization
XML	Extensible Markup Language

Contributions

This section describes the outcomes obtained during this research. The main contributions of this dissertation were:

- A systematic review on computerized cancer registries in Europe;
- An evaluation of a Portuguese cancer registries system that results in a publication in the 12th Iberian Conference on Information Systems and Technologies (CISTI) available in IEEE digital library. (Santos-Pereira, 2017)*.
- Insights for a new cancer registry model that outcomes from the results of the two previous studies enunciated.
- A set of specific Key Performance Indicators to evaluate the transition between the actual model and the model proposed.

*Article published

Santos-Pereira, C. (2017) 'A qualitative research evaluation of a Portuguese computerized cancer registry', in *2017 12th Iberian Conference on Information Systems and Technologies (CISTI)*, pp. 1–6. doi: 10.23919/CISTI.2017.7975827. Full paper is available in APPENDIX E.

1 Introduction and motivation

This section introduces and contextualize the dissertation subject, which intends to better understand the cancer registry including its goals, limits and legal aspects. This project was developed at HealthySystems.

Medical registries are described as a systematic collection of a clearly defined set of health and demographic data for patients with specific health characteristics, held in a central database for a predefined purpose (Bianconi *et al.*, 2012).

A cancer registry is a standardized tool to produce population- based data on cancer incidence and survival (O.M. Jensen, D.M. Parkin, R. MacLennan, 1991a). Cancer registries can retrieve and store information on all cancer cases occurring in a defined population. Its data can be used in a wide variety cancer control ranging from etiological research, through primary and secondary prevention to health-care planning and patient care, so benefiting both the individual and society. Cancer registries are evolving to provide a high level of clinical details, and to improve their capability to provide an evaluation of health interventions in oncology. Diagnosis, stage, and treatment information is registered with increasing frequency and higher level of clinical detail. Thus, the cancer registry is evolving as a tool to support planning and evaluation of cancer control strategies. However, the traditional cancer registry is retrospective or historical in its nature since it is presently limited/bound to investigate variables routinely determined in health archives (Bianconi *et al.*, 2012).

The storage of large quantity of cancer registries is possible. However, the real problem in expanding the cancer registry scope is the difficulty to access an increasing number of variables from different sources (Lenzerini, 2002).

Because of the new evaluation goals and changes in cancer care, additional data sources are required for cancer registries to maintain completeness and validity of information. For instance, linkage with screening archives is useful to identify screen-detected cancers and improve the ability to evaluate screening activities. Similarly, ambulatory care, diagnostic tests, and drug prescription files are increasingly necessary due to the wide use of out-patient care and to calculate quality of care indicators. This important effort that is the evolution of the cancer registry into an intelligence unit for surveillance and evaluation of oncological care must be accomplished without losing timeliness of data diffusion (Bianconi *et al.*, 2012).

Even though the final achievements in terms of data validity and timeliness depend on the level of automation of the various sources needed for cancer registration, the IT (Information Technology) used by a cancer registry is also essential to make the best of the available data and to allow the rapid production of results (O.M. Jensen, D.M. Parkin, R. MacLennan, 1991b).

Regarding data protection, last year the European Commission proposed to replace the Directive (95/46/EC) (Conseil, 1995) by the General Data Protection Regulation (Commission, 2016). The overall intention of this reform is to protect personal data and to facilitate a free flow of data within the European Union (EU); also, aiming to overcome problems alluded to by the research community concerning data sharing across borders for research purposes. The outcome of the data protection reform is crucial to all epidemiological activities and clinical quality control in the EU In contrast to a directive, a regulation is binding by itself and does not need implementing legislation by the Member States. It implies a harmonization of data protection measures across the EU.

On the one hand harmonization may facilitate valuable data sharing for research purposes, but on the other, excessive regulation can easily disable even simple monitoring of cancer, with disastrous consequences for public health information (Hakulinen *et al.*, 2011).

1.1 Problem Description

In Europe, cancer registration is challenged by significant disparities in the quality and coverage of cancer registries (CRs), by insufficient harmonization and comparability of procedures and data, by heterogeneous legislation that limits CR's abilities for networking, collaboration, and participation in research (A. Forsea, 2016).

Furthermore, in Portugal the Health Ministry is working on an unique cancer registry for Portugal – RON - *Registo Oncológico Nacional* (SNS, 2017). This new system will integrate the data from the actual regional cancer registries RORENO (IPO-Porto, 2017b), ROR-Centro (IPO-Coimbra, 2017) and ROR-Sul (IPO-Lisboa, 2017)), following the recommendations of the WHO (World Health Organization) (WHO/IARC, 2014). Many discussions arise around this topic specially in terms of data protection issues.

Therefore, this is the right time to make a state of the art and analysis of the actual solutions implemented in the market in the Europe, identify the main problems and concerns that arise in the last years and use this knowledge to build new systems that fulfil user needs in balance with the legislation rules.

1.2 Research Questions

Driven by this moment of change, this MSc dissertation focus on an evaluation of the actual solutions for cancer registries and understanding its needs.

The main research questions that this study proposes to answer are:

a) Which are the main characteristics of the implemented cancer registries?
    a. These characteristics/ functionalities fulfil the cancer registries requirements and needs?
    b. Which are the main cancer registries problems, concerns and solutions?
b) Which are the main functionalities, users and stakeholders of RORENO?
    a. Which are the main problems/ constraints felt by the users?
c) Which are the mandatory cancer registry characteristics/ functionalities that fulfil the users' needs?

1.3 Objectives

The objectives were defined by following the research questions, resulting in two main objectives:

1) To analyse the main characteristics and concerns about CRs available in the literature.
2) To analyse a specific computerized cancer registry implemented *in "Instituto Português de Oncologia do Porto"* - RORENO. For example, to characterize the system environment, such as its main functionalities and core processes, team involved and different healthcare institutions in the regional network. Furthermore, to identify its main problems and difficulties that arose RORENO in the last years.

1.4 Study and Project Development at Healthy Systems

HealthySystems is a spin-off company of the University of Porto. Their products and services include optimization of cyber-security systems, the development of security technologies using mobile devices, checking the quality of data in databases and messaging in real time optimization of infrastructure networks, security audits and consulting on performance and data protection laws. HealthySystems works in collaboration with *"Instituto Português de Oncologia do Porto – IPO-Porto"* in ODISSEIA project.

The scope and objectives of ODISSEIA project are to support the collaboration and sharing of knowledge among all actors involved in prevention, diagnosis, treatment and research of cancer disease in the North of Portugal, as well as the collection, processing and availability of information for the multiple stakeholders that form part of the entire network that has developed around oncological disease.

ODISSEIA is associated with an innovative application and paradigm change, so it is important to reinforce that ODISSEIA is not traced in the unknown. On the contrary, ODISSEIA is:

- Based on solid experience and knowledge consolidated over many years by the IPO-Porto professionals (medical doctors and RORENO users).
- It has the presupposition of involving the various actors, namely the institutions and health professionals, identifying their needs and creating value for each one of parts.

1.5 Report outline

This MSc thesis is organized into eight chapters. The present chapter "*Introduction and Motivation*" introduces the theme of this thesis, describing the research questions and details the objectives of the execution of this work.

The second chapter "*Background*" presents some key-concepts important to retain before starting the studies, for example what is a cancer registry, an overview about data protection legislation and its implication in this area.

Then, it is presented the "*Methodology*" in chapter three. The fourth chapter "*Study A – Literature Review*" presents the results of a systematic review on "computerized cancer registries" and the main findings.

Chapter five presents "*Study B – RORENO qualitative research evaluation*", in this chapter it is presented the main results of RORENO evaluation as well as the concerns of the system users. This study was published in "*CISTI'2017 - 12ª Conferência Ibérica de Sistemas e Tecnologias de Informação*" held in Lisbon – Portugal during June 2017.

Chapter six "*RORENO – going forward*" aggregate the main findings of Study A and B and propose a direction for the future of this system and a set of key performance indicators to evaluate the performance of a new system compared to the actual system.

Finally, in chapter seven "*Conclusion and Recommendation*s" is presented the main findings, limitations of the work and final conclusions.

2 Background

This section describes the state of the art, introducing some important concepts and current developments in Europe. A comprehensive reflection about the theme is later described in chapter five corresponding to Study A: Systematic Review.

2.1 Cancer registration

A cancer registry is an essential part of a programme of cancer control. By definition, a cancer registry is an information system designed for the collection, storage, management, and analysis of data on persons with cancer, usually covering a hospital or group of hospitals (O.M. Jensen, D.M. Parkin, R. MacLennan, 1991a).

The main purposes are:

i. To establish and maintain a cancer incidence reporting system;
ii. To be an informational resource for the investigation of cancer and its causes; and
iii. To provide information to assist public health officials and agencies in the planning and evaluation of cancer prevention and cancer control programs.

Cancer registry information may be used in a multitude of areas, and the value of the data increases if comparability over time is maintained.

This information is important to many people for a wide variety of reasons. Researchers need accurate, up-to-date cancer data to study possible causes of cancer. Medical administrators use cancer data to make decisions regarding equipment purchases and developing programs for cancer prevention. Departments of Health use cancer data to investigate potential cancer clusters and their causes.

The methodology for cancer registration comprise the following requirements (Silva, 1999):

a) Clear definition of the catchment population. The registry should be able to distinguish between residents of the area and those who have come from outside and it should be able to register cases in residents treated outside the area.
b) Availability of reliable population denominators from the census or other statistical offices.
c) Generally available medical care and ready access to medical facilities, so that the great majority of cancer cases will come into contact with the health care system at some point in their illness and, therefore, will be correctly diagnosed.
d) Easy access to case-finding sources such as hospitals, pathology departments, death certificates and other sources of clinical data within the catchment area and in the surrounding areas.

2.2 Cancer registry content

The main source of data on cancer cases usually include: treatment, diagnostic facilities (oncology centres or hospital departments, pathology laboratories, or imaging facilities etc.) and the official territorial death registry. Additional data sources like ambulatory, private clinics, elderly care homes and general practitioners' networks increase the completeness of data, but also the logistics and expenses (Silva, 1999).

Hospital-based cancer registries are more numerous and widespread than population-based cancer registries. The primary purpose of these registries is to contribute to patient care by providing accessible information on the patients with cancer, the treatment they received and its results. The data may also be used for clinical research and, to a certain extent, for epidemiological purposes. One of the main advantages of hospital registries are the availability and the completeness of medical records. The data collected by a hospital registry tend to be more extensive than those collected by a population registry (Silva, 1999).

The top benefits of a computerized cancer registry include: 1) more complete treatment information, 2) less time for case finding and data entry, 3) more available time for data retrieval and analysis, 4) improved completeness, accuracy, and timeliness, 5) better patient tracking for follow-up, and 6) improved workflow efficiency (Contiero *et al.*, 2008).

The main challenges of the utilization of a computerized system within a cancer registry are 1) lack of adequate funding, 2) lack of medical staff to support the system, 3) changing data standards, 4) lack of full-time commitments, and 5) lack of a standardized data exchange (Colquitt, Clements and Hart-hester, 2012).

The minimum set of data recommended to be collected by all cancer registries was formulated by the International Agency for Research on Cancer (IARC) (WHO/IARC, 2014) and by the European Network of Cancer Registries (ENCR) (ENCR, 2016) and are described below:

- Cancer patient's personal identification,
- Tumour site,
- Tumour histology, classified by the icd-o,
- Tumour stage,
- Tumour diagnosis, related to the icd9,
- Tumour therapy,
- Further treatment,
- Follow up,
- Individual history,
- Family history,
- Death, including autopsy results, if any.

These items correspond to sensitive personal and medical information (Coebergh, 2015).

Table 1 - ENCR recommendation for optional set of data for CRs (ENCR, 2016).

Variables	Description
The Person	
Personal identification	In some countries a unique ID number, in others full name combined with date of birth and sex
Date of birth	Given as day, month and year (dd/mm/yyyy)
Sex	Male (M) or Female (F)
Ethnic group	As the population mixture increases this variable will increase in importance also to study inequality. [May be difficult to agree a classification which can be applied across the whole of Europe]
Address including postal (or zip) code	Needed for ID purpose and for geographical based studies
Vital status & date	It may be of value to indicate whether known or assumed (e.g. based on linkages to death certificates) (dd/mm/yyyy)
Date of death	Needed to study survival and follow-up (dd/mm/yyyy)
Last follow-up date	Needed to study follow-up (dd/mm/yyyy). Registry should indicate whether date refers to active or passive follow-up.
The Tumour	
Incidence date	This date should be given priority as outlined by the ENCR recommendations as indicated here A–D. (Optional: To have comparability more dates should be collected, preferably all included in the definition)
A: Date of first histological/ cytological confirmation of the tumour	Date of biopsy or date of pathology or date of pathology report (dd/mm/yyyy)
B: Date of first hospital admission or contact	May be the date of first out-patient visit for the disease (dd/mm/yyyy)
C: Other date of diagnosis	e.g. GP visit (dd/mm/yyyy)
D: Date of death	For cases discovered at death/autopsy or unknown (dd/mm/yyyy)
Primary tumour site	This should as a minimum be according to the ICD-O (*International Classification of Diseases for Oncology*)
Laterality	This should be recorded for all paired organs, but as a minimum for breast, eye, ovary, testis, and kidney (but observe the multiple primary rules)
Primary tumour histology	This should as a minimum be according to the ICD-O

2.3 Classification and coding

Classification of neoplasms (tumour) involves their arrangement or distribution in classes according to a method or system. Neoplasms can be classified in many ways but, for cancer registry and clinician alike, the two most important items of information are the anatomical location of the tumour in the body and the morphology; i.e., the appearance of the tumour when examined under the microscope (histology and cytology), as this indicates its behaviour (malignant, benign, in situ, and uncertain) (O.M. Jensen, D.M. Parkin, R. MacLennan, 1991a).

Since it was first published in 1976, the International Classification of Diseases for Oncology (ICD-O) has been internationally recognized as the definitive classification of neoplasms. It is used by cancer registries throughout the world to record incidence of malignancy and survival rates, and the data produced are used to inform cancer control, research activity, treatment planning and health economics (WHO, 2017).

The classification of neoplasms used in ICD-O links closely to the definitions of neoplasms used in the WHO/IARC Classification of Tumors series which are compiled by consensus groups of international experts and, as such, the classification is underpinned by the highest level of scientific evidence and opinion.

The third edition of ICD-O (ICD-O-3) has been available in printed format since 2000. In September 2011, following approval by the WHO/IARC Committee for ICD-O-3, the classification was updated with several new or modified codes and terms (ICD-O-3 First Revision, or ICD-O-3.1) (WHO, 2013).

The ICD-O coding is used in cancer registries for coding the site (topography) and the histology (morphology) of neoplasms, usually obtained from a pathology report. It gives a multi-axial classification of the site, morphology, behaviour, and grading of neoplasms.

2.4 Data quality

Two main issues should be considered when evaluating the quality of the data in a cancer registry: its completeness and its validity (Silva, 1999).

Completeness is the extent to which all the incident cancers occurring in the population are included in the registry database. Incidence rates and survival proportions will be close to their true value if maximum completeness in case-finding procedures can be achieved. A population based-registry should, by definition, register every single case that occurs in its catchment population. However, case ascertainment is rarely complete. Various methods, such as comparisons with death certificates and hospital records, have been used to determine the degree of completeness of registration (Zanetti et al., 2015).

Validity or accuracy refers to the proportion of cases in the registry with a given characteristic that truly have that attribute, and depends on the precision of source documents and the level of expertise in abstracting, coding and recoding. The validity of the data can be assessed in various ways. The proportion of cases with microscopic verification of diagnosis is a very useful index, as is the proportion registered during life (not simply from a death certificate) (Silva, 1999).

Other concepts can be assessed to evaluate the data quality such as comparability and timeless.

Comparability of the statistics generated for different population groups (registries, geographical areas, etc.), and over time, is essential to their meaningful interpretation. A basic requirement is the standardization of practices concerning classification and coding of new cases, and consistency in basic definitions of incidence, such as rules for the recording and reporting of multiple primary cancers occurring in the same individual (Y Bhurgri, A Bhurgri, 2002).

Timeliness of reporting of cancer registry results is an aspect of registry quality that can be considered as a separate issue, although this clearly influences the extent to which data are complete and accurate. Access to recent data is perceived as a priority by users, but, since registries are constantly updating their database as reports are received, and some notifications arrive long after the case was diagnosed, statistics for the recent periods will be incomplete, and will need future updates. There is, therefore, some conflict between the requirement for timely data, and other aspects of data quality, particularly completeness (Zanetti et al., 2015).

Cancer registries should develop their own internal quality control checks so that attention is drawn to missing information and inconsistent data. Many registries frequently re-abstract and re-code a sample of cases to assess the quality of their data (Silva, 1999).

2.5 European initiatives and data

Cancer registration dates back to the first half of the 20th century and have been expanding in Europe since the early 1900's. Over the last three decades cancer registration has become an important element of the EU's strategy against cancer, promoted within the framework of the European Action against Cancer Programme (1985–2008), the European Partnership for Action Against Cancer (EPAAC) (2009–2014). The last data shows that nearly 200 population-based cancer registries (PCRs) are active in Europe and they are members of European Network of Cancer Registries (ENCR) (Coebergh, 2012). The quality registration coverage of population by national cancer registries are available in 22 European countries. High quality registration of 10–50% of the population are available in France, Italy, Switzerland, Spain, Germany, and Serbia (EPAAC, 2014). High quality registration of <10% of the population are available in Poland and Portugal (EPAAC, 2014). EU member states Romania, Greece, and Hungary had as per 2012 only regional or partial data although legislation is in place to allow national cancer registration (EPAAC, 2014).

2.6 European data protection legislation

Cancer case reporting is mandatory by law in most of the European countries with high quality registers that included in the cancer incidence in five continents series (Bray *et al.*, 2014). Personal data protection legislation has a major impact on electronic cancer registry. In a pan-European survey within the EUROCOURSE project, 20–35% of responding cancer registries reported legal-related barriers to cancer registration across most of Europe, while in the South-West region these barriers amounted to up to 60% (Siesling *et al.*, 2015). Particularly concerning are the nationally variable barriers in the linkage of cancer registries with other health-related databases like mass screening programmes, biobanks, vital status, and causes of death databases (Siesling *et al.*, 2015) (Andersen and Storm, 2015).

Protection of personal data has a long history originating from the Nuremberg code in 1947, followed by the first Helsinki declarations by the World Medical Association since 1964, the Belmont report in the United States (US) in 1979 and the Council of Europe convention 108 in 1981. In 1995, the EU adopted the European data protection directive (95/46/EC) (Conseil,

1995) on protection of individuals with regard to the processing of their personal data and the free movement of such data. In general, EU directives lay down end results that must be achieved in every Member State. National authorities have to adapt their laws to meet these goals, but are free to decide how to do so. While Directive (95/46/EC)) set out European data protection objectives and standards for the collection, storage and use of personal data, Member State implementation has led to rather heterogenic regulation regarding the use of data for public health research across the EU, moreover hindering data sharing for research purposes.

In 2014 the European Commission proposed to replace the Directive 95/46/CE (Conseil, 1995) by the General Data Protection Regulation (Commission, 2016). The overall intention of this reform is to protect personal data and to facilitate a free flow of data within the European Union (EU). This initiative will also help to overcome problems alluded by the research community concerning about data sharing across borders for research purposes.

The outcome of the data protection reform is crucial to all epidemiological activities and clinical quality control in the EU. In contrast to a directive, a regulation is binding by itself and does not need implementing legislation by the Member States. It implies a harmonization of data protection measures across the EU, including use of data for public health purposes such as prevention and evaluation of screening programmes. However, on the one hand harmonisation, may facilitate valuable data sharing for research purposes, but on the other hand, excessive regulation can easily disable even simple monitoring of cancer, with disastrous consequences for public health information (Commission, 2016).

2.7 Technology support for cancer registries

Historically, the process for data entry into cancer registries was completed by a certified cancer register using a manual or electronic process for documentation into paper or electronic forms. The path for data abstracting within a cancer registry begins with the initial patient evaluation and diagnosis followed by entry into a hospital registry database. A hospital-based cancer registry collects information on all cancer patients who receive services from a healthcare organization (Houser *et al.*, 2012).

Because of the new evaluation goals and changes in cancer care, additional data sources are required for cancer registries to maintain completeness and validity of information. For instance, linkage with screening archives is useful to identify screen-detected cancers and improve the ability to evaluate screening activities. Similarly, ambulatory care, diagnostic tests, and drug prescription files are increasingly necessary due to the wide use of out-patient care and to calculate quality of care indicators (Gliklich RE, Dreyer NA, Leavy MB, 2014).

The main difficult with expanding the cancer registry scope is the difficulty to access an increasing number of variables from a big number of sources (Tognazzo, 2006). This difficulty may heavily influence the data quality and/or time of data production. The usefulness of evaluation studies depends essentially on timeliness since health technologies and cancer care are rapidly changing (Bray F, 2009).

Even though the final achievements in terms of data validity and timeliness depend on the level of automation of the various sources needed for cancer registration, the information technology used by a cancer registry is also essential to make the best of the available data and to allow the timely production of results.

Computerized systems have been widely adopted in the realization and management of cancer

registries. The International Association of Cancer Registries (IACR) has an open-source tool to input, store, check, and analyze cancer registry data (CanReg5) (IACR, 2017a). However, computerized records implementation may create benefits as well as challenges for cancer registries in areas such as policies and regulations, data quality, reporting, management, staffing, and training.

3 Methodology

This chapter presents the methodology adopted for this research organized in three parts. The first one introduces some research methodologies and the motivation to choose a qualitative approach. Second presents some concepts about the methodology adopted and finally the third presents the research design.

3.1 Research methodologies analysis and motivation to choose a qualitative approach

Scholars face many approaches from which to choose when they conduct a research study. To Neuman (Neuman, 2014) there are two categories for data collection: quantitative, which provides data in the form of numbers and qualitative which provide data in form of words and pictures.

Both approaches use multiple research techniques (e.g., survey, interview, ethnography) to gather and analyse empirical data. Despite some real differences between quantitative and qualitative research, they overlap a great deal.

Quantitative research it aims to achieve absolute an undeniable truth and it is measured in absolute, quantifiable terms. As opposed to quantitative methods, qualitative research methods are flexible, context-specific and situational, furthermore rather than avoiding involvement of the researcher, they recommend it.

Within this research, the focus was towards the use of qualitative methods, due to its objectives, and the increased recognition of the benefits of a qualitative approach to health care research (Luborsky and Rubinstein, 1995).

Table 2 – Comparison between quantitative and qualitative approaches (Neuman, 2014)

Quantitative approach	Qualitative approach
Measure objective facts	Construct social reality, cultural meaning
Focus on variables	Focus on interactive processes, events
Reliability the key factor	Authenticity the key factor
Value free	Values present and explicit
Separate theory and data	Theory and data fused
Independent of context	Situationally constrained
Many cases, subjects	Few cases, subjects
Statistical analysis	Thematic analysis
Researcher detached	Researcher involved

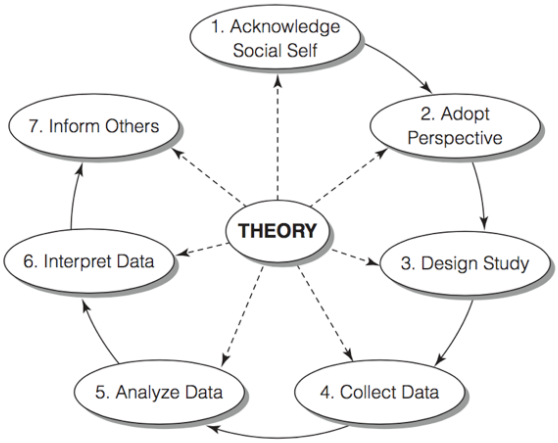


Figure 1 - Steps to perform a qualitative research (Neuman, 2014)

3.2 Qualitative methods

This sub-section describes key-concepts about the methods chosen in this research.

Literature review

An early and essential step in doing a study is to review the accumulated knowledge on your research question. This applies to all research questions and all types of studies. Clichés reinforce this advice: Do not waste time "reinventing the wheel" and remember to "do your homework" before beginning an endeavour. Doing a literature review builds on the idea that knowledge accumulates and that we can learn from and build on what others have done (Neuman, 2014). The main goals of a literature review are:

- *To demonstrate a familiarity with a body of knowledge and establish credibility*. A review tells a reader that the researcher knows the research in an area and knows the major issues. A good review increases a reader's confidence in the researcher's professional competence, ability, and background.
- *To show the path of prior research and how a current project is linked to it*. A review outlines the direction of research on a question and shows the development of knowledge. A good review places a research project in a context and demonstrates its relevance by making connections to a body of knowledge.
- *To integrate and summarize what is known in an area*. A review pulls together and synthesizes different results. A good review points out areas in which prior studies agree, disagree, and major questions remain.
- *To learn from others and stimulate new ideas*. A review tells what others have found so that a researcher can benefit from the efforts of others. A good review identifies blind alleys and suggests hypotheses for replication.

The conduct of a systematic review depends heavily on the scope and quality of included studies: thus, systematic reviewers may need to modify their original review protocol during its conduct. Any systematic review reporting guideline should recommend that such changes can be reported and explained without suggesting that they are inappropriate. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statement acknowledges this iterative process. Aside from Cochrane reviews, all of which should have a

protocol, only about 10% of systematic reviewers' report working from a protocol (Moher D, Tetzlaff J, Tricco AC, Sampson M, 2007). Without a protocol that is publicly accessible, it is difficult to judge between appropriate and inappropriate modifications (Moher D, Liberati A, Tetzlaff J, 2009). The aim of the PRISMA Statement is to help authors improve the reporting of systematic reviews and meta-analyses.

Interviews

This is the most common format of data collection in qualitative research (Neuman, 2014).

Semi-structured interviews are those in-depth interviews where the respondents have to answer open-ended questions and thus are widely employed by different healthcare professionals in their research.

Semi-structured interviews are based on semi-structured interview guide, which is a schematic presentation of questions or topics and need to be explored by the interviewer.

To have the interview data captured more effectively, recording of the interviews is considered an appropriate choice but sometimes a matter of controversy among the researcher and the respondent. Hand written notes during the interview are relatively unreliable, and the researcher might miss some key points. The recording of the interview makes it easier for the researcher to focus on the interview content and the verbal prompts and thus enables the transcriptionist to generate "verbatim transcript" of the interview (A. Oakley, 1998).

Observation

Observation is a type of qualitative research method which not only included participant's observation, but also covered ethnography and research work in the field. In the observational research design, multiple study sites are involved. Observational data can be integrated as auxiliary or confirmatory research (Gray, 2009).

To perform an observation the data collected, tend to be either seen and written down, or recorded on a computer. Observations may be made and recorded immediately as observations, or the data may be recorded 'raw' and analysed later (Kawulich, 2005).

A participant observation has some limitations namely (McLeod, 2015):

a) It can be difficult to get time / privacy for recording. For example, with covert observations researchers can't take notes openly as this would blow their cover. This means they have to wait until they are alone and reply on their memory. This is a problem as they may forget details and are unlikely to remember direct quotations.
b) If the researcher becomes too involved they may lose objectivity and become bias. There is always the danger that we will "see" what we expect (or want) to see. This is a problem as they could selectively report information instead of noting everything they observe. Thus, reducing the validity of their data.

3.3 Research Design

This sub-section presents the study design adopted in this research organized into two studies: (1) systematic review and (2) RORENO evaluation. By combining both studies it will answer the research questions and objectives proposed.

In order to explain this synergy, it was defined four main steps: problem understanding; identification of key concerns; understanding the problem solving and finally discussion

about a new model. Figure 2 shows a diagram that systematize the steps and relation between the studies.

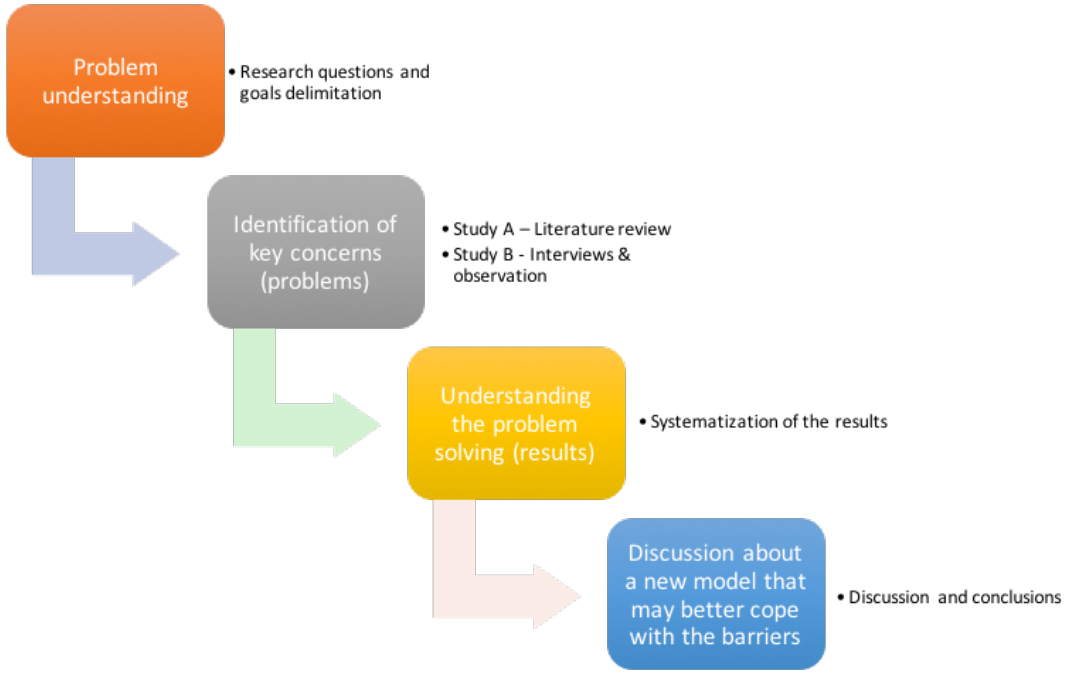


Figure 2 - Research design

Problem understanding

In order to better understand the problem, a research was conducted by gathering the key concepts about this theme and the current initiatives in Europe. The results of this research are presented in chapter 2 – "Background".

Identification of key concerns

The second step of this research design comprehends the identification of key concerns. It reflects a comprehensive examination of the problem in two scenarios: European scenario described in literature and regional scenario (North of Portugal).

The first scenario was explored in study A, a literature was performed in order to understand the gap between users' needs and actual implemented CRs systems.

Regarding the second scenario, RORENO system was analysed through final users interviews and participant observation of the system working.

Understanding the problem solving (results)

After the collection of evidence in study A and B, these evidences were analysed and systematized into different categories. The results of each study are described in chapter 4 and 5.

Discussion about a new model

The results of study A gave a better vision about Europe needs about CR helping to better understanding about RORENO needs in comparison with Europe.

In chapter 6, was presented a reflection about possible improvements to RORENO to increase its performance and quality.

4 Study A - Literature review

This chapter presents Study A: Literature review and is organized into introduction, methods, results and discussion.

4.1 Introduction

Important steps have been taken at European Union-level in recent years towards mapping and understanding challenges, identifying best practices, and creating the policy frameworks and the tools for cooperation and information sharing.

Although cancer has now become the second cause of death in Europe, one third of the population still lacks quality cancer registration, mostly in the regions with lowest resources and health status (Leal *et al.*, 2016). It is therefore imperative that the efforts to support the development of CRs continue, and that the wealth of knowledge and vision acquired in this area is transformed into action.

This section explores the first objective of this dissertation:

(1) To analyse the main characteristics and concerns about CRs available in the literature.

And answered the following research questions:

(1) Which are the main characteristics of the implemented cancer registries?
    a. These characteristics/ functionalities fulfil with cancer registries requirements and needs?
    b. Which are the main cancer registries problems, concerns and solutions documented in the literature?

4.2 Methods

To achieve this objective a systematic review was conducted. The protocol used was the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (Moher D, Liberati A, Tetzlaff J, 2009).

PRISMA is an evidence-based minimum set of items for reporting in systematic reviews and meta-analyses. PRISMA focuses on the reporting of reviews evaluating randomized trials, but can also be used as a basis for reporting systematic reviews of other types of research, particularly evaluations of interventions (Moher D, Liberati A, Tetzlaff J, 2009). The following topic describes the systematic review protocol.

Following the PRISMA checklist, that includes items deemed essential for transparent reporting of a systematic review, the items below present systematic review methodology.

Eligibility criteria

Studies describing or evaluating the use of information systems to perform cancer registry were selected.

Review team

The review team were composed by three medical informatics specialists.

Information sources

Studies were searched in April 20, 2017 in bibliographic databases. Four distinct bibliographic databases were searched: Medline (via PubMed) (Medline, 2017); ISI (ISI Web of Knowledge) (Knowledge, 2017); IEEE (IEEE Xplore) (IEEE, 2017) and Scopus (Elsevier, 2017).

Search methods

Only articles written in English were included. No criteria for publication date were established.

The query search string used in Medline® was:

"cancer registries"[All Fields] AND computerized [All Fields].

A similar query was used in the other databases and was adapted to the search engine.

Study selection

The study selection process is illustrated in Figure 3. This diagram follows PRISMA statement steps namely: Identification, Screening, Eligibility and Included. The character *"n"* represents the number of records selected in each step.

SCREENING

The first selection was based on its title and abstract. The selected articles were reviewed by two distinct reviewers that collaborate in the whole process. The study was considered eligible when at least one of the reviewers decided that the title/abstract mentioned the key concept of cancer registries solutions.

In cases of disagreement, a consensus meeting was held to decide whether the article should be selected.

The second phase of study selection was based on the full text. In this stage, articles were excluded based on the following criteria:

(1) Language (some articles had title and abstract in English but full-text in other language).

(2) Relevant articles for the study based on the development of the following key-concepts:

   a. Data collection;

   b. Completeness, validity and comparability;

   c. Standards;

   d. Data protection legislation;

   e. Data exploration.

Eligible articles were tabulated and used in the qualitative synthesis.

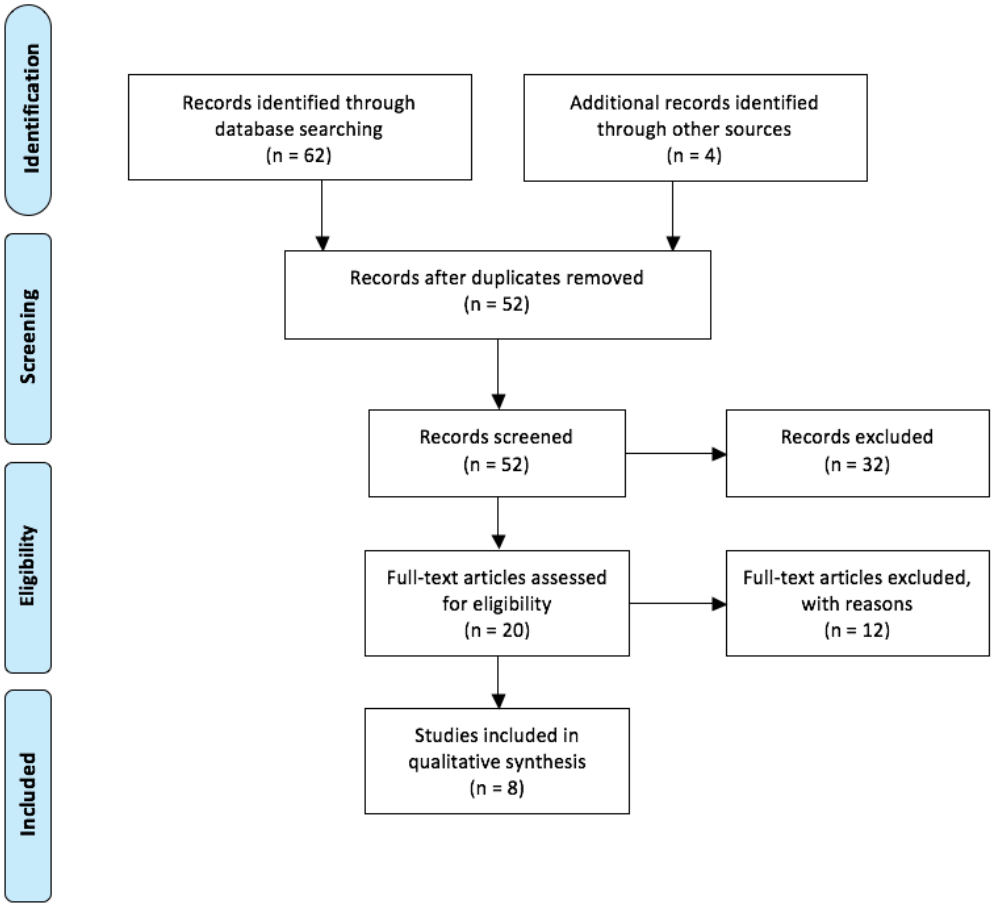


Figure 3 – Flow of information through the different phases of the systematic review following PRISMA.

Data collection process

For each article, two of the reviewers completed the data collection form. If reviewers disagreed, a third reviewer adjudicated. As our analysis concerned only published data, we did not seek to obtain further data from authors.

Data analysis process

In this phase, it was analysed the records content. Each of the eight papers was analysed and was collected key findings about the five key-concepts.

It was decided to organize the main findings according with: (1) data collection, (2) completeness, validity and comparability, (3) standards, (4) data protection and (5) data exploration.

The results were systematized into a table using visual symbols to quickly understand papers scope.

4.3 Results – qualitative synthesis

This sub-section presents the literature review results. First is presented the study selection, a report of key-concepts as described by authors and then a comprehensive analysis of this key-concepts.

Study selection

The selection process is illustrated in the follow table (Table 3). The eight articles were organized chronologically and characterized using symbols (see description below).

Table 3 - Results resume

Reference	Data collection	Completeness, validity and comparability	Standards	Data protection	Data exploration
(Tafazzoli *et al.*, 2002)	💻		✓		🗀
(Contiero *et al.*, 2008)	💻🗐	●	🖹		
(Bjugn, Casati and Norstein, 2008)	💻		✓🖹		
(Siesling *et al.*, 2015)				🔒	🗀
(Andersen and Storm, 2015)				🔒🔒	
(Coebergh, 2015)	💻🗐	●☑⚖		🔒	
(Rossi *et al.*, 2015)		●☑⚖	🖹	🔒	
(A.-M. Forsea, 2016)	💻🗐	●☑⚖		🔒	

💻 - automated collection | 🗐 - manual collection

● - completeness | ☑ - validity | ⚖ - comparability

✓ - communication standards | 🖹 - codification standards

🔒 - mentioned data protection legislation | 🔒🔒 - full dedicated to data protection legislation

🗀 - data exploration

Study characteristics presented by authors

In this sub-section is presented for each key-concept the main findings collected from the papers presented in Table 3.

DATA COLLECTION

- **(Tafazzoli et al., 2002**) presents the GTDS (Giessen Tumour Documentation System). This system supports the management of the "*Tumorbasisdokumentation*", a comprehensive data standard for hospital CR in Germany. GTDS receives automatically data from the hospital admission, discharge and transfer (ADT) server and from various ancillary systems like clinical laboratory, radiology etc.

- **(Contiero et al., 2008)** studies an automated software cancer registration called Open Registry adopted by the Varese (population-based) Cancer Registry since 1997. The authors described that this system collects data information from various sources by a mixed reporting. Automated case registration uses electronic files that contain coded data associated with demographic data, usually with a unique identification code (typically social security number). In some cases, the information is not available in electronic format (for example pathology reports), in these cases, the information is manually inserted in the database. The manual system requires three health professionals to perform case finding, two clerks for data input and linking, one for case generation, a physician to verify cases and a computer specialist/ statistician to manage the database. The authors explained that this passive reporting, that generates 41% of the cases, have a total estimated (annual) cost of 250,000€ that includes costs with hardware and software licenses, and human resources. One the other hand, active reporting required only one clerk for checking information sources and linkage and one and a half health professionals were required for cases verification, a physician was still required to resolve complex cases and a computer specialist to manage information acquisition, record linkage and manage programs. Adding software and hardware the cost estimated is 150,000€.

- **(Bjugn, Casati and Norstein, 2008)** presents the project Cancer Registry and the Norwegian Society for Pathology that aims to (1) develop standardized templates in database format for histopathology reports on cancer resection specimens and (2) develop an Extensible Markup Language (XML) standard to facilitate future electronic transfer of cancer reports from hospitals to the Cancer Registry. The template is based on international guidelines and classification systems and is fully integrated into software being used by all pathology laboratories in Norway. This work establishes a proposal for a template to promote automatic integration.

- **(Coebergh, 2015)** study presents a historical and longitudinal developments of the roughly 160 cancer registries that emerged since 1927 and accelerating since the late 70s especially in southern and continental Europe. Regarding data collection, they explain that data collection and coding activity of the cancer registries are, for reasons of uniformity and complexity, done largely by registration clerks (connecting with the many specialties) and low profile data analysts. Although data collectors are currently continuously adapting to hospital-specific applications of Electronic Patient Records (EPR). Author explains that there is a potential for all sorts of data linkages of cancer registries with other clinical and public health research cohorts. They claim that is

essential, but thus far appeared rather dependent on national interpretation of the EU guideline on Data Protection of 1995, only gradually converging.

- **(A. Forsea, 2016)** outlines the situation of population-based cancer registration in Europe, highlighting its challenges and opportunities, to support the various efforts involved in improving, expanding, and harmonising cancer registration. In this paper, she states that both active and passive methods of data collection are used in Europe. Automatic registration is mostly used by registries in North-Western Europe, where cancer case reporting is also mandatory by law and facilitated by the wide use of digital records. Manual registration is more frequent in Central and Eastern European countries. In countries with multiple regional cancer registries, different methods or combinations may be used, as in Italy, France, or Switzerland.

COMPLETENESS, VALIDITY AND COMPARABILITY

- **(Contiero *et al.*, 2008)** study assessed the completeness of the automatically generated data by comparison with a gold standard of all cases identified by manual and automatic systems for the year 1997 when the automated system was introduced, and the manual system was still in operation. The results revealed that the automatic procedure lost 1.3% of cases compared to the gold standard incidence. More than half (0.8%) of these lost cases were registered in 1998, though diagnosed in 1997. The residual 0.5% of lost cases seems an acceptable figure, as it compares favourably with the 2.2% up to 41% of cases reported lost in the other completeness of other studies and to the 10% of cases lost by the manual system of the cancer registry in 1997.

- **(Coebergh, 2015)** cites a EUROCOURSE paper on completeness and timeliness of cancer registries (Zanetti *et al.*, 2015) justifying the need for a systematic and timely input checks, often in need of involvement of a diverse medical source. In terms of comparability the author focus on a need of clear algorithms for accurate estimations of risk, detection, prognosis and side-effects. (Coebergh, 2015) explains that given the main purposes of a population-based cancer registries, it is essential paying attention for good practices prioritises and removing obstacles for completeness, timeliness and validity, especially undue attention for privacy. Attaining completeness of a cancer registry is crucial for its added value to research with selected patients and its external validity. Examples are analyses of clustering, of time trends in incidence, and for studies of process and outcome for example survival studies. A lack of completeness promotes mistrust and lack of collaboration among institutions and professionals, and a cost-effectiveness.

- **(Rossi et al., 2015)** describes in their work data check procedures providing cancer registries data quality indicators. They studied individual records for all cancer cases diagnosed up to 2007 with vital status updated to 31st December 2008 anonymized provided by EUROCARE web-portal. Records covered 99 eligible registries representing 29 countries (including Portugal). The study includes an automated data quality checking procedure organized in three different phases: (1) formal adherence to the range of validity of each variable was verified; (2) consistency of combinations of two or more fields was checked. From these phases 68,000 records with errors and warnings were identified. And the last phase included a more detailed examination of specific variables.

- Quality control like completeness and validity are highly variable across Europe as mentioned by **(A. Forsea, 2016)**. In 2011, a EUROCOURSE survey analysed the responses of 116 of the 179 European general cancer registries, covering 280 million inhabitants of 32 European countries. A 12% of all responding cancer registries reportedly did not conduct any completeness assessment, mostly because of lack of resources. The newest, complex, software-based methods of evaluation are used by a minority of cancer registries. These results prompted the expert consortium of EUROCOURSE to formulate a set of recommendations for improving and harmonising quality control by cancer registries.

COMMUNICATION AND CODIFICATION STANDARDS

- **(Tafazzoli *et al.*, 2002)** explores GTDS architecture. The authors explained that they used HL7 (Health Level 7) standard to integrate data from Hospital Information System (HIS) to the GTDS (i.e. ADT and laboratory data, pathology reports etc.). They implemented also BDT that is a specific German ASN.1- like format for the transmission of medical record data between physician's office computer systems and has become a *quasi* standard in Germany.

- **(Contiero *et al.*, 2008**) in this paper shows that the information collected follows the Standard International Classification of Disease (ICD-9) (WHO, 2017) and according to ICD-O-2 (IACR/WHO, 2017) in the case of diagnosis date, and demographic information. Regarding pathology reports the terminology adopted is SNOMED morphology codes. Then they need to transform the SNOMED codes on the pathology files into ICD-9 codes, for these they wrote a subroutine in Open Registry that makes use of specifically compiled tables.

- **(Bjugn, Casati and Norstein, 2008)** in their work propose a Norwegian adaptation (Norwegian Directorate of Health, 2012) of the Systematized Nomenclature of Medicine (SNOMED) codes (SNOMED, 2017) is integrated into the software. Also, proposes an Extensible Markup Language (XML) standard to facilitate future electronic transfer of cancer reports from hospitals to the Cancer Registry as referred before.

- The study protocol presented by **(Rossi et al., 2015)** included demographic characteristics, dates of diagnosis, death or last known life status (including day), life status, tumour topography and morphology collected according to the International Classification of Diseases for Oncology, 3rd edition (ICD-O-3) and basis for diagnosis. This is the only terminology mentioned in this work.

DATA PROTECTION

- **(Siesling *et al.*, 2015)** presents the results of an extensive survey of cancer registration practices and data use was conducted among 161 population-based cancer registries across Europe. In this study, they find out that the conditions for cancer registration varied substantially among the responding cancer registries according to European sub-regions. (Siesling *et al.*, 2015) registered that in about 25% of all cancer registries an informed consent appeared to be required to register a cancer patient, but this was often implicit and to be waived (no signature required). In case of research if the data

remained non-identifiable for data users some cancer registries could work through "opt out" choice for patients.

- **(Andersen and Storm, 2015)** focuses their study in the reform of the European data protection framework and the challenges for cancer registration. GDPR states articles concerning processing of personal data concerning health (article 9), the right to be forgotten (article 17) and processing of data for historical, statistical and research purposes (article 83). These articles contain exceptions without which cancer and other routine monitoring of diseases, survivorship, treatment outcome and research into risk factors for diseases will terminate abruptly. For example, article 9.2 prohibits any processing of personal data concerning health, but exceptions are made for cancer registration and public health research. (Andersen and Storm, 2015) explains that epidemiological research depends on the balance between preserving patients' integrity and anonymity while also enabling important research to improve people's health and the quality of care. They exemplify that pseudo-anonymization and encryption systems for example, are influenced by minor errors in the data used in the anonymization process. Such errors will appear unrelatedly how perfect a system is rated and increase the risk of missed linkages of data on single individuals.

- Concerning EU regulation **(Coebergh, 2015)** study focuses on Directive 95/46/EC (Conseil, 1995) problems explaining the political diversity of data protection practices of cancer registries across Europe (allowed by the Directive). They exemplify that is essential learn from incidental threats to continuity of population-based cancer registries, i.e. in Hamburg (1982–1991), in former East Germany (1991), and more recently in Estonia, Slovakia and Bulgaria. Authors focuses that Germany and France are still suffering from the late effects of policies by Hitler and Napoleon, respectively violating and overemphasising individual rights, and for this make life difficult for epidemiologic cancer registries, making the process slow, frustrating and costly. By contrast, patients expect to be offered adequate care and surveillance, also at long term, they expect its quality to be secured statistically.

- **(Rossi et al., 2015)** mention data protection as an issue to data quality procedures. They refer that European Union decision makers should try to harmonize confidentiality and ethical issues related to personal health data to balance the right to privacy and the right to health. Data quality, assessed through standard indicators, was generally high. The quality of follow-up data is particularly important in survival estimates. They refer that in some countries access to death certificates may be limited or hampered by regulations on data protection, causing missed data linkages.

- **(A. Forsea, 2016)** explains that cancer cases reporting is mandatory by law in most of the European countries with high quality registers that included in the cancer incidence in five continents series. Personal data protection legislation has a major impact on cancer registries functioning. The author shows with a EUROCOURSE survey, that 20–35% of responding cancer registries reported legal-related barriers to cancer registration across most of Europe, while in the South-West region these barriers amounted to up to 60%. Particularly concerning is the nationally variable barriers in the linkage of cancer registries with other health-related databases like mass screening programmes, biobanks, vital status, and causes of death databases.

DATABASE EXPLORATION

- **(Tafazzoli *et al.*, 2002)** explained that GTDS is generally fed into epidemiological cancer registries, so they decided to implement integrated knowledge-based functions to monitor data quality, especially the adherence to the constraints defined by IARC. The authors think that alerting registrars while they are still involved in a tumour case, the effort involved in responding to a constraint violation is reduced. In this work, they explained that in the future, they intend to collaborate with epidemiological cancer registries in performing an historical comparison to quantify the effect on data quality.

- **(Siesling *et al.*, 2015)** study reports that 92% of their cancer registries sample routinely reported cancer incidence rates, 60% collected follow-up data for estimating cancer survival rates; 79% also exhibited or contributed to, likely more reliable, cancer mortality rates, mainly through national or provincial statistical offices.

4.4 Results analysis

For the sample, it was identified some problems and concerns about cancer registration in Europe. The concerns were organized into: time lag and manual registration; data protection harmonization; data quality; standards harmonization and data exploration.

- **Time lag and manual registration**: (Bjugn, Casati and Norstein, 2008) explained that time lag in receiving required information from health care providers, receipt of paper-based information that requires time-consuming manual registration into electronic systems, and variation in parameters reported and staging systems used by health care providers. The consequences of this are time-delayed registry information, costly procedures, and loss of quality. The author reinforces that in an "ideal" world, all histopathology reports on each particular type of cancer should contain information on the same parameters. Reports should be constructed with the aim of being imported into a database and be electronically transferred to the cancer registry.
  (A. Forsea, 2016) support that a transition to automatic registration will require financial and skilled manpower resources allocated to obtain good quality cancer incidence data using automated registration systems, and that some degree of manual checking will always be required; nevertheless, after initial setup our automated system required less manpower than the manual system.

- **Data protection harmonization:** (Siesling *et al.*, 2015) described that there are barriers to cancer registration and research based on CR data following varying national interpretations of the Data protection Directive of 1995, especially in Germany and France. It might change in the near future with the current version of the emerging new regulation in the EU also affected by the increasing role of big data (Commission, 2016). However this new regulation brings new concerns such as the new restrictions about data protection as described by (Andersen and Storm, 2015). Losing access to register data, compromise data quality and the analysis hereof at the expense of confidentiality and data protection regulations will neither serve the individual nor the society.
  (Siesling *et al.*, 2015) reflected about another concern. An optimal condition for CR is likely to have substantial cost implications for the registration and linkage processes. It therefore remains a big challenge to staff of many CRs and their users to formulate

and define exceptions for population-based clinical and public health research to the current new regulation. Furthermore, it seems obvious that an ENCR working group prepares adequate implementation of the pending data protection regulation. Moreover, the group should advise the CRs and respective patient groups and authorities in the various member states to bring software development in line rather than being overwhelmed by untested proposals.

- **Data quality:** (Rossi *et al.*, 2015) thinks that additional efforts are needed to harmonise cancer registration in Europe (varying registration systems with varying data quality) and to support the collection of accurate clinical information needed to improve data comparability (cancer biology, risk patterns, use of diagnostic tests and screening, treatments).
  (A. Forsea, 2016) rationalized that budget is obviously a fundamental factor influencing the quality and performance of CRs. In her study, she found out that 12% (n=116) of respondents don't implement routines to assess have cancer registries completeness mostly because of a lack of resources.
  (Coebergh, 2015) reinforce this importance saying that is crucial for the value of research its external validity. A lack of completeness promotes mistrust and lack of collaboration among institutions and professionals. (Coebergh, 2015) also adds that is essential paying attention also to timeliness, validity and data protection.
  (Contiero *et al.*, 2008) exposed that a completeness of 98.7% indicates in automatic procedure is a valid alternative to manual methods.

- **Standards harmonization:** Though the literature analyses it is possible to understand that the International Association of Cancer Registries (IACR) have already pronounced about an harmonization about codification standards recommending for cancer registries the use of the International Classification of Diseases for Oncology (ICD-O) (WHO, 2017) to code the topography (site of primary tumour) and morphology (histological type) of the tumours. This terminology was presented by (Rossi et al., 2015) study. However this seems not to be the only terminology adopted. (Contiero *et al.*, 2008) reported in his work the use of ICD-9 and also SNOMED. (Bjugn, Casati and Norstein, 2008) expose that in Norwegian they used an adaptation of SNOMED. Regarding standards in terms of data integration there wasn't identified a recommendation for a standard.
  (Tafazzoli *et al.*, 2002) showed a German system that adopted HL7 standard to integrate a HIS with CR system.
  (Bjugn, Casati and Norstein, 2008) proposed a XML scheme designed by them to simplify the electronic integration from HIS to CR.

- **Data exploration:** All the studies analysed described that the aim of data collected is to elaborate incidence and survival reports as required by WHO.
  (Tafazzoli *et al.*, 2002) reported that they intend to collaborate with epidemiological cancer registries in performing an historical comparison to quantify the effect on data quality. This sample do not identify other potential uses that this data could be not only in an external level but also inside an organization. At this point, this sample do not answer this issue.

4.5 Discussion

Systematic reviews and meta-analyses of individual participant data have been recognized as a gold standard approach from the early days of systematic review (Chalmers, 1993). For this reason, this was the methodology chosen to achieve the first objective of this dissertation: To analyse the main characteristics and concerns about CRs available in the literature.

It was followed the PRISMA protocol in order to select the articles that answer the objectives.

The results totalized eight articles that were analysed according with different key-concepts.

In this study, it was found out different concerns and problems in this field. One of the main concerns that were identified was the time lag in registration. There was a consensus about this issue, it was clear that the lack of automatic integrations between systems difficult cancer registration. Manual registration requires human resources and may imply a higher number of errors. The answer to this problem seems to be known by the stakeholders, the solution will be the automatization of data collection from different sources. A short-term solution could be each cancer registry group decide to invest in this issue, chose a terminology, chose data integration standards, chose data quality indicators, and integrate their own data sources with their cancer registry. However, a better solution would be that the European decision makers define how to do this process in order to allow a harmonization between countries and later extend this scope allowing to integrate information between different cancer registries. In both cases, it will require a highly effort in defining a new model. This effort will implicate the stakeholders and financial investments.

In the last years, a new concern arose in this field that was the need to comply with GDPR. In the past, the concern of stakeholders was the disparities between countries in the application of Directive 95/46/EC. With GDPR this problem seems to be solved since GDPR seems to be more precise and clear (without double meanings) and requires a full application of it articles. Analysing GDPR articles against the point of view of the reviewed articles, it was possible to understand that the restrictions imposed by GDPR will jeopardize public health studies. The restructuring of processes and systems will also of course imply financial investment.

Another finding of this study is that the data collected are only used to elaborate incidence and survival reports. The sample studied does not show a secondary use of this data for example to improve healthcare services. This could be due to GDPR constraints and the difficult to understand: what data is legitimate to treat? who can treat it? in what purpose? what authorization are required? etc.

This study results should be interpreted against its limitations, the English language limitation prevented that other studies were analyzed.

5 Study B - RORENO qualitative research evaluation

This chapter presents the Study B: RORENO qualitative research evaluation. This chapter is organized into four subsections: introduction, methods, results and discussion.

5.1 Introduction

The RORENO (IPO-Porto, 2017b) workgroup begins its activity in 1988 answering the governmental rule 35/88 (República, 1988). This group is held in "*Instituto Português de Oncologia do Porto Francisco Gentil, EPE – IPO-Porto*", a public reference oncologic Hospital (IPO-Porto, 2017a). Since 1988 has been reporting cancer cases that includes data sent by different healthcare institutions of the north of Portugal.

The RORENO geographical area comprehends the following districts: Porto, Braga, Viana do Castelo, Vila Real, Bragança, some cities in Aveiro district (Albergaria-a-Velha, Arouca, Castelo de Paiva, Espinho, Estarreja, Murtosa, Oliveira de Azeméis, Ovar, S. João da Madeira, Santa Maria da Feira and Vale de Cambra), in Viseu district (Cinfães, S. João da Pesqueira Armamar, Lamego, Moimenta da Beira, Penedono, Resende, Sernancelhe, Tabuaço and Tarouca) and in Guarda district (Vila Nova de Foz Côa) (IPO-Porto, 2017b).

In 2005, this workgroup acquires a software named RORENO that facilitates the communication between this network. This system allows to accomplish the core competencies of the group:

- Collection of the most complete and up-to-date data on all new cases of tumours among residents of the Northern Region of Portugal.
- Production, analysis and interpretation of impact indicators of oncological disease and respective publication (national and international level).
- Provide information to all health professionals, researchers, policy makers, health care organizations in the effort to contribute to the prevention and control of cancer diseases (IPO-Porto, 2017b).

The RORENO workgroup is expert in epidemiology and is constituted by a multi-disciplinary team of medical doctors, anatomical pathology technicians, statistical technicians and informatics professionals. The RORENO team is composed by thirteen professionals working in cancer registry of the North of Portugal.

RORENO workgroup follows a confidentiality policy, they think that data confidentiality is extremely important in the operation and maintenance of a registry. Given the need to access individual and detailed information, cancer registries have always paid close attention to the protection of information privacy.

In order to protect the privacy of each cancer patient and to ensure that the information provided is not misused, RORENO workgroup adopts as its own the privacy policy published by the International Agency for Research of Cancer (IARC) (IACR, 2017b) and ENCR (European Network of Cancer Registries) (ENCR, 2016).

RORENO group produces two kind of registries: RO and ROR. IPO-Porto belonging to the list of healthcare organization in the North of Portugal that produce oncological information needs to register ROs (*registo oncológicos* – oncologic records) centered in the historical of a

patient inside the institution. On the other hand, IPO-Porto is also responsible to produce regional reports as mentioned before, so RORENO group as to produce RORs (*registo oncológico regional* – regional oncologic records) that aggregate RO data from all the North healthcare institutions belonging to oncologic network.

In resume, a patient could have different ROs if he/she access different healthcare institutions in the North and only one ROR. Figure 4 shows an example of a patient path through different institutions and the number of RO created and a unique ROR.

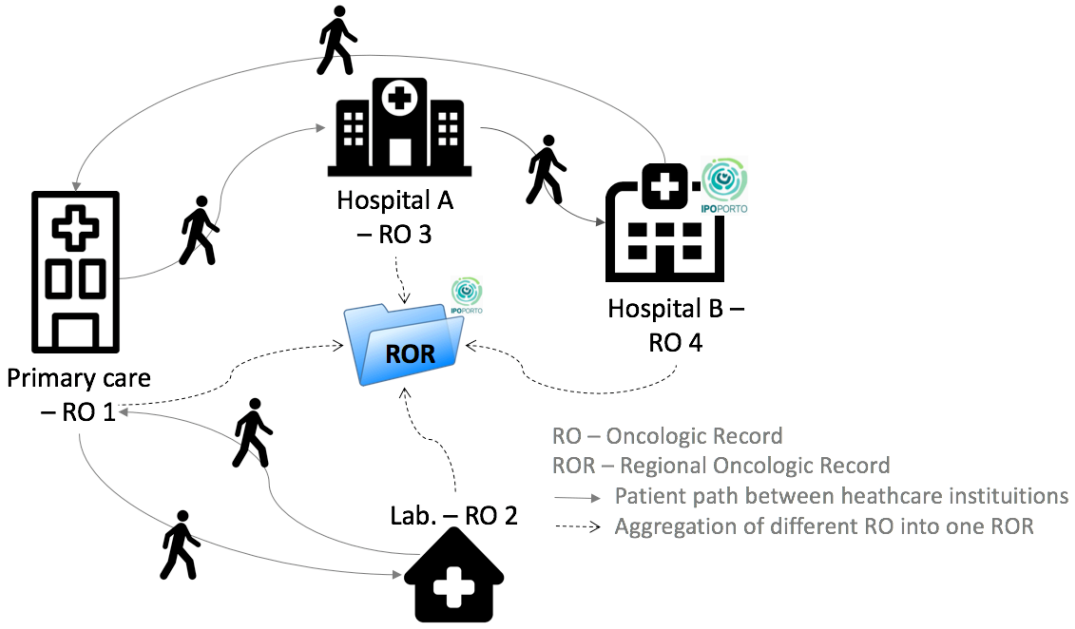


Figure 4 - Exemplificative scheme of patient ROs and ROR

The aim of this study is to evaluate the computerized cancer registry implemented in "*Instituto Português de Oncologia do Porto*" using a qualitative research. The main goal is to characterize the environment of the system, such as the main functionalities and core processes, team involved and different healthcare institutions in the regional network. Then identify the main problems and difficulties that arose in the last year.

With this objective, it is intended to answer the research questions proposed before:

a) Which are the main functionalities, users and stakeholders of RORENO?
    a. Which are the main problems/ constraints felt by the users?
b) Which are the necessary cancer registry characteristics/ functionalities that fulfil the users' needs?

5.2 Methods

In this study, it was applied two kinds of qualitative techniques to acquire information about the health information system in study: semi-structured interview and participant observation. The following sub-section explains how these techniques were applied.

Semi-structured interview

Semi-structured interviews are in-depth interviews where the respondents have to answer open-ended questions and thus are widely employed by different healthcare professionals in their research (Jamshed, 2014). In this study, we collect data among RORENO team through semi-structured interviews. The main topics explored in the interviews were:

- Which are the RORENO personas;
- Which are RORENO roles and permissions characterization;
- Which are RORENO main functionalities;
- How is designed RORENO architecture in terms of data integrations;
- Explanation of the core processes including:
    - How the data is collected;
    - Which are the main sources (both institutions and information systems applications);
    - How is created a new case;
    - How is de-duplicated and validated the data;
    - How is reported a follow-up.

It was realized five interviews with the duration of one hour each, between April and May of 2016. In each interview were present:

- RORENO responsible (n=1) in every interview;
- Statistics experts responsible by ROR (n=2) in every interview;
- Anatomical pathology technicians responsible by RO (n=4), some were present in one interview and another in other.

The information was annotated into paper records, and then systematized into schemes and BPM diagrams. In the next meeting/ interview it was asked to correct the information collected in the meeting before.

Observation

Observation is a type of qualitative research method which not only included participants' observation, but also covered ethnography and research work in the field. Observational data can be integrated as auxiliary or confirmatory research (Jamshed, 2014).

In this study, it was used observation to validate the information collected in the interviews, mainly the core processes workflow. For this together, with the end-user (mostly anatomical pathology technicians and statistical technicians), it was observed how the system works and the different steps to perform a task. After it was designed core processes using BPM in an iterative way until we had an agreement of the correct flow of steps.

None information was recorded due to confidentiality and privacy protection of patient information.

5.3 Results

This section, presents the results of this research. This section is divided into the following sub-sections: RORENO Personas; roles characterization; description of the main architecture; main functionalities, RORENO main processes; service blueprint; RORENO data integration, data quality and discussion.

RORENO Personas

A persona defines an archetypical user of a system, an example of the kind of person who would interact with it (Modelling, 2014).

In RORENO scenario the system was already designed and is in execution so, in this sub-section it is presented an example of users. In this set of users, it is possible to organize the users into two groups: active and passive users.

Active users are the users that interact daily with the system, on the other hand passive users are users that indirectly uses the system. Basically, passive users have access needs however they don't have a suitable area for them. In this scenario, active users are: project director, anatomical pathology team and statistic team; passive users are: administrator/ managers and healthcare professionals / researchers (see Figure 5).

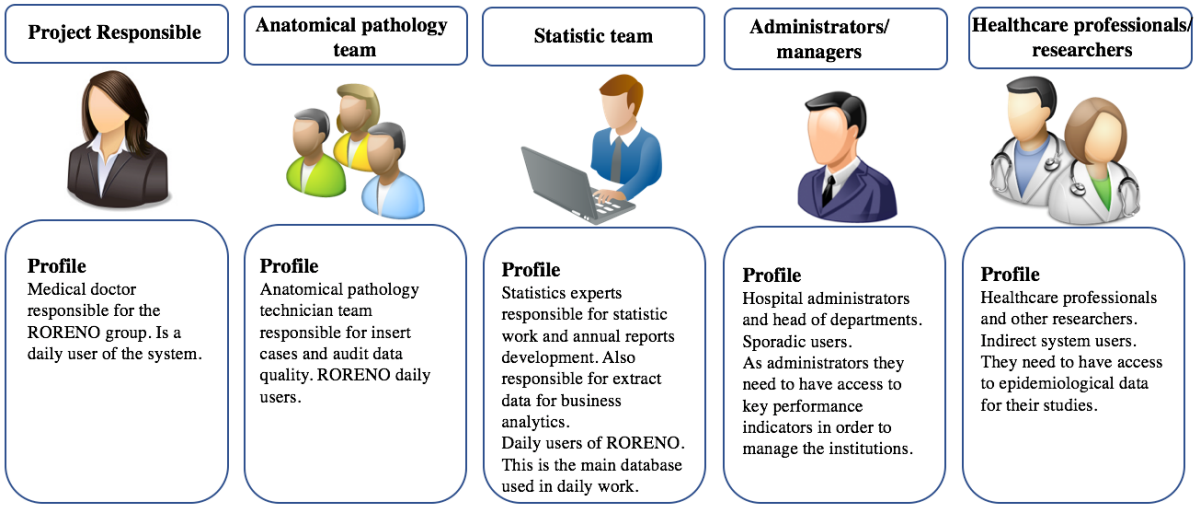


Figure 5 - RORENO Personas

RORENO roles characterization

This sub-section, identifies the main profiles of users and its permissions/tasks (active users) in RORENO system.

In terms of front-end management, we identified: administrator, responsible for entering data and external institution profile.

- ***Role Administrator (ADMIN)***: This role is responsible for the management of users and have the permission to create new variables in the system (allowing the adaptation of the system). Furthermore, of these special functionalities administrator inherences all the permissions of the other roles (RORENO, Data source Institution, default user).
- ***Role RORENO****:* integration of files (Excel) provided by external institutions and data validation (manually and using IARC Check tool (WHO/IARC, 2014)), data deduplication and validation. This role inherence permissions from default user.
- ***Role Data Source Institution***: Each external institution in the region has credentials to access this system and fulfil the cancer registry forms (manually or automatically). They can also insert information about the patients' follow-up and can export and explore the data. The main difference between this role and the role before is that each external institution can only view information about their own patients. This role inherence permissions from default user.
    - ***Responsible for entering data*****:** This role is performed by several users and has in charge the responsibility to insert data in the system. Due to interoperability problems, sometimes they need to insert variable by variable in the system manually. RORENO team need to insert this information from different institutions. They also have to search follow-up information about the patients (e.g. vital state). RORENO team search (manually) in national patient databases such as RNU, contact by phone other institutions, city responsible, families etc. They also use the IARC check tool (WHO/IARC, 2014) to validate the data quality. And have permissions to export and explore data to produce the incidence and survival reports.
- ***User****:* This role defines the transversal functionalities that all the roles described above have: register (first time a user uses the system, login, follow up update, patients' search and data mining).

In terms of backend, the informatics department of the healthcare institution that holds the cancer registry (IPO-Porto) ensure the availability, security of the system and the backup copies to use in case of failure. Currently this system does not have maintenance by the software provider.

These characterization is systematized in the following UML use case (Figure 6).

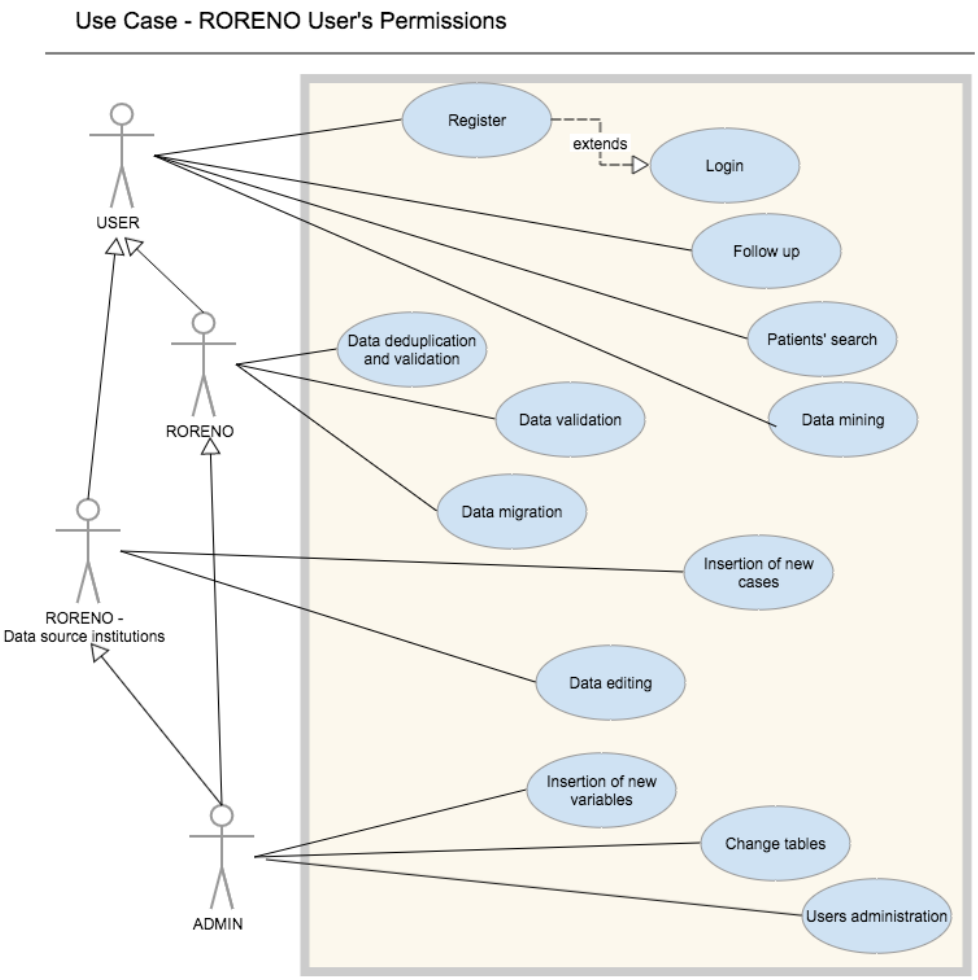


Figure 6 - Use case - users' RORENO

RORENO main architecture

Figure 7 presents a general overview of the RORENO architecture. This system is stored in a reference oncologic hospital – IPO-Porto and managed by RORENO team. RORENO system can be accessed by different healthcare institutions (e.g. hospitals, primary care...) in the North region and are within the national private healthcare network (*Rede Informática da Saúde)*. *Rede Informática da Saúde* is a private multimedia network of the Ministry of Health that interconnects as local networks of its agencies and services. *Rede Informática da Saúde* arise with the need to exchange information and concern for the safety, security and safety of health institutions. Only the public institutions are within this informatics network.

IPO-Porto is an important source of information to collect for RORENO since is the reference oncologic hospital in the North region and it's also responsible to collect the cancer data between the north heath institutions to feed the RORENO system (ROs).

As we can see in Figure 7 RORENO is only integrated with IPO-Porto administrative database that includes administrative information about a patient such as demographic information and schedule (appointments, surgeries, vital state...). The others important sources of information such as departmental information systems, laboratory information system (LIS), radiologic information system (RIS) and drug information system doesn't have an automatic integration with RORENO. So, in this case, RORENO team needs to access these systems one by one, patient by patient and collect manually the information needed.

In the case of other national relevant healthcare databases such as *Plataforma de Dados da Saúde* (PDS - Health Data Platform) (SPMS, 2013) *and Registo Nacional de Utentes* (RNU - national patients' registration database) (ACSS, 2010) that have crucial information to complete the registries, RORENO also doesn't have an automatic integration. The team needs to access these platforms and collect the needed data manually.

After collected the data, RORENO team fulfil a paper form (example available in Appendix A) and then transcribe for RORENO electronic form.

The external institutions collection of data is explained better in sub-section data integration.

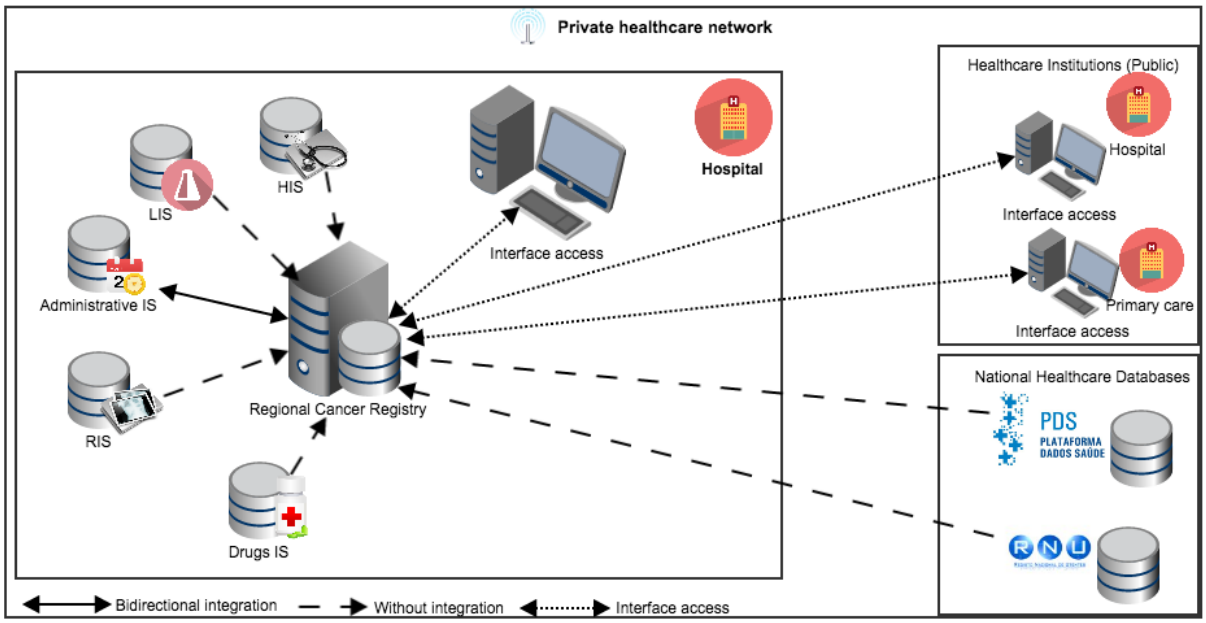


Figure 7 - RORENO architecture scheme – main integrations

RORENO main functionalities

RORENO information system has many functionalities, the main identified in this study are:

- Forms to insert information of an individual case (RO);
- Suggestion for data aggregation (information from several institutions from the same patient and same diagnosis);
- Excel export option (important to perform data analysis and produce incidence and survival reports);
- Excel import option (allows to insert excel data files);
- Follow-up forms (possible to receive and send automatically this information from/to the hospital administrative system).
- Worklist of new patients accepted in the hospital (information sent by the administrative hospital system) that is needed to complete medical information.

Figure 8 describes a RORENO process map. A process map comprehends a workflow diagram to bring forth a clearer understanding of a process or series of parallel processes (White and Cicmil, 2016). In Figure 8 was represented the main processes to complete activities displaying the trigger and the relations between them and the output. The left column represents different triggers, and the right column represents outputs. The boxes in the middle are core processes of this system and its relations.

By analysing the Figure 8 it is easy to understand that several activities are handheld such as information collection to code a new case (ROR) and search information about vital state to complete a follow-up. To complete this information in some cases the team needs to telephone to familiars or either contact the local or national authorities.

It is possible to observe also that they upload some information sent by other institutions through excel files (RO). After upload, they need to validate the data quality using a IARC tool. This tool isn't integrated in RORENO system.

In a quick analysis, it is possible to recognise that these activities that support core processes are time-consuming tasks and easy to add errors. This fact justifies some time-delay in reports production.

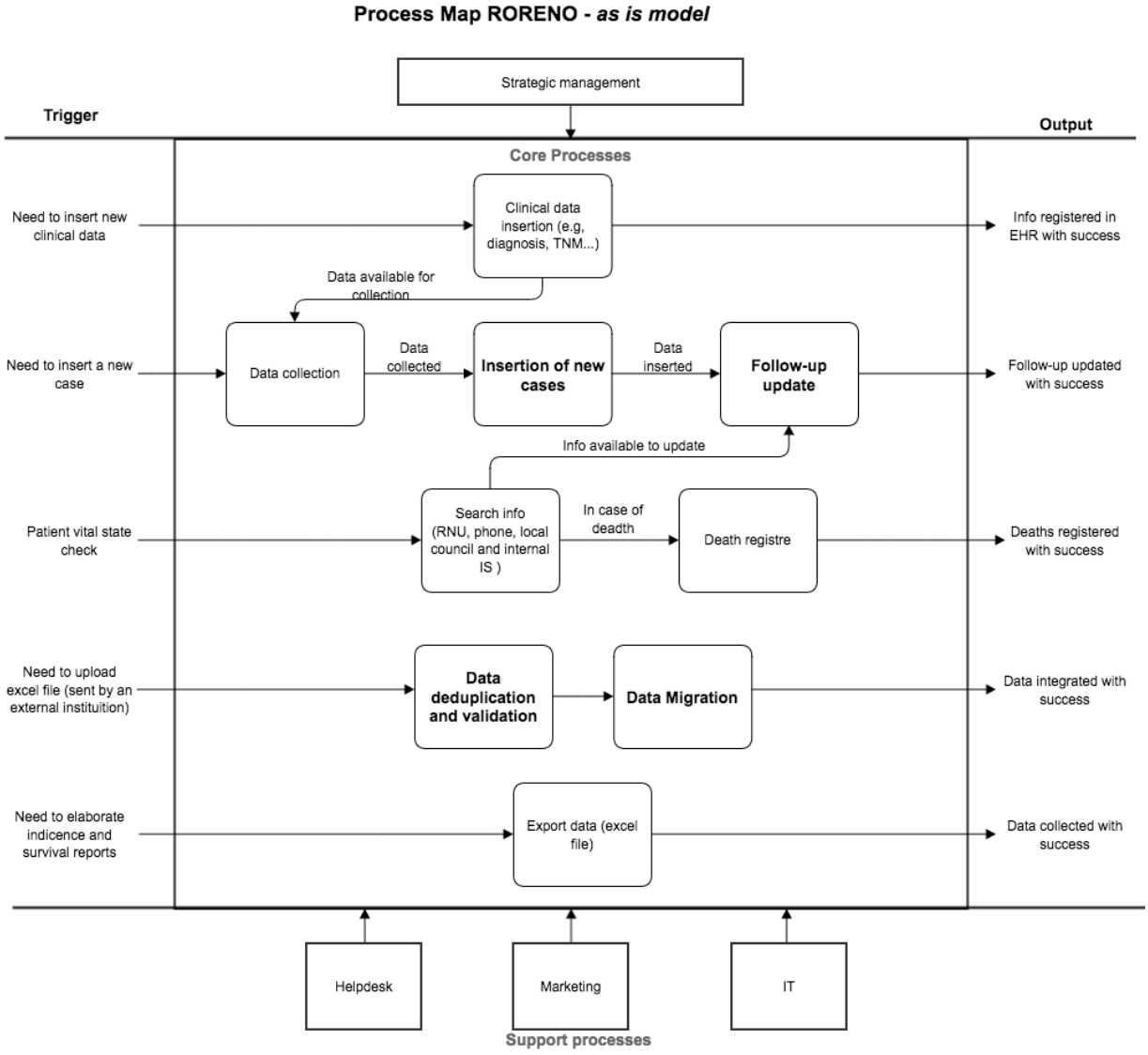


Figure 8 - Process map RORENO

The next sub-section describes step by step the core processes highlighted with bold in Figure 8.

RORENO core processes

Core process can be defined by a process with a set of related and interdependent activities that transform an input to a system to an output with added value to a customer in this case a patient. In this sub-section will be present four core processes: (1) insertion of a new case; (2) follow-up update; (3) data deduplication and validation and (4) data migration.

NEW CASE

Figure 10 available in Appendix B represents through a Business Process Management (BPM) diagram the different steps to accomplish an insertion of a new case in RORENO system.

Table 4 presents a fact sheet with the different phases; trigger; main activities; actors involved and expected results.

Table 4 - New case process - fact sheet

FACT SHEET	
Process ID	P1
Process name	New Case
Process description	This activity allows the insertion of a new case related with a patient. This process includes: search of new patients, data collection and manual codification of the information collected (phase 1), selection of patients to be codified (phase 2) and creation of a new oncologic record (phase 3).
Trigger	Need for case codification
Main activities	Phase 1 – Search new patient (unclassified); Phase 2 – Patient selection; Phase 3 – New record
Actors involved	Anatomical pathology technician
Expected results	Data codified with success – new record

This process was reported by RORENO team as a crucial process for cancer registration. In some scenarios RORENO team need to manual insert cases into RORENO system. The two main scenarios are: (1) external institutions that sent data through paper records (e.g. exams reports) and (2) RORENO team needs to manually collect and insert IPO Porto data into RORENO.

RORENO team reported that data collection activity is one of the most time-consuming activity since an anatomical pathology technician needs to access patient by patient from different database systems and fill an organized paper form (Appendix A) as mentioned before.

UPDATE FOLLOW-UP

Figure 11 available in Appendix B represents through a BPM diagram the different steps to accomplish update follow-up in RORENO system.

Table 5 presents a fact sheet the different phases; trigger; main activities; actors involved and expected results.

Table 5 - Update follow-up process - fact sheet

FACT SHEET	
Process ID	P2
Process name	Update follow-up
Process description	This activity allows the obit confirmation of the oncologic patients to elaborate the annual survival reports. This process includes first the exportation of data from the year in study (phase 1), vital state confirmation (phase 2) and follow-up register (phase 3).
Trigger	Need for elaboration of annual survival report.
Main activities	Phase 1 – export cases; Phase 2- obit confirmation; Phase 3- follow up register.
Actors involved	Anatomical pathology technician and Statists experts
Expected results	Follow up updated with success. Data available to produce survival reports.

In this process, a manual search on RNU system is required to update follow-up on RORENO system. This could be interpreted as a bottleneck in the whole process.

Here it is possible identify as a new functionality the automatic integration between RORENO and RNU as occurs in other scenarios. For example, in the case of admission-discharge-transfer (ADT) system (central administrative database in healthcare institutions), in some hospitals are already integrated with RNU, allowing integration of updated demographic data.

In RORENO scenario, they need access RNU manually and search patient by patient the follow-up information required to update RORENO system.

DATA MIGRATION

Figure 12 available in Appendix B represents through a BPM diagram the different steps to accomplish data migration in RORENO system.

Table 6 presents a fact sheet the different phases; trigger; main activities; actors involved and expected results.

Table 6 - Data migration process - fact sheet

FACT SHEET	
Process ID	P3
Process name	Data migration
Process description	This activity allows the integration of data provided by external institutions. This data is provided in excel files and needs to be validated (codified, corrected, completed…) (phase 1) and then integrated in RORENO database (phase 2).
Trigger	Need for case codification
Main activities	Phase 1 – Codification /Complete / Correct data; Phase 2 – Integration
Actors involved	Statistics expert
Expected results	Dada integrated in RORENO with success

This process is required to response two distinct scenarios: (1) some institutions does not have access to RORENO system since they are private institutions; and (2) some institutions do not have the means to use RORENO system, because they do not have the resources available to collect data and they do not have means to integrate automatically their own data sources with RORENO system.

For these reasons RORENO team needs to lead with excel files from multiple institutions. The main problem reported by RORENO team, is that each institution sends an excel file formatted in their own way.

There are some issues that obligated RORENO team to validate carefully all data, for example: different codification standards for the same variable, different variables in a single cell, the use of age as a variable instead the date of birth and etc. This manual validation implies an extra effort for RORENO team. So, they reported that a new system should automate this process, keeping only difficult data conflicts to be solve by a human resource.

DATA DEDUPLICATION AND VALIDATION

Figure 13 available in Appendix B represents through a BPM diagram the different steps to accomplish data deduplication and validation in RORENO system.

Table 7 - Data deduplication and validation process - fact sheet

presents a fact sheet the different phases; trigger; main activities; actors involved and expected results.

As the previous process, this process could improve with automatic validation rules. RORENO team receives data from different kind of institutions in the north of Portugal: hospitals, private clinics, laboratories etc. In some cases, they receive the same data about a patient from different institutions, for example information about the same exam or patient demographic data. In order to catch these cases, RORENO team needs to export data to an excel file from RORENO system and then manually search these cases and again manually change case by case in RORENO system in order to correct the data in RORENO system.

RORENO team reported this issue as an important issue to address in a new system. This manual identification of duplicated data implies an extra effort from RORENO team.

Table 7 - Data deduplication and validation process - fact sheet

FACT SHEET	
Process ID	P4
Process name	Data deduplication and validation
Process description	This activity allows to associate different administrative data from the same patient (phase 1), organize this data (elimination of duplicated data) and validate in a unique regional oncological register (phase 2).
Trigger	Need for case deduplication to regional oncological register validation
Main activities	Phase 1 – cases association Phase 2 – deduplication and validation
Actors involved	Anatomical pathology technician and Statists experts
Expected results	New regional oncological register (ROR)

RORENO service blueprint – new case

A service blueprint is an applied process chart which shows the service delivery process from the customer's perspective. In this sub-section is presented an example of a service blueprint for an important activity performed in RORENO by a user. Figure 14 available in Appendix C shows this diagram, the activity illustrated is the insertion of a new case also described in Figure 10 as a BPM diagram.

In this diagram, it is possible to understand the interaction between the user, in this case an anatomical pathology technician responsible by the insertion of new cases (RO), the RORENO frontstage (interface), RORENO backstage (backend) and support processes.

The number of "clicks" needed to perform this task are low, in the interviews the users didn't complained about it. They only complained about the step of insertion by hand of data that requires excessive time and effort.

So, in general they feel that if the system integrates this information automatically this task would much more easy and quick, avoiding human errors.

RORENO data integration

The main goal of RORENO system is to aggregate cancer registries from different healthcare institutions in the region. The set of external institutions that belongs to RORENO circle are presented in Appendix D. In 2011, they totalize thirteen-five (35) healthcare institutions that together contributed with 37 802 records (last data provided by RORENO group) (RORENO, 2017).

The system is almost a stand-alone system as mentioned before and RORENO team need to insert manually the data in the system.

Although the institutions had the possibility to use the RORENO forms to insert their own data, for some institutions is more convenient to send the information by email (through an excel file) or by post office (paper-based records) and ask some assistance to the professionals that work with the system to insert this data.

Figure 9 shows a scheme that summarizes the three kinds of processes selected by external Institutions. Figure 9 classifies the institutions into three main types:

- Type 1 classifies healthcare institutions that insert data in RORENO (manual or automatic – in this case it wasn't possible to get this information);
- Type 2 refers to institutions that sent an excel file with data through email to be imported by RORENO and;
- Type 3 shows the case of institutions that sent through post office paper based records (e.g. laboratory reports).

In a total of thirteen-five (35) institutions that belongs to this region, six (6) are a healthcare institution type 1, twenty-three (23) are institutions type 2 and six (6) are institutions type 3.

These institutions include public hospitals and health centres and private pathological anatomy laboratories. Since private pathological anatomy laboratories cannot access the private network of the Portuguese Ministry of Health they cannot share their data through a direct integration. The only options they have is to send the information (through excel file or paper).

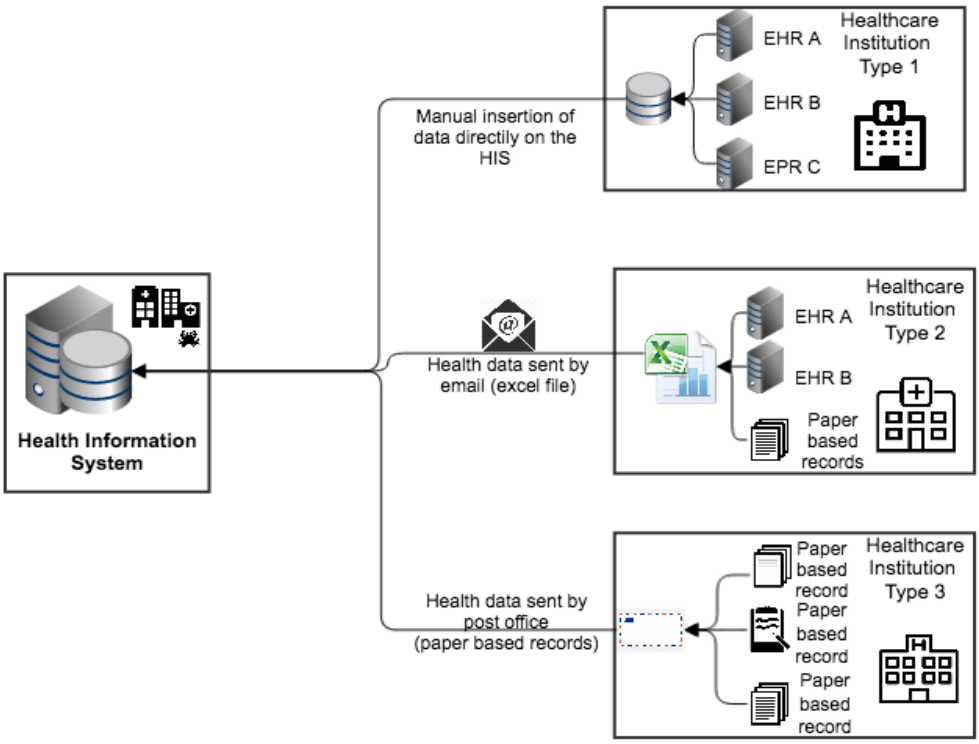


Figure 9 – RORENO data integration model

Data quality

Since they receive a significant amount of data by excel files to integrate in the system, they use a tool (IARC check (IACR, 2017b)) to scan the file and verify some inconsistencies between variables, for example (IPO-Porto, 2017b):

- Age, Incidence Date, Date of Birth
- Age, Site, Histology (ICD-O-3 classification)
- Site, Histology (ICD-O-3 classification)
- Sex, Site (ICD-O-3 classification)
- Sex, Histology (ICD-O-3 classification)
- Behaviour, Site (ICD-O-3 classification)
- Behaviour, Histology (ICD-O-3 classification)
- Grade, Histology (ICD-O-3 classification)
- Basis of diagnosis, Histology (ICD-O-3 classification)

This tool isn´t integrate in RORENO system. They have to “scan” the excel files, correct the problems and then migrate to RORENO.

5.4 Discussion

From the interviews with the end-users of the system, the IARC check (IACR, 2017b) isn't enough to verify the data quality, they need to make an extra effort to manually check if there are: duplicate data, incomplete data, and other types of errors. This data validation plus the insertion of almost all data manually is a process very slow that requires time from various professionals.

One of the reasons that RORENO team think other institutions don't use RORENO system is because a lack of budget to have professionals dedicated to insert the data in the system since they do not have an automatically integration implemented.

By now RORENO team published the reports of regional cancer incidence and cancer survival of 2011 year (RORENO, 2015) and are working on 2012 data. Since we are in 2017 this delay is a problem. However, this seems to be a world-wide problem. In this research it was searched for the newest cancer incidence report and survival reports in Europe and the most recent numbers are from 2013-2014 (International, 2017)(UK, 2014)(Ferlay *et al.*, 2015).

Users also identified minor errors in RORENO software such as lack of keyboard shortcuts, like click in "enter" to jump to the next box, hidden buttons etc. This usability problems were considered as minor problems however it has an impact in a tool since is daily used by the team.

Another issue reported by the team, is the impossibility of other users in IPO-Porto to access some statistics. Every week they receive requests from healthcare professionals such as medical doctors and pharmacists. In this case, after an approval of responsible, they have to select the data manual and gave access through excel or another format. Sometimes, different people ask for the same data. They feel that this could be automatize in the future, with a formal authorization of access.

Regarding follow-up date, RORENO team suggests the automatic integration with the national database – SICO (*Sistema de Informação dos Certificados de Óbito* - Information System for Death Certificates). SICO was released in 2014 with the purpose to allow the entities involved in the process of certification of deaths to articulate, in order to promote an adequate use of resources, to improve the information quality and speed of access to data in a secure environment and respect for the privacy of citizens .

In this study, it was made a maturity diagnosis of the computerized cancer registry RORENO implemented in 2005. Since 2005 the informatics in the healthcare institutions have been evolving and the sources of data that were paper-based are nowadays robust electronic systems that already supports standardized protocols for integration like HL7 (*Health Level 7*) and DICOM (*Digital Imaging and Communications in Medicine*). This scenario reflected not only the healthcare institutions but also the national projects promoted by the Portuguese Health Ministry like national patients' database.

6 RORENO – going forward

This chapter, intends to introduce some enhancements as a proposal to improve RORENO system performance and consequently an improvement of RORENO team performance. This chapter is divided into two sub-sections: (1) improvements proposal and (2) key performance indicators.

6.1 Improvements proposal

The upgrade to a new system is an important step that will have direct impact in the well-being of more than 3 million persons that lives in the North of Portugal and an indirect impact in all Portuguese population (see Table 10) (RORENO, 2017).

By analysing the results of study A and B, was formulated a set of recommendations for RORENO improvement. The proposal could be applied in the actual version of RORENO system or included in a new CR system adopted by IPO-Porto.

The recommendations are divided in the follow categories: automatic integrations; data protection, business intelligence functions and data quality and are presented as requirements.

Automatic integrations

- The data collection should be embedded in the health care processes themselves, promoting an automatic and synchronous collection of information;
- Inside IPO-Porto, the variables needed to complete a RO should be connected to the IPO-Porto integration platform that posteriorly allows system integration;
- In an institution level, it is recommended that each institution should integrate automatically their data in the system;
- In a national level, it is recommended to integrate with PDS, RNU, SICO and RON.
- The communication interface shall use the official codification terminologies (ICD-9, ICD-O-3 and SNOMED CT), or those that SPMS (entity that manages PDS, RNU, SICO and RON) determines necessary for integration.
- The communication interface shall use HL7 standard.

Data quality

- The system should include a rule mechanism capable of allowing the selection of cases that meet certain criteria based on the patient variables and are associated with computable guidelines.
- The system should include rules described using appropriate languages for clinical decision support systems, namely Arden Syntax or Guidelines Definition Language (GDL);
- The system shall provide quality control procedures with measures such as:
    - Detection of duplicate records;
    - Detection of lack of data and use of unknown or poorly defined codes;
    - Number of tumour records submitted;
    - Agreement rates.
- Quality control routines must be automatic and periodic;

- The system should allow automatic detection of potentially duplicate cases, based on patient name, date of birth, age and national health number, creating a worklist for manual resolution
- The system must perform automatic error checking on the data using the rules defined by JRC-ENCR, making it impossible to introduce major errors and issuing automatic minor error alert warnings (ENCR, 2016).

Business intelligence

As in any other organization, monitoring the key registry operations is very important and leads to considerable saving in overall costs of registry activities.

It is recommended to add to this system a BI tool to monitored certain indicators, for example clinical indicators related with treatments, pathologies etc., or management indicators for example time to schedule a consultation, consume and cost of new drugs.

A BI tool should include the following functions:

- Monitor a set of management indicators and specific clinical indicators;
- The set of management indicators should comprehend the following operational areas: new patients; outpatient; inpatient; patient charges; logistics and pharmacy; billing; inpatient waiting list; diagnostic and treatment procedures waiting list and so on.
- Provide a differentiated section for management indicators and clinical indicators;
- Create interactive and complete reports;
- Only authorized users should have access to this tool, after a valid authentication;
- Should generate notifications / alerts if the results are not within a pre-defined threshold (kpi target).

Data protection

- System should be compliant with GDPR from May 25, 2018;
- It is recommended to perform a PIA (Privacy impact assessment) in order to understand which are the security risks demonstrating GDPR compliance.
    - PIA is performed by analyzing the information stored and exchanged between different systems of this project. This assessment aims to evaluate the purpose of these treatments, the data that are maintained in the systems, the security measures implemented for data protection, and the evaluation of data access rights.
- System should provide security mechanism to ensure data protection regulation (GDPR) compliance such as:
    - Authentication service that is responsible for ensuring the authentication and identity validation layer for all ODISSEIA systems. The following authentication methods are supported either in isolation or in combination (e.g. factor 2):
        - Login / password - Including connectors for Lightweight Directory Access Protocol (LDAP) and Active Directory (AD) validation.
        - Smartcard Citizen Card (CC) and Card of the Physicians Order.

6.2 Key Performance Indicators

Key performance indicators (KPIs) can defined as quantifiable measurements of the improvement or deterioration in the performance of an activity critical to the success of a business. In this context, KPIs are an important tool to measure the improvement in different domains in the implementation of a new model compared with an old model.

In this subsection, it is proposed a set of KPIs measuring improvements in RORENO core processes as wells as in RO IPO-Porto core processes and automatic integrations.

The KPIs proposed cover the following domains: time; workload and quality.

Table 8 - KPI proposal

To be model	Process	Time	Workload	Quality
Core processes - RORENO	New case	# Average time spent creating a ROR / diagnosis need # Average time spent creating a ROR / need for surgery # Average time spent creating a ROR / chemotherapy	# Number of ROR created / month # Number of ROR created / institution # Number of duplicate cases # Number of eliminate cases	# Number of records with conflicting information # Number of manual editions of individual records # Number of manual editions of cases (ROR)
	Follow-up	#Average time processing additional records to the worklist	# Number of cases added to the worklist (by PDS and RNU) # Number of cases added to the worklist by PDS / month # Number of cases added to the worklist by RNU / month	# Number of automatically integrated orders # Number of requests added to the worklist
Core processes – RO IPO-Porto	RO IPO-Porto – new case	# Average time spent creating a RO / diagnosis need # Average time spent creating a RO / need for surgery # Average time spent creating a RO /	# Number of ROs created / day # Number of ROs created / week # Number of ROs created / month	# Number of records with conflicting information # Number of records with conflicting information / reason (that triggers the conflict)

Table 8 - KPI proposal

To be model	Process	Time	Workload	Quality
		chemotherapy		# Average number of integrated systems / request (creation + change) # Number of manual editions of individual records
Automatic integrations	(indirectly all processes)	#Time delay	# Number of different sources / created new case # Number of cases diagnosed in one institution and treatment in another # Number of cases with death from cancer without database registration # Number of records of new cases manually / month # Number of records of new cases manually / year	# Completeness of registered data (number of hospitals filling out the entire form) # Provision of data services # Effort in the transition of data to the North Cancer Registry

7 Conclusions and recommendations

This final chapter presents research main findings, limitation of this work, recommendations, future work and final conclusions.

7.1 Main findings

This sub-section presented the main findings of this research. In order to better expose its main findings, the research questions were used to present the main findings.

The first question: "Which are the main characteristics of the implemented cancer registries?" was awswered by means of a literature review in study A. This main question was complemented by two sub-questions: "These characteristics/ functionalities fulfil with the cancer registries requirements and needs?" and "Which are the main cancer registries problems, concerns and solutions?".

In Europe cancer registries are evolving due to WHO requirements. Cancer registries born in the half of 20th century but only in the middle 90's had a greater technological promotion. Since then until now, despite having technological support, cancer registration processes require many humam resourses to perform manual data collection. The reason because this happens seems to be a lack of consensus between decision makers and of course a lack of financial support to invest in this process. Authors explained in their works that the progress in this field should imply a consensus between organization, and a strategy definition. This strategy should include the recommendation of proper data quality routines and reccomendation of data communication and codification standard.

More recently, legislation about data protection had an update with GDPR. GDPR seems to solve a problem reported by authors that was the lack of harmonization between Europe member-states. Although this regulation will bring new concerns in this field namely higher restrictions in data processing and patient involvement in this process at least the member-states will have an equal response. GDPR brings a new concern for individual organization since they have now the responsibility for their data protection and the sanctions for data breaches could lead to fines up to €20 million or 4% of global annual revenue for the preceding financial year (Commission, 2016), being collected by data watchdogs (in Portugal CNPD – *Comissão Nacional de Proteção de dados*) (CNPD, 2017).

So, in summary the strategy for a near future will necessarily imply a finantial investment in this field. The first priority should be the GDPR that will affect all the data types of information (paper based and elctronic based) and then automatize the cancer registration processes to comply in a more effeciency way.

Concerning the second third main question of this research: "Which are the main functionalities, users and stakeholders of RORENO?" that includes the sub-question: "Which are the main problems/ constraints felt by the users?" and third main question: "Which are the mandatory cancer registry characteristics/ functionalities that fulfil the users' needs?" were answered through study B.

In this second study, the idea was to make a study of a Portuguese system and then compare it to the first study (A).

RORENO analysis confirm some problems already found out by Study A. RORENO has few data integrations which implies an extra effort by RORENO workgroup to manually insert

data and validate data quality in order to answer WHO requirements. Users also complain about some minor problems in RORENO software namely usability issues.

In the future, a new system should include as main requirement data source integration, namely among LIS, RIS, ADT and with external institutions in North of Portugal.

IPO-Porto already started this path with ODISSEIA project. IPO-Porto strategy is in a near future to integrate data sources and automatize data processing. Their strategy implies to focus human resources work in studying the knowledge obtained this data and use this to improve their health service in terms of patient experience (treat well and fast) that will result in an improvement of cost-effectiveness.

7.2 Limitations of the work

This sub-section presents the limitations of this research.

Regarding the systematic review – Study A, results should be interpreted as a whole with the study limitations. The limitation to English language prevented the capture of other studies. Some studies that have an English abstract seems to be interesting however the full paper was written in other language for example German. This limitation should be addressed in a further research.

Concerning Study B, the RORENO workgroup involvement during the sessions allowed an in-depth study. RORENO workgroup had a special interest in this research, because as daily users they felt that their opinion was listened and the requirements elicitation was not only focused on decision makers and informatics. This last point leads as a research limitation. In this study, the stakeholders involved were limited to daily (final) users. To complete requirements elicitation other stakeholders needs to be involved namely external institutions representatives. During the dissertation period was impossible to get authorization to have a meeting with these stakeholders.

7.3 Recommendations and future work

Regarding study A – systematic review it is recommended to extend the study. Future work includes to study citations made by each of the eight papers that result from this research and also seek for papers where these eight papers were cited. Using this technique, it will be possible to follow authors that study this subject and try to get a more complete and updated answer of research questions. After gather updated results, this research will be published in a scientific journal of this field.

Concerning RORENO and ODISSEIA project, the next phase of this work implies to full describe a complete set of functional and non-functional requirements for a new cancer registry software that will serve the North of Portugal. One important requirement that this new system must comply is the integration with the new National Oncologic System (RON) that will be released in the future. It is recommended that this specification consider the results of this dissertation, bringing new features that solves current issues and anticipate new issues using recommended standards instead of non-standardized solutions.

After requirements specification, the following phases should include development, implementation and evaluation of the new system. The evaluation should include requirements verification and KPI assessment. To answer GDPR requirements, new system provider and IPO-Porto should work on a PIA explaining security and privacy features ensuring GDPR compliance, avoiding issues with an audit.

7.4 Conclusions

In Europe, the progress that has been made deserves to continue, and the current momentum of attention and support for quality cancer registration should be maintained. The reduction of disparities in the quality and function of cancer registries is a cornerstone, as are the efforts to harmonize, standardize, and bring together in a comparable and understandable way the wealth of cancer data across the continent.

Epidemiological research depends on the balance between preserving patients' integrity and anonymity while also enabling important research to improve people's health and the quality of care, until 2018 the healthcare institutions should be prepared for comply the changes in the data protection (GDPR), so this system should comply with these new privacy and security requirements.

The results of this research thesis constitute the starting point to define a new cancer registry system for the North of Portugal. The findings obtained in this research should be considered in this new model. The actual system RORENO has some limitations as identified in study B, however these limitations are not only felt by this oncologic group but also shared by other European groups as finding in study A.

It is time to learn with these issues, define and implement a new system to fulfill the future needs. As it will be an important financial investment for institutions, in or case IPO-Porto it is important to bring new uses of data (not transgressing GDPR recommendations). It will be important to transform data into knowledge in order to improve processes inside the organization.

The secondary use of data principle has the purpose to enhance individuals' health care experiences, expand knowledge about diseases and treatments, strengthen understanding of health care systems' effectiveness and efficiency, support public health and security goals, and aid businesses in meeting customers' needs (Safran *et al.*, 2007). A good practice of secondary use of data is its usage for business intelligence however it is important to assure the quality of data. If the data quality is ensured it can be easily use for other purposes in order to provide a better service to the patients and reduce its costs for example monitoring the waiting times, eliminate the wasting processes, monitoring drugs adverse reactions, use the knowledge (based in evidence) to better negotiate with suppliers etc.

References

A. Oakley (1998) 'Gender methodology and people's ways of knowing: Some problems with feminism and the paradigm debate in social science.', *Sociology*.

ACSS (2010) *Registo Nacional de Utentes (RNU)*, *Diretório de Informação em Saúde*. Available at: http://dis.dgs.pt/2010/07/08/registo-nacional-de-utentes-rnu/.

Andersen, M. R. and Storm, H. H. (2015) 'Cancer registration , public health and the reform of the European data protection framework : Abandoning or improving European public health research ?', *European Journal of Cancer*. Elsevier Ltd, 51(9), pp. 1028–1038. doi: 10.1016/j.ejca.2013.09.005.

Bianconi, F. *et al*. (2012) 'Information Technology as Tools for Cancer Registry and Regional Cancer Network Integration', *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBER NETICS—PART A: SYSTEMS AND HUMANS*, 42(6), pp. 1410–1424.

Bjugn, R., Casati, B. and Norstein, J. (2008) 'Structured electronic template for histopathology reports on colorectal carcinomas: a joint project by the Cancer Registry of Norway and the Norwegian Society for Pathology', *Human Pathology*, 39(3), pp. 359–367. doi: http://dx.doi.org/10.1016/j.humpath.2007.06.019.

Bray, F. *et al*. (2014) *Cancer Incidence in Five Continents Vol . X*.

Bray F, P. D. (2009) 'Evaluation of data quality in the cancer registry: principles and methods. Part I: comparability, validity and timeliness.', *Eur J Cancer*. doi: 10.1016/j.ejca.2008.11.032.

Chalmers, I. (1993) 'The Cochrane collaboration: preparing, maintaining, and disseminating systematic reviews of the effects of health care.', *Annals of the New York Academy of Sciences*.

CNPD (2017) *Comissão Nacional de Proteção de Dados*.

Coebergh, J. (2015) 'EUROCOURSE lessons learned from and for population-based cancer registries in Europe and their programme owners: Improving performance by research programming for public health and clinical evaluation.', *Eur J Cancer*. doi: 10.1016/j.ejca.2015.02.018.

Coebergh, W. (2012) 'In search of Program Owners of the cancer registry, EUROCOURSE WP1, Deliverable 1.6', *ENCR*.

Colquitt, S., Clements, K. and Hart-hester, S. (2012) 'The Impact of Electronic Health Record Usage on Cancer Registry Systems in Alabama', *Perspect Health Inf Management*, 9.

Commission, E. (2016) 'General Data Protection Regulation', *European Comission*.

Conseil, P. européen et du (1995) 'Directive 95/46/CE', pp. 0031–0050.

Contiero, P. *et al*. (2008) 'Comparison with manual registration reveals satisfactory completeness and efficiency of a computerized cancer registration system', *Journal of Biomedical Informatics*, 41(1), pp. 24–32. doi: https://doi.org/10.1016/j.jbi.2007.03.003.

Elsevier (2017) *Scopus*. Available at: https://www.scopus.com/freelookup/form/author.uri.

ENCR (2016) *European Network of Cancer Registries*. Available at: http://www.encr.eu/.

EPAAC (2014) 'Proposal of indicators for the European Cancer Information System WP9 Deliverable 2 In', *ECIS*.

Ferlay, J. *et al*. (2015) 'Cancer incidence and mortality worldwide : Sources , methods and major patterns in GLOBOCAN 2012', *International Journal of Cancer*, 386(15). doi: 10.1002/ijc.29210.

Forsea, A. (2016) 'Cancer registries in Europe — going forward is the only option', *Open access cancer Journal*, pp. 1–15. doi: 10.3332/ecancer.2016.641.

Forsea, A.-M. (2016) 'Cancer registries in Europe—going forward is the only option', *ecancermedicalscience*. Cancer Intelligence, 10, p. 641. doi: 10.3332/ecancer.2016.641.

Gliklich RE, Dreyer NA, Leavy MB (2014) 'Data Sources for Registries', in *Registries for Evaluating Patient Outcomes: A User's Guide [Internet]*. 3rd editio. Rockville (MD): Agency for Healthcare Research and Quality (US).

Gray, D. (2009) 'Doing Research in the Real World.', *California: Sage Publications*, 2nd ed.

Hakulinen, T. *et al*. (2011) 'Harmonization may be counterproductive - At least for parts of Europe where public health research operates effectively', *European Journal of Public Health*, 21(6), pp. 686–687. doi: 10.1093/eurpub/ckr149.

Houser, S. H. *et al*. (2012) 'The Impact of Electronic Health Record Usage on Cancer Registry Systems in Alabama', *Perspectives in Health Information Management / AHIMA, American Health Information Management Association*. American Health Information Management Association, 9(Spring), p. 1f.

IACR (2017a) *CanReg5 software*. Available at: http://www.iacr.com.fr/index.php?option=com_content&view=category&layout=blog&id=68&Itemid=445.

IACR (2017b) *IACR - International Association of Cancer Registries*. Available at: http://www.iacr.com.fr/index.php?option=com_content&view=article&id=72:iarccrgtools&catid=68&Itemid=445.

IACR/WHO (2017) 'Internacional Classification of Diseases for Oncology'.

IEEE (2017) *IEEE Xplore digital database*. Available at: http://ieeexplore.ieee.org/Xplore/home.jsp.

International, W. cancer research fund (2017) *Data for cancer frequency by country*, *World cancer research fund international*. Available at: http://www.wcrf.org/int/cancer-facts-figures/data-cancer-frequency-country.

IPO-Coimbra (2017) *ROR-Centro*. Available at: http://www.rorcentro.com.pt/.

IPO-Lisboa (2017) *ROR-Sul*. Available at: http://www.ror-sul.org.pt/pt.

IPO-Porto (2017a) *IPO-Porto*. Available at: http://www.ipoporto.pt/en/.

IPO-Porto (2017b) *RORENO - Registo Oncológico Regional do Norte*. Available at: http://www.roreno.com.pt/.

Jamshed, S. (2014) 'Qualitative research method-interviewing and observation', *Journal of Basic and Clinical Pharmacy*. India: Medknow Publications & Media Pvt Ltd, 5(4), pp. 87–88. doi: 10.4103/0976-0105.141942.

Kawulich, B. (2005) 'Participant Observation as a Data Collection Method', *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 6(2).

Knowledge, I. W. of (2017) *Web of Science*. Available at: http://apps.webofknowledge.com/WOS_GeneralSearch_input.do?product=WOS&search_mode=GeneralSearch&SID=U2YhonnWze7oNVVQL51&preferencesSaved=.

Leal, J. *et al*. (2016) 'Economic Burden of Bladder Cancer Across the European Union', *European Urology*, 69(3), pp. 438–447. doi: https://doi.org/10.1016/j.eururo.2015.10.024.

Lenzerini, M. (2002) 'Data Integration: A Theoretical Perspective', in *Proceedings of the Twenty-first ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*. New York, NY, USA: ACM (PODS '02), pp. 233–246. doi: 10.1145/543613.543644.

Luborsky, M. R. and Rubinstein, R. L. (1995) 'Sampling in Qualitative Research: Rationale, Issues, and Methods', *Research on aging*, 17(1), pp. 89–113. doi: 10.1177/0164027595171005.

McLeod, S. (2015) *Observation Methods*, *Simply Psychology*. Available at: https://www.simplypsychology.org/observation.html.

Medline (2017) *Pubmed*. Available at: https://www.ncbi.nlm.nih.gov/pubmed/.

Modelling, A. (2014) *Personas: An agile introduction*.

Moher D, Liberati A, Tetzlaff J, A. D. (2009) 'Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement.', *The PRISMA Group*. doi: https://doi.org/10.1371/journal.pmed.1000097.

Moher D, Tetzlaff J, Tricco AC, Sampson M, A. D. (2007) 'Epidemiology and Reporting Characteristics of Systematic Reviews.', *PLoS Med*. doi: https://doi.org/10.1371/journal.pmed.0040078.

Neuman, W. L. (2014) *Social Research Methods: Qualitative and Quantitative Approaches*. Pearson Ed.

Norwegian Directorate of Health (2012) *Direktoratet for e-helse*.

O.M. Jensen, D.M. Parkin, R. MacLennan, C. S. M. and R. G. S. (1991a) 'Cancer Registration : Principles and Methods', *IARC Scientific Publication*, 95.

O.M. Jensen, D.M. Parkin, R. MacLennan, C. S. M. and R. G. S. (1991b) *Cancer Registration: Principles and Methods*. 95th edn. IARC Scientific Publication.

República, D. da (1988) 'Portaria 35/88'.

RORENO (2015) 'Registo Oncológico Regional do Norte 2010', *IPO-Porto*.

RORENO (2017) *ROERNO Publication 2011*.

Rossi, S. *et al*. (2015) 'The EUROCARE-5 study on cancer survival in Europe 1999–2007: Database, quality checks and statistical analysis methods', *European Journal of Cancer*, 51(15), pp. 2104–2119. doi: http://dx.doi.org/10.1016/j.ejca.2015.08.001.

Safran, C. *et al*. (2007) 'Toward a National Framework for the Secondary Use of Health Data: An American Medical Informatics Association White Paper', *Journal of the American Medical Informatics Association : JAMIA*. American Medical Informatics Association, 14(1),

pp. 1–9. doi: 10.1197/jamia.M2273.

Santos-Pereira, C. (2017) 'A qualitative research evaluation of a Portuguese computerized cancer registry', in *2017 12th Iberian Conference on Information Systems and Technologies (CISTI)*, pp. 1–6. doi: 10.23919/CISTI.2017.7975827.

Siesling, S. *et al*. (2015) 'Uses of cancer registries for public health and clinical research in Europe : Results of the European Network of Cancer Registries survey among 161 population-based cancer registries during 2010 – 2012', *European Journal of Cancer*. Elsevier Ltd, 51(9), pp. 1039–1049. doi: 10.1016/j.ejca.2014.07.016.

Silva, I. dos S. (1999) 'The Role of Cancer Registries Cancer Epidemiology: Principles and Methods', *International Agency for Research on Cancer*.

SNOMED (2017) *SNOMED International*.

SNS (2017) *Registo Oncológico Nacional*, *SNS - Serviço Nacional de Saúde*. Available at: https://www.sns.gov.pt/noticias/2016/12/19/registo-oncologico-em-2017/.

SPMS (2013) 'PDS – Plataforma de Dados da Saúde'.

Tafazzoli, A. G. *et al*. (2002) 'Integrated decision support in a hospital cancer registry', *Artificial Intelligence in Medicine*, 24(3), pp. 243–255. doi: http://dx.doi.org/10.1016/S0933-3657(01)00107-5.

Tognazzo, S. (2006) 'Quality control of automatically defined cancer cases by the automated registration system of the Venetian Tumour Registry. Quality control of cancer cases automatically registered', *The European Journal of Public Health*. doi: 10.1093/eurpub/cki035.

UK, C. research (2014) *Cancer statistics for the UK*, *Cancer research UK,*. Available at: http://www.cancerresearchuk.org/health-professional/cancer-statistics.

White, G. R. T. and Cicmil, S. (2016) 'Knowledge acquisition through process mapping: Factors affecting the performance of work-based activity', *International Journal of Productivity and Performance Management*. Emerald, 65(3), pp. 302–323. doi: 10.1108/IJPPM-01-2014-0007.

WHO (2013) 'ICD-O', in *International Classification of Diseases for Oncology (ICD-O)*.

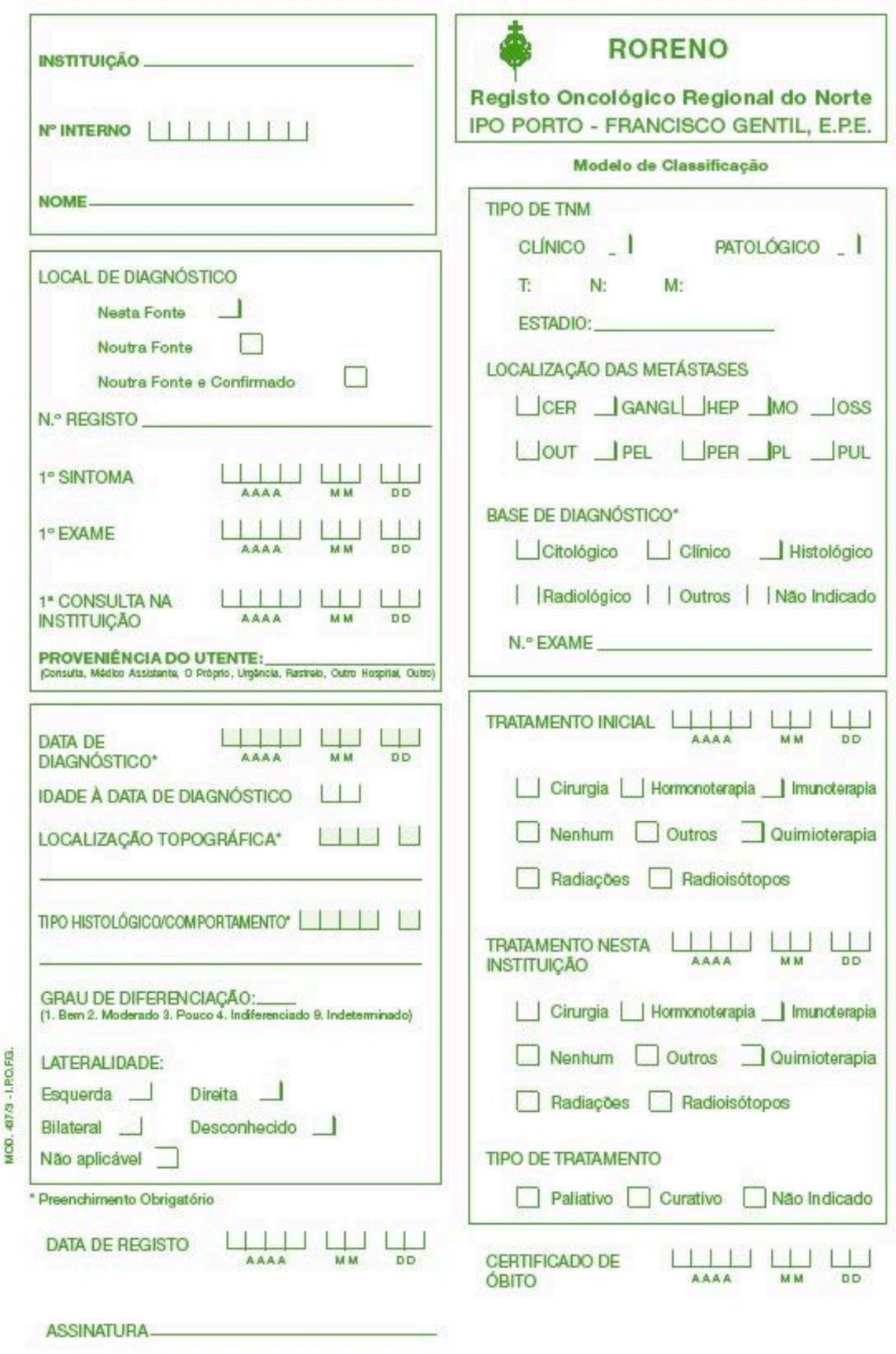
WHO (2017) *International Classification of Diseases (ICD) Information Sheet*. Available at: http://www.who.int/classifications/icd/factsheet/en/.

WHO/IARC (2014) 'International Agency for Research on Cancer', *WHO*.

Y Bhurgri, A Bhurgri, S. H. H. (2002) 'Comparability and Quality Control in Cancer Registration; Karachi (data monitoring 1995-2001).', *JPMA. The Journal of the Pakistan Medical Association*, 52 7, pp. 301–7.

Zanetti, R. *et al*. (2015) 'Completeness and timeliness: Cancer registries could/should improve their performance', *European Journal of Cancer*, 51(9), pp. 1091–1098. doi: http://dx.doi.org/10.1016/j.ejca.2013.11.040.

APPENDIX A: Form template for cancer registration (available in (RORENO, 2015))



INSTITUIÇÃO _____

Nº INTERNO |_|_|_|_|_|_|_|

NOME _____

RORENO

Registo Oncológico Regional do Norte

IPO PORTO - FRANCISCO GENTIL, E.P.E.

Modelo de Classificação

LOCAL DE DIAGNÓSTICO

Nesta Fonte ☐

Noutra Fonte ☐

Noutra Fonte e Confirmado ☐

N.º REGISTO _____

1º SINTOMA |_|_|_|_| AAAA |_|_| MM |_|_| DD

1º EXAME |_|_|_|_| AAAA |_|_| MM |_|_| DD

1ª CONSULTA NA INSTITUIÇÃO |_|_|_|_| AAAA |_|_| MM |_|_| DD

PROVENIÊNCIA DO UTENTE: _____

(Consulta, Médico Assistente, O Próprio, Urgência, Rastreio, Outro Hospital, Outro)

TIPO DE TNM

CLÍNICO _ | PATOLÓGICO _ |

T: N: M:

ESTADIO: _____

LOCALIZAÇÃO DAS METÁSTASES

☐CER ☐GANGL ☐HEP ☐MO ☐OSS

☐OUT ☐PEL ☐PER ☐PL ☐PUL

BASE DE DIAGNÓSTICO*

☐Citológico ☐Clínico ☐Histológico

☐Radiológico ☐Outros ☐Não Indicado

N.º EXAME _____

DATA DE DIAGNÓSTICO* |_|_|_|_| AAAA |_|_| MM |_|_| DD

IDADE À DATA DE DIAGNÓSTICO |_|_|

LOCALIZAÇÃO TOPOGRÁFICA* |_|_|_| |_|

_____

TIPO HISTOLÓGICO/COMPORTAMENTO* |_|_|_|_| |_|

_____

GRAU DE DIFERENCIAÇÃO: ____

(1. Bem 2. Moderado 3. Pouco 4. Indiferenciado 9. Indeterminado)

LATERALIDADE:

Esquerda ☐ Direita ☐

Bilateral ☐ Desconhecido ☐

Não aplicável ☐

MOD. 497/3 - I.P.O.P.G.

* Preenchimento Obrigatório

DATA DE REGISTO |_|_|_|_| AAAA |_|_| MM |_|_| DD

ASSINATURA _____

TRATAMENTO INICIAL |_|_|_|_| AAAA |_|_| MM |_|_| DD

☐ Cirurgia ☐ Hormonoterapia ☐ Imunoterapia

☐ Nenhum ☐ Outros ☐ Quimioterapia

☐ Radiações ☐ Radioisótopos

TRATAMENTO NESTA INSTITUIÇÃO |_|_|_|_| AAAA |_|_| MM |_|_| DD

☐ Cirurgia ☐ Hormonoterapia ☐ Imunoterapia

☐ Nenhum ☐ Outros ☐ Quimioterapia

☐ Radiações ☐ Radioisótopos

TIPO DE TRATAMENTO

☐ Paliativo ☐ Curativo ☐ Não Indicado

CERTIFICADO DE ÓBITO |_|_|_|_| AAAA |_|_| MM |_|_| DD

APPENDIX B: RORENO Core processes

New case

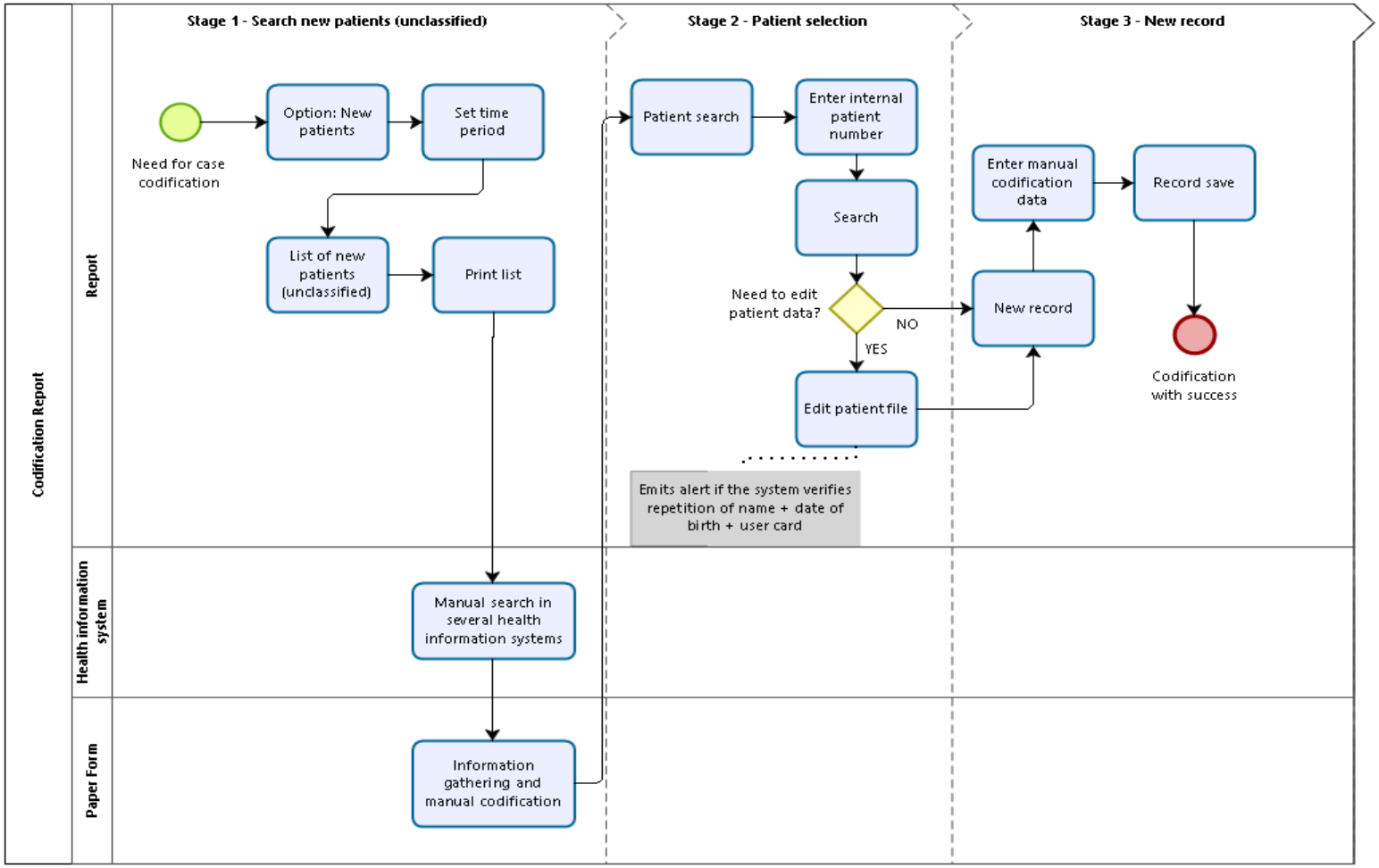


Figure 10 – New case insertion process

Update follow-up

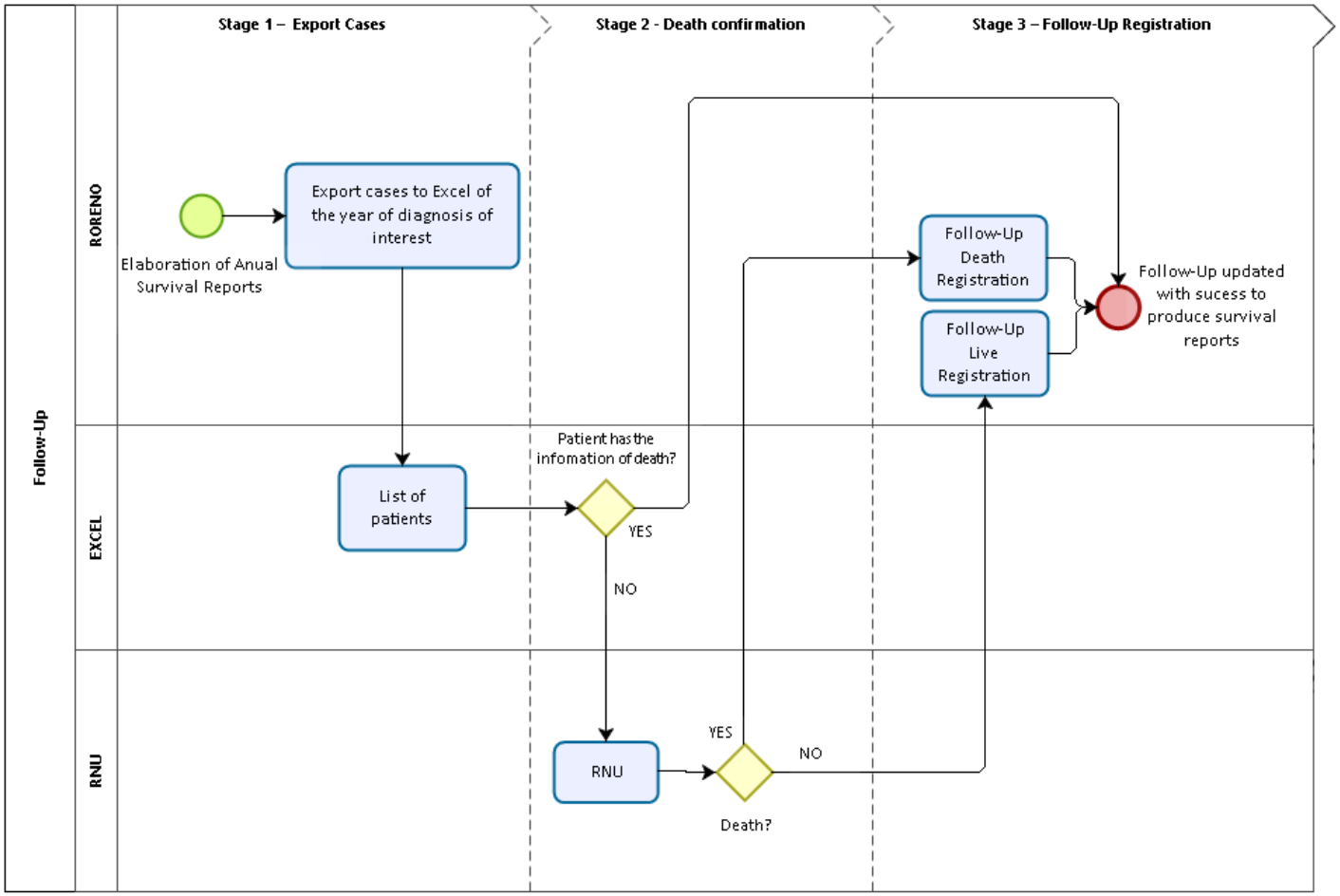


Figure 11 - Follow-up process

Data migration

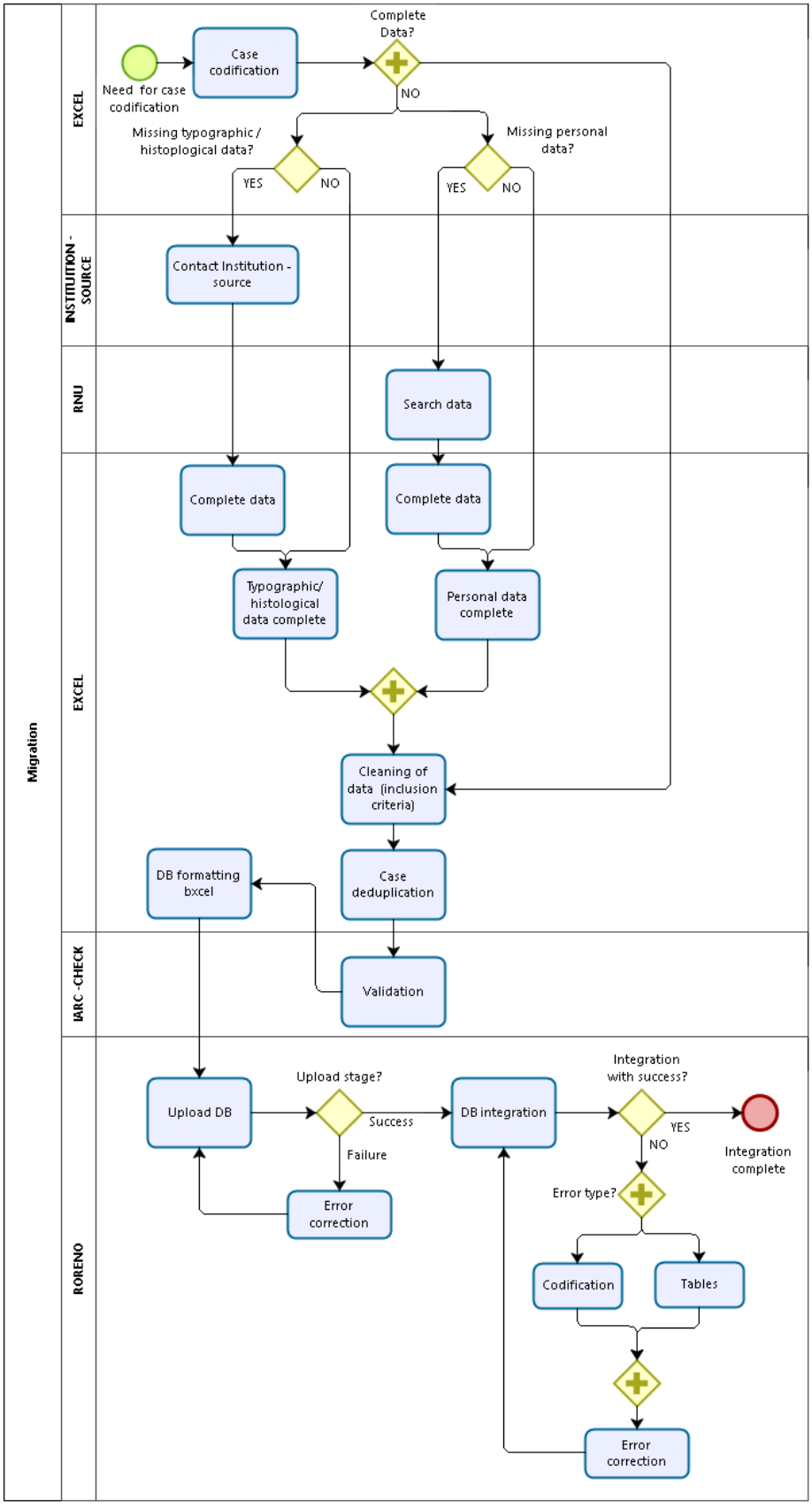


Figure 12 - Data migration process

Data deduplication and validation

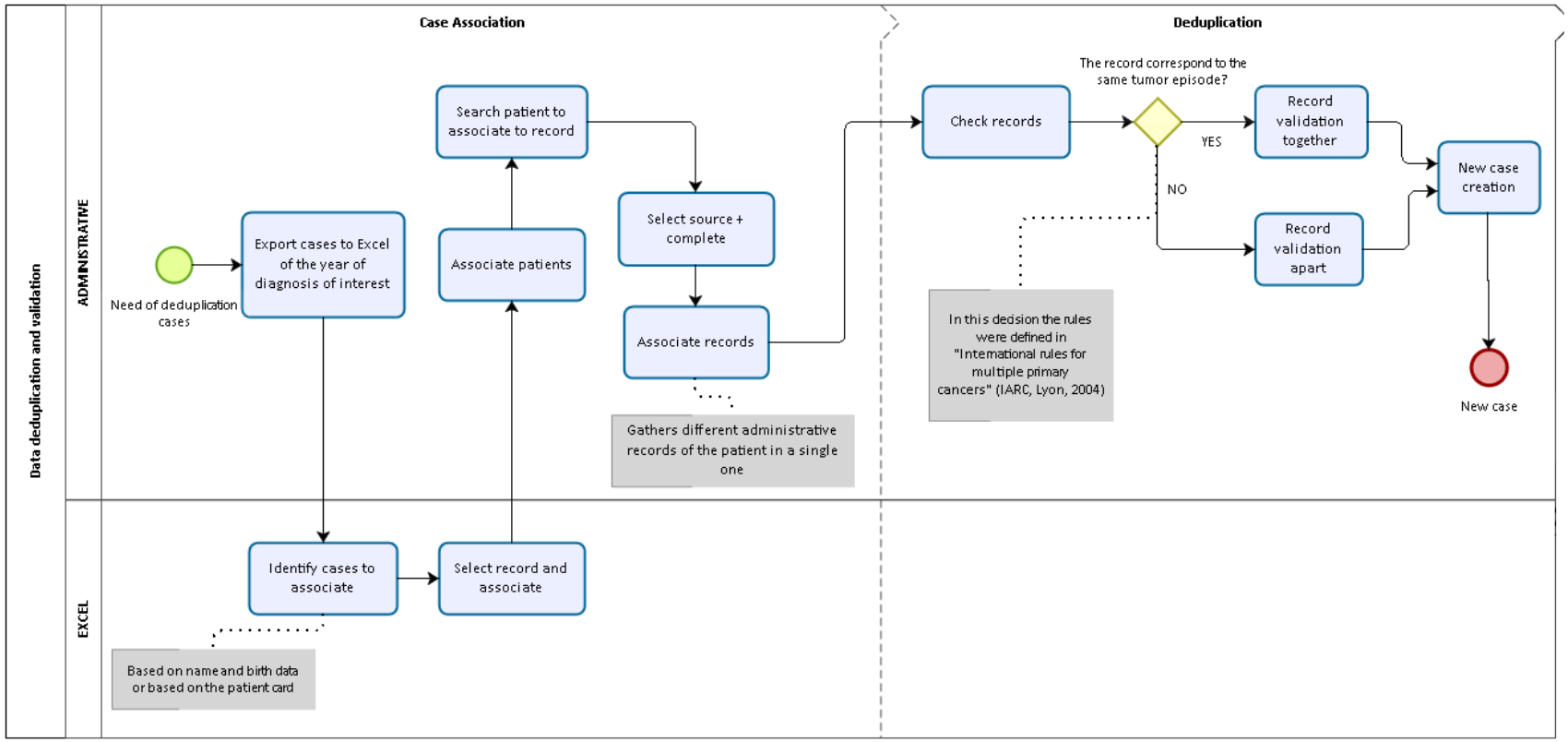


Figure 13 - Data deduplication and validation process

APPENDIX C: Service blueprint

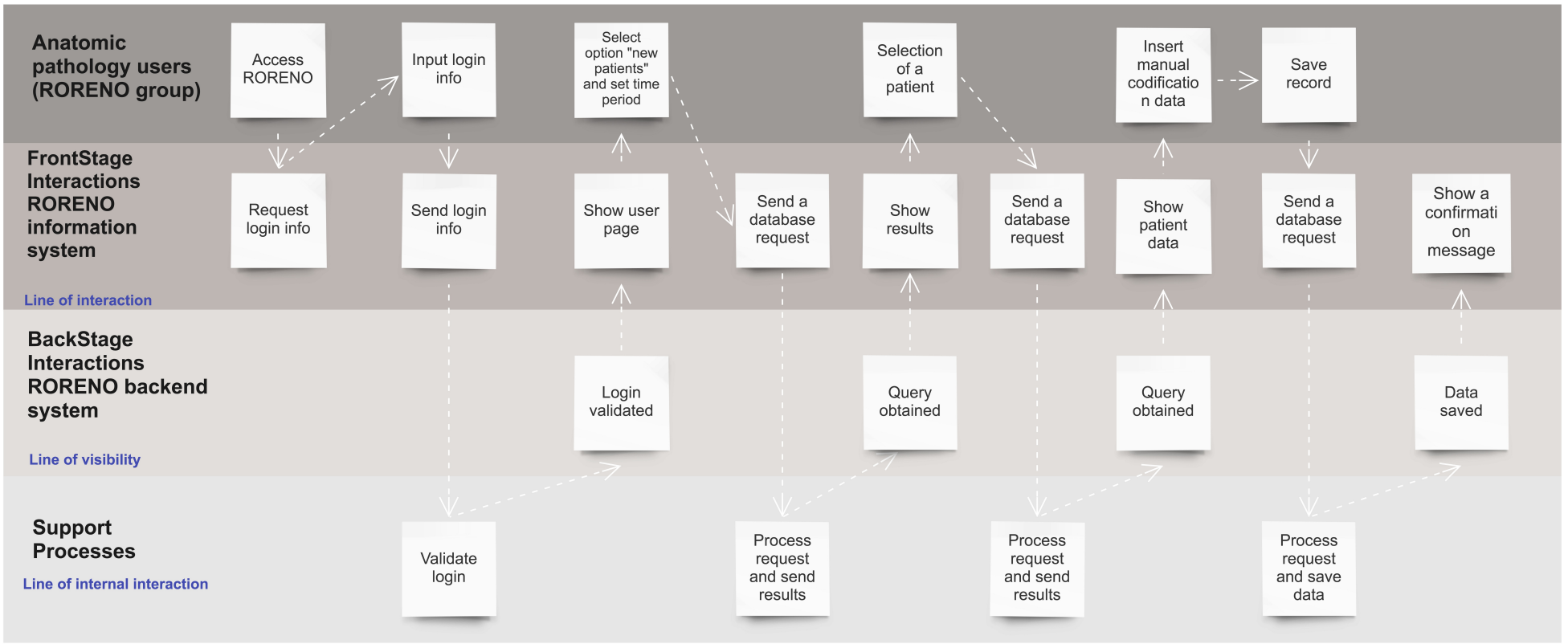


Figure 14 - Service Blueprint: Codification of a new case in RORENO

APPENDIX D: List of institutions that belongs to RORENO network

The following table (Table 9) shows the list of healthcare institutions that belongs to RORENO network. This list includes not only public institutions but also private institutions and the volume of its records sent to RORENO. This information is available in (IPO-Porto, 2017b).

Table 9 - List of institutions that delivered records to RORENO in 2011

INSTITUIÇÕES	RESPONSÁVEIS	Nº REGISTOS
Instituto Português de Oncologia do Porto, E.P.E.	Doutora Ana Filipa Gonçalves	8181
Hospital Geral de São João, Porto	Dra. Isabel Carvalho	2985
Lab. Anatomia Patológica do Dr. J. A. Macedo Dias	Dr. J. A. Macedo Dias	2431
Lab. Anatomia Patológica do Dr. Eduardo S. Ferreira	Dr. Eduardo S. Ferreira	2208
C. H. Porto - Hospital Geral de Santo António	Dr. José Ramón Vizcaíno	1989
Registo Oncológico Regional do Centro	Dr. Manuel António Silva	1785
C. H. Vila Nova de Gaia/Espinho	Dr. Agostinho Sanches	1502
Hospital São Marcos, Braga	Dr. Fernando Pardal	1458
C. H. de Vila Real/Peso da Régua	Dra. Maria José del Rio	1304
C. H. Tâmega e Sousa - Hospital Padre Américo	Dra. Filipa Carneiro	1209
Lab. Anatomia Patológica da Dra. Isabel Macedo Pinto	Dra. Isabel Macedo Pinto	1158
C. H. Alto Minho, Viana do Castelo/Ponte de Lima	Dr. Manuel Veiga	1143
Hospital Pedro Hispano, Matosinhos	Dra. Mrinalini Honavar	1102
Lab. Anatomia Patológica - HICISLAB	Dr. António Paiva Correia	864
C. H. Alto Ave - U. H. Guimarães	Dra. Camila Coutinho	848
Lab. Anatomia Patológica do Dr. Vicente Gonçalves	Doutor Vicente Gonçalves	840
C. H. Entre Douro e Vouga - Hospital São Sebastião	Prof. Doutor António Araújo	827
C. H. Médio Ave - U. H. Famalicão	Dra. Marta Novais Silva	805
Lab. Anatomia Patológica do Dr. Fernando Pardal	Dr. Fernando Pardal	594
Lab. Anatomia Patológica do Dr. Caspurro	Dr. Silva Caspurro	593
Hospital da Luz Póvoa de Varzim	Dr. Luís Grangeia	566
C. H. Trás-os-Montes e Alto Douro - U. H. Chaves	Dra. Maria José del Rio	386
C. H. Nordeste - U. H. Bragança	Dra. Lília Meireles	371
C. H. Póvoa de Varzim/Vila do Conde	Dra. Adelaide Graça	238
Lab. Anatomia Patológica do Prof. Carlos Lopes	Prof. Doutor Carlos Lopes	238
Hospital Santa Maria Maior, Barcelos	Dra. Marta Gomes	231
Hospital da Luz Arrábida	Dr. Leal da Silva	229
Instituto CUF	Dr. Carlos Sottomayor	206
IPATIMUP	Prof. Doutor Fernando Schmitt	205
C. H. Nordeste - U. H. Mirandela	Dra. Lília Meireles	197
Centro de Dermatologia EPIDERMIS	Prof. Doutor Osvaldo Correia	127
Registo Oncológico Regional do Sul	Dra. Ana Miranda	108
C. H. Trás-os-Montes e Alto Douro - U. H. Lamego	Dr. Alexandre Hoffmann Castela	95
C. H. Nordeste - U. H. Macedo Cavaleiros	Dra. Lília Meireles	51
C. H. Médio Ave - U. H. Santo Tirso	Dra. Marta Novais Silva	8

Table 10 - Population resident in North of Portugal in 2011 (estimative data): source (RORENO, 2017)

GRUPOS ETÁRIOS	SEXO MASCULINO	SEXO FEMININO	TOTAL
< 1	16190	15686	31876
1 - 4	66646	64382	131028
5 - 9	94395	90772	185167
10 - 14	106420	101052	207472
15 - 19	108272	103493	211765
20 - 24	108142	106717	214859
25 - 29	115075	116666	231741
30 - 34	129164	135357	264521
35 - 39	142691	151079	293770
40 - 44	138642	147587	286229
45 - 49	141152	151335	292487
50 - 54	129291	138017	267308
55 - 59	114239	125805	240044
60 - 64	99709	111412	211121
65 - 69	78318	95254	173572
70 - 74	67072	86650	153722
75 - 79	55269	78699	133968
80 - 84	34210	57352	91562
≥85	21453	46748	68201
TOTAL	1766350	1924063	3690413

APPENDIX E: Article published

A qualitative research evaluation of a Portuguese computerized cancer registry

Cátia Santos-Pereira[1,2,3], Ricardo Cruz-Correia[1,4,5], António Carvalho Brito[2,3], Alexandre B. Augusto[1], Manuel E. Correia[1,2], Maria José Bento[6], Luís Antunes[6]
HealthySystems - HLTSYS, Lda[1]
Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência - INESC TEC[2]
Faculdade de Engenharia da Universidade do Porto – FEUP[3]
Center for Research in Health Technologies and Information Systems – CINTESIS[4]
Faculdade de Medicina da Universidade do Porto - FMUP[5]
Instituto Português de Oncologia Francisco Gentil do Porto – IPO-Porto[6]
Porto, Portugal
catiapereira@hltsys.pt; rcorreia@med.up.pt; acbrito@fe.up.pt; aaugusto@hltsys.pt; mcc@hltsys.pt; mjbento@ipoporto.min-saude.pt; luisantunes@ipoporto.min-saude.pt

Abstract — **A cancer registry is a standardized tool to produce population-based data on cancer incidence and survival. Cancer registries can retrieve and store information on all cancer cases occurring in a defined population. The main sources of data on cancer cases usually include: treatment and diagnostic facilities (oncology centres or hospital departments, pathology laboratories, or imaging facilities etc.) and the official territorial death registry. The aim of this paper is to evaluate the north regional cancer registry (RORENO) of Portugal using a qualitative research. We want to characterize: the main functionalities and core processes, team involved, different healthcare institutions in the regional network and an identification of issues and potential improvements. RORENO links data of thirteen-two healthcare institutions and is responsible for the production of cancer incidence and survival report for this region. In our semi-structure interviews and observation of RORENO we identified a serious problem due to a lack of an automatic integration of data from the different sources. Most of the data are inserted manually in the system and this implies an extra effort from the RORENO team. At this moment RORENO team are still collecting data from 2011. In a near future it is crucial to automatize the integration of data linking the different healthcare institutions in the region. However, it is important to think which functionalities this system should give to the institutions in the network to maximize the engagement with the project. More than a database this should be a source of knowledge available to all the collaborative oncologic network.**

Keywords – cancer registries; information system; qualitative research; BPMN, qualitative research, semi-structure interview, observation.

I. INTRODUCTION

Medical registries are described as a systematic collection of a clearly defined set of health and demographic data for patients with specific health characteristics, held in a central database for a predefined purpose [1].

A cancer registry is a standardized tool to produce population-based data on cancer incidence and survival [2]. Cancer registries can retrieve and store information on all cancer cases occurring in a defined population. Its data can be used in a wide variety cancer control ranging from etiological research, through primary and secondary prevention to health-care planning and patient care, so benefiting both the individual and society. Cancer registries are evolving to provide a high level of clinical details, and to improve their capability to provide an evaluation of health interventions in oncology. Diagnosis, stage, and treatment information is registered with increasing frequency and higher level of clinical detail. Thus, the cancer registry is evolving as a tool to support planning and evaluation of cancer control strategies. However, the traditional cancer registry is retrospective or historical in its nature since it is presently limited/bound to investigate variables routinely determined in health archives [1].

The storage of large quantity of cancer registries is possible. However, the real problem in expanding the cancer registry scope is the difficulty to access an increasing number of variables from different sources.

The Portuguese health Ministry are working on an unique cancer registry for Portugal – RON - *Registo Oncológico Nacional* [3]. This new system will integrate the data from the actual three regional cancer registries (RORENO, ROR-Centro and ROR-Sul), following the recommendations of the WHO (World Health Organization) [4]. So, in this moment of change we believe it is important to evaluate the actual systems understanding its needs. The outcome of this work could be an important source of knowledge for the definition of the next systems with the same scope.

The aim of this paper is to evaluate the computerized cancer registry implemented *in "Instituto Português de Oncologia do Porto"* using a qualitative research. The main goal is to characterize the environment of the system, such as the main functionalities and core processes, team involved and different

978-989-98434-7-9/17/$31.00

healthcare institutions in the regional network. Then identify the main problems and difficulties that arose in the last year.

II. BACKGROUND

A. Cancer registry content

The main source of data on cancer cases usually include: treatment, diagnostic facilities (oncology centres or hospital departments, pathology laboratories, or imaging facilities etc.) and the official territorial death registry. Additional data sources like ambulatory, private clinics, elderly care homes and general practitioners' networks increase the completeness of data, but also the logistics and expenses [5].

Hospital-based cancer registries are more numerous and widespread than population-based cancer registries. The primary purpose of these registries is to contribute to patient care by providing accessible information on the patients with cancer, the treatment they received and its results. The data may also be used for clinical research and, to a certain extent, for epidemiological purposes. One of the main advantages of hospital registries are the availability and the completeness of medical records. The data collected by a hospital registry tend to be more extensive than those collected by a population registry [5].

The top benefits of a computerized cancer registry include: 1) more complete treatment information, 2) less time for case finding and data entry, 3) more available time for data retrieval and analysis, 4) improved completeness, accuracy, and timeliness, 5) better patient tracking for follow-up, and 6) improved workflow efficiency.

The main challenges of the utilization of a computerized system within a cancer registry identified by respondents are 1) lack of adequate funding, 2) lack of medical staff to support the system, 3) changing data standards, 4) lack of full-time commitments, and 5) lack of a standardized data exchange [5].

The minimum set of data recommended to be collected by all cancer registries was formulated by the International Agency for Research on Cancer (IARC) [6] and by the European Network of Cancer Registries (ENCR) and are: cancer patient's personal identification; tumour site; tumour histology, classified by the ICD-O; tumour stage; tumour diagnosis, related to the ICD9; tumour therapy; further treatment; follow up; individual history; family history; and death, including autopsy results, if any. These items correspond to sensitive personal and medical information [7].

B. European initiatives

Cancer registries have been expanding in Europe since the early 1900's. Over the last three decades cancer registration has become an important element of the EU's strategy against cancer, promoted within the framework of the European Action against Cancer Programme (1985–2008), the European Partnership for Action Against Cancer (EPAAC) (2009–2014) [8]. The last data shows that nearly 200 population-based cancer registries (PCRs) are active in Europe and they are members of European Network of Cancer Registries (ENCR) [9].

The quality registration coverage of population by national cancer registries are available in 22 European countries. High quality registration of 10–50% of the population are available in France, Italy, Switzerland, Spain, Germany, and Serbia [10]. High quality registration of <10% of the population are available in Poland and Portugal [10]. EU member states Romania, Greece, and Hungary had as per 2012 only regional or partial data although legislation is in place to allow national cancer registration [10].

C. Europeran Legislation.

Cancer case reporting is mandatory by law in most of the European countries with high quality registers that included in the cancer incidence in five continents series [11].

Personal data protection legislation has a major impact on electronic cancer registry. In a pan-European survey within the EUROCOURSE project, 20–35% of responding cancer registries reported legal-related barriers to cancer registration across most of Europe, while in the South-West region these barriers amounted to up to 60% [12]. Particularly concerning are the nationally variable barriers in the linkage of cancer registries with other health-related databases like mass screening programmes, biobanks, vital status, and causes of death databases [12][13].

In 2014 the European Commission proposed to replace the Directive 95/46/CE [14] by the General Data Protection Regulation [15]. The overall intention of this reform is to protect personal data and to facilitate a free flow of data within the European Union (EU). This initiative will also help to overcome problems alluded by the research community concerning about data sharing across borders for research purposes.

The outcome of the data protection reform is crucial to all epidemiological activities and clinical quality control in the EU. In contrast to a directive, a regulation is binding by itself and does not need implementing legislation by the Member States. It implies a harmonization of data protection measures across the EU, including use of data for public health purposes such as prevention and evaluation of screening programmes. However, on the one hand harmonisation, may facilitate valuable data sharing for research purposes, but on the other hand, excessive regulation can easily disable even simple monitoring of cancer, with disastrous consequences for public health information.

D. Portuguese computerized cancer registy in study

The RORENO [16] workgroup begins his activity in 1988 answering the governmental rule 35/88 [17]. This group is held in "*Instituto Português de Oncologia do Porto Francisco Gentil, EPE* – IPO-Porto", a public reference oncologic Hospital [18]. Since 1988 has been reporting cancer cases sent by different healthcare institutions of the north of Portugal. The geographical area comprehends the following districts: Porto, Braga, Viana do Castelo, Vila Real, Bragança, some cities in Aveiro district (Albergaria-a-Velha, Arouca, Castelo de Paiva, Espinho, Estarreja, Murtosa, Oliveira de Azeméis, Ovar, S. João da Madeira, Santa Maria da Feira and Vale de Cambra), in Viseu district (Cinfães, S. João da Pesqueira

Armamar, Lamego, Moimenta da Beira, Penedono, Resende, Sernancelhe, Tabuaço and Tarouca) and in Guarda district (Vila Nova de Foz Côa) [16].
In 2005, this workgroup acquires a software named RORENO that facilitates the communication between this network. This system allows to accomplish the core competencies of the group:

- Collection of the most complete and up-to-date data on all new cases of tumours among residents of the Northern Region of Portugal.
- Production, analysis and interpretation of impact indicators of oncological disease and respective publication (national and international level).
- Provide information to all health professionals, researchers, policy makers, health care organizations in the effort to contribute to the prevention and control of cancer diseases [16].

The RORENO workgroup is expert in epidemiology and is constituted by a multi-disciplinary team of medical doctors, pathological anatomy technicians, statistical technicians and informatics professionals. The RORENO team is composed by thirteen professionals working in cancer registry of the North of Portugal.

III. Methods

In this research, it was applied two kinds of qualitative techniques to acquire information about the health information system in study: semi-structured interview and observation. In the following sub-section, will be explained how these techniques were applied.

A. *Semi-structured Interview*

Semi-structured interviews are in-depth interviews where the respondents have to answer open-ended questions and thus are widely employed by different healthcare professionals in their research [19].
In this study, we collect data among RORENO team through semi-structured interviews. The main topics explored in the interviews were: roles and permissions characterization; main functionalities of the health information system; explanation of the core processes including how the data is collected and the which are the main sources (both institutions and information systems applications). It was realized five interviews with the duration of one hour each, between April and May of 2016.

B. *Observation*

Observation is a type of qualitative research method which not only included participant's observation, but also covered ethnography and research work in the field. Observational data can be integrated as auxiliary or confirmatory research [19].
In this study, we decided to use observation to validate the information collected in the interviews, mainly the core processes workflow. For this together, with the end-user (mostly pathological anatomy technicians and statistical technicians), we observed how the system works and the different steps to perform a task. After this we designed core processes using BPMN in an iterative way until we had an agreement of the correct flow of steps.

IV. Results

In this section, we presented the results of this research. This section is divided into the following sub-sections: description of the main architecture; main functionalities, data integration and roles characterization, data quality and problems identification.

A. *RORENO main architecture*

Figure 1 presents a general overview of the RORENO architecture. This system is stored in a reference oncologic hospital – IPO-Porto and managed by RORENO team. RORENO can be accessed by different healthcare institutions (e.g. hospitals, health centres…) in the North region and are within the national private healthcare network (RIS – *Rede*

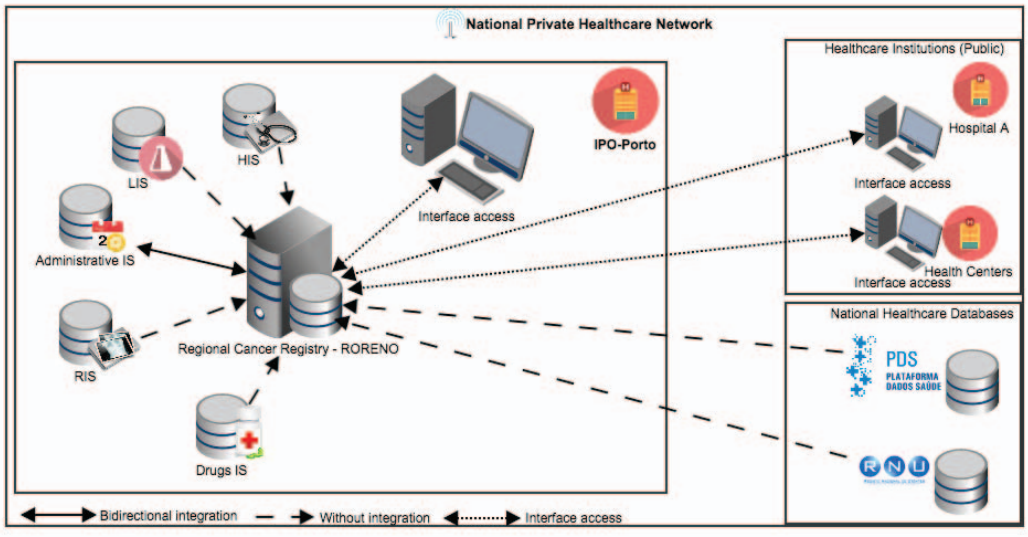


FIGURE 1. RORENO ARCHITECTURE

Informação de Saúde). Only the public institutions are within this private network. IPO-Porto is an important source of information to collect for RORENO since is the reference oncologic hospital in the North region and its also responsible to collect the cancer data between the north heath institutions to feed the RORENO system.

As we can see in Figure 1 RORENO is only integrated with IPO-Porto administrative database that includes administrative information about a patient such as demographic information and schedule (appointments, surgeries, vital state…). The others important sources of information such as departmental information systems, laboratory information system (LIS), radiologic information system (RIS) and drug information system doesn´t have an automatic integration with RORENO. So, in this case, RORENO team needs to access these systems one by one, patient by patient and collect manually the information needed.

In the case of other national relevant healthcare databases such as *Plataforma de Dados de Saúde* (PDS - Health Data Platform) and *Registo Nacional de Utentes* (RNU - national patients' registration database) that have crucial information to complete the registries, RORENO also doesn´t have an automatic integration. The team needs to access these platforms and collect the needed data manually. After collected the data, RORENO team fulfil a paper form (example available in [20]) and then transcribe for RORENO electronic form.

The external institutions collection of data is explained better in sub-section data integration.

B. *RORENO main functionalities*

The main functionalities identified in the study are:

- Forms to insert information of an individual case;
- Suggestion for data aggregation (information from several institutions from the same patient and same diagnosis);
- Excel export option (important to perform data analysis and produce incidence and survival reports);
- Excel import option (allows to insert excel data files);
- Follow-up forms (possible to receive and send automatically this information from/to the hospital administrative system).
- Worklist of new patients accepted in the hospital (information sent by the administrative hospital system) that is needed to complete medical information.

Figure 2 describes through a process map the main functionalities important to complete activities and what is the trigger, what are the relations between them and the output. The four boxes in the left column represent four different triggers, and the four boxes in the right represent four outputs. The boxes in the middle are core processes of this system and its relations.

By analysing the Figure 2, the three first core processes could be improved with a full automatic integration with main databases in IPO-Porto, integration with PDS and RNU database and an integration with external institutions. The performance of these three core processes in terms of time is very time and resource consuming and incurs a high risk of errors (due to manually interaction).

C. *RORENO data integration*

The main goal of this system is to aggregate cancer registries from different healthcare institutions in the region. However, the system is almost a stand-alone system and professionals needs to insert manually the data in the system. Although the institutions had the possibility to use the forms in the systems to insert their own data, for some institutions is more convenient to send the information by email (through an excel file) or by post office (paper-based records) and ask some assistance to the professionals that work with the system to insert this data. Figure 3 shows a scheme that summarizes these three kind of processes selected by external Institutions.

In a total of thirteen-three institutions that belongs to this region, six are a healthcare institution type 1, twenty-one are institutions type 2 and six are institutions type 3. These institutions include public hospitals and health centres and

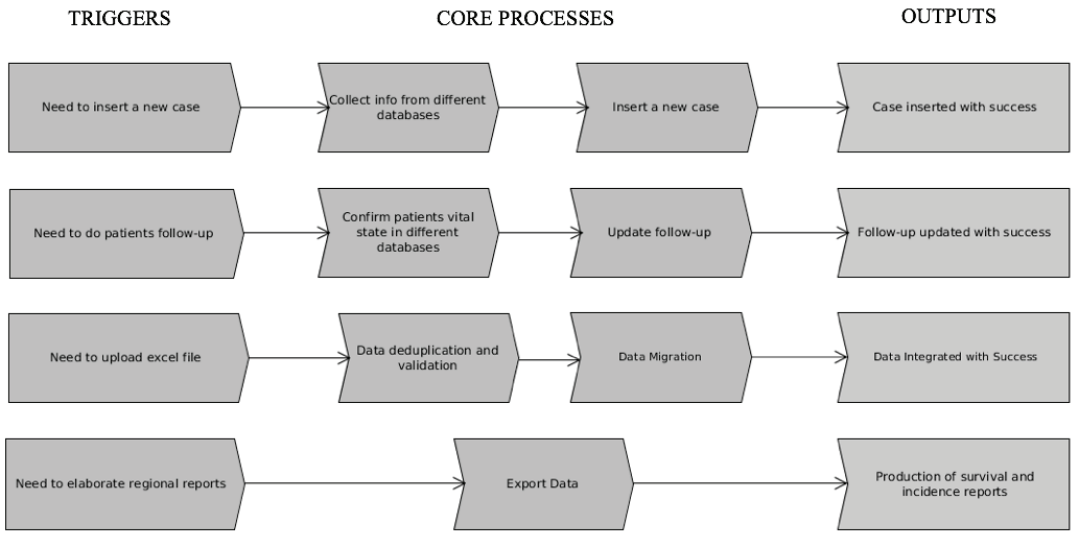


FIGURE 2. RORENO PROCESS MAP

private pathological anatomy laboratories. Since private pathological anatomy laboratories cannot access the private network of the Portuguese Ministry of Health they cannot share their data through a direct integration. The only options they have is to send the information (through excel file or paper).

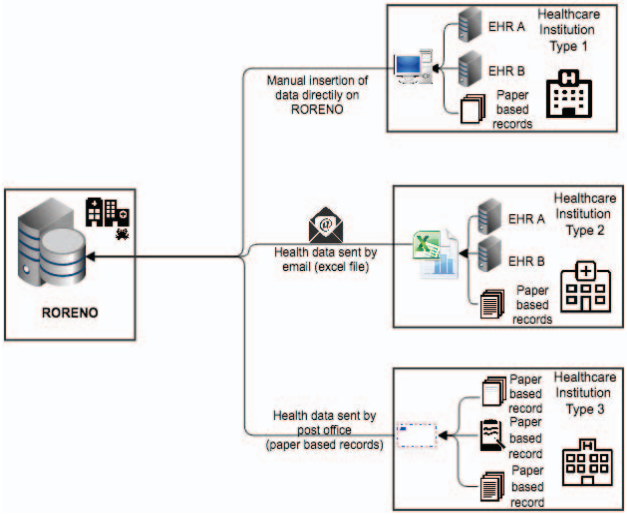


FIGURE 3. DIFFERENT PROCESSES TO INTEGRATE DATA IN RORENO

D. *RORENO roles characterization*

In this sub-section, we identified the main profiles of users and its permissions/tasks. These users belong to RORENO team. In terms of front-end management, we identified: administrator, responsible for entering data and external institution profile.

- Administrator: This role is responsible for the management of users, integration of files (Excel) provided by external institutions, data validation (manually and using IARC Check tool) and have the permission to create new variables in the system (allowing the adaptation of the system).
- Responsible for entering data: This role is performed by several users and has in charge the responsibility to insert data in the system. Due to interoperability problems, sometimes they need to insert variable by variable in the system manually. RORENO team need to insert this information from different institutions. They also have to search follow-up information about the patients (e.g. vital state). RORENO team search (manually) in national patient databases such as RNU, contact by phone others institutions, city responsible, families etc. They also use the IARC check tool [21] to validate the data quality. And have permissions to export and explore data to produce the incidence and survival reports.
- External Institution: Each external institution in the region has credentials to access this system and fulfil the cancer registry forms (manually or automatically). They can also insert information about the patients' follow-up and can export and explore the data. The main difference between this role and the role before is that each external institution can only view information about their own patients.

In terms of back-end, the informatics department of the healthcare institution that holds the cancer registry ensure the availability, security of the system and the backup copies to use in case of failure.

E. *Data quality*

Since they receive a significant amount of data by excel files to integrate in the system, they use a tool (IARC check [21]) to scan the file and verify some inconsistencies between variables, for example [16]:

- Age, Incidence Date, Date of Birth
- Age, Site, Histology (ICD-O-3 classification)
- Site, Histology (ICD-O-3 classification)
- Sex, Site (ICD-O-3 classification)
- Sex, Histology (ICD-O-3 classification)
- Behaviour, Site (ICD-O-3 classification)
- Behaviour, Histology (ICD-O-3 classification)
- Grade, Histology (ICD-O-3 classification)
- Basis of diagnosis, Histology (ICD-O-3 classification)

F. *Main problems identified in RORENO*

From the interviews with the end-users of the system, the IARC check [21] isn't enough to verify the data quality, they need to make an extra effort to manually check if there are: duplicate data, incomplete data, and other types of errors. This data validation plus the insertion of almost all data manually is a process very slow that requires time from various professionals. One of the reasons that RORENO team think other institutions don't use RORENO is because a lack of budget to have professionals dedicated to insert the data in the system since they do not have an automatically integration implemented.

By now RORENO team published the reports of regional cancer incidence and cancer survival of 2010 year [20] and are working on 2011 data. Since we are in 2017 this delay is a problem. However, this seems to be a world-wide problem. In this research it was searched for the newest cancer incidence report and survival reports in Europe and the most recent numbers are from 2013-2014 [22][23][10]. We don't have clues about the problems in other countries but in our case-study an investment improving the actual RORENO version could be a path to attenuate the problem.

V. DISCUSSION

In Europe, the progress that has been made deserves to continue, and the current momentum of attention and support for quality cancer registration should not be lost. The reduction of disparities in the quality and function of cancer registries is a cornerstone, as are the efforts to harmonise,

standardise, and bring together in a comparable and understandable way the wealth of cancer data across the continent [9].

In this study, we made a maturity diagnosis of a computerized cancer registry implemented in 2005. Since 2005 the informatics in the healthcare institutions has been evolving and the sources of data that were paper-based or even now are robust electronic systems that already supports standardized protocols for integration like HL7 (*Health Level 7*) and DICOM (*Digital Imaging and Communications in Medicine*). This scenario reflected not only the healthcare institutions but also the national projects promoted by the Portuguese Health Ministry like national patients' database.

Epidemiological research depends on the balance between preserving patients' integrity and anonymity while also enabling important research to improve people's health and the quality of care [13], until 2018 the healthcare institutions should be prepared for comply the changes in the data protection (GDPR), so this system should comply with these new privacy and security requirements.

VI. Future Work

After the diagnosis of RORENO *as is model* we pretend to study which improvements this system should implement. It is crucial to automatize the integration of data linking the different healthcare institutions in the region. However, it is important to think which functionalities this system should give to institutions in the network in order to maximize the engagement with the project. More than a database this should be a source of knowledge available to all the collaborative oncologic network.

Support tools like business intelligence (BI) and an I&D platform could be implemented within this new project. Some key performance indicators like effectiveness in screening programs, impact of new drugs, adherence of drugs, effectiveness of different approaches (surgical/ radiotherapy/ chemotherapy) for each type of neoplasia could be tracked almost in real-time. This monitoring could have a great impact in daily decision making.

Acknowledgements

This work was supported by project ODISSEIA - Oncology Disease Information System [POCI-02-0550-FEDER-012618] and project NanoSTIMA - Macro-to-Nano Human Sensing: Towards Integrated Multimodal Health Monitoring and Analytics [NORTE-01-0145-FEDER-000016]. It was also supported by FCT (*Fundação de Ciência e Tecnologia*) through the doctoral fellowship [SFRH/BDE/105533/2014].

References

1. Bianconi F, Brunori V, Valigi P, Rosa F La, Stracci F. Information Technology as Tools for Cancer Registry and Regional Cancer Network Integration. IEEE Trans. Syst. MAN, CYBER NETICS—PART A Syst. HUMANS,. 2012;42(6):1410–1424.
2. O.M. Jensen, D.M. Parkin, R. MacLennan CSM and RGS. Cancer Registration : Principles and Methods. IARC Sci. Publ. 1991;95.
3. SNS. Registo Oncológico Nacional [Internet]. SNS - Serviço Nac. Saúde. 2017;Available from: https://www.sns.gov.pt/noticias/2016/12/19/registo-oncologico-em-2017/
4. WHO/IARC. International Agency for Research on Cancer. WHO. 2014;
5. Colquitt S, Clements K, Hart-hester S. The Impact of Electronic Health Record Usage on Cancer Registry Systems in Alabama. Perspect Heal. Inf Manag. 2012;9.
6. Jerzy E. Tyczynski, Eva Démaret DMP. Standards and Guidelines for Cancer Registration in Europe. IARC Tech. Publ. 2003;(40).
7. Blobel B. Clinical Record Systems in Oncology . Experiences and Developments on Cancer Registers in Eastern Germany. :37–55.
8. EPAAC. EPPAC: European Partnership for Action Against Cancer [Internet]. 2011;Available from: http://www.epaac.eu/cancer-data-and-information
9. ENCR. European Network of Cancer Registries [Internet]. 2016;Available from: http://www.encr.eu/
10. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, et al. Cancer incidence and mortality worldwide : Sources , methods and major patterns in GLOBOCAN 2012. Int. J. Cancer. 2015;386(15).
11. Bray F, Brewster DH, Mbalawa CG, Kohler B, Swaminathan R, Ferlay J. Cancer Incidence in Five Continents Vol . X. 2014.
12. Siesling S, Louwman WJ, Kwast A, Hurk C Van Den, Callaghan MO. Uses of cancer registries for public health and clinical research in Europe : Results of the European Network of Cancer Registries survey among 161 population-based cancer registries during 2010 – 2012. Eur. J. Cancer. 2015;51(9):1039–1049.
13. Andersen MR, Storm HH, Work E. Cancer registration , public health and the reform of the European data protection framework : Abandoning or improving European public health research ? Eur. J. Cancer. 2015;51(9):1028–1038.
14. Conseil P européen et du. Directive 95/46/CE. 1995;0031–0050.
15. Commission E. General Data Protection Regulation. Eur. Com. 2016;
16. IPO-Porto. RORENO - Registo Oncológico Regional do Norte [Internet]. 2017;Available from: http://www.roreno.com.pt/
17. República D da. Portaria 35/88. 1988;
18. IPO-Porto. IPO-Porto [Internet]. 2017;Available from: http://www.ipoporto.pt/en/
19. Jamshed S. Qualitative research method-interviewing and observation. J. Basic Clin. Pharm. 2014 Sep;5(4):87–88.
20. RORENO. Registo Oncológico Regional do Norte 2010. IPO-Porto. 2015;
21. IACR. IACR - International Association of Cancer Registries [Internet]. 2017;Available from: http://www.iacr.com.fr/index.php?option=com_content&view=article&id=72:iarccrgtools&catid=68&Itemid=445
22. Data for cancer frequency by country [Internet]. World cancer Res. fund Int. 2017;Available from: http://www.wcrf.org/int/cancer-facts-figures/data-cancer-frequency-country
23. UK C research. Cancer statistics for the UK [Internet]. Cancer Res. UK,. 2014;Available from: http://www.cancerresearchuk.org/health-professional/cancer-statistics