PARENT
cross-border
PAtient REgistries iNiTiative

Co-funded by the European Commission

Discussion paper for eHealth Network

"Guidelines on patient registries and supporting tools" [1]

Background

Article 14(2)b of the Directive on patients' rights in cross-border healthcare states that among the objectives of the eHealth Network will be to ***draw up guidelines on effective methods for enabling the use of medical information for public health and research***.

The efforts of the eHealth Network in the areas of Patient Summaries and ePrescriptions have thus far successfully promoted the interoperability of data generated during the provision of care. Ensuring the interoperability of existing data collections in the context of their secondary use is the logical next step, closing the circle in the added value generated by eHealth tools and applications.
Patient registries are repositories of medical information which, due to their sheer number and large volume, present significant potential for research and public health improvements in the EU. Examples of patient registries' utilization areas include the monitoring of patients' treatments and their safety assessment, health technology assessment and evidence-based policy creation utilizing "real world" data and the trends of translational medicine (e.g. registry-based clinical trials, personalized medicine).

What are patient registries?

For the purpose of PARENT Joint Action work, a **patient registry is defined** as an organised system that collects, analyses and disseminates data on a group of people defined by a particular disease, condition, exposure, or health-related service, and that serves scientific, clinical and/or public health (policy) purposes.
Registries should be designed and evaluated with respect to their intended purpose(s) which can broadly be described in terms of patient outcomes. Some of the main **purposes** for establishing and running a patient registry are to **describe the natural history of a disease**, to **determine clinical effectiveness and/or cost-effectiveness**, to **assess safety or harm**, to **measure quality of care** and to serve **public health surveillance and disease control**. In broad terms, patient registries contribute to the improvement of patient care and healthcare planning, to social, economic and quality of life outcomes monitoring and to the production of other health indicators (e.g. access to health-care, health status, subjective and objective quality, healthcare systems financing, DRG settings etc.)

[1] Prepared by the Patient Registries iNiTiative Joint Action (PARENT), www.patientregistries.eu

Challenges

The quality and structure of data currently held in patient registries is however inconsistent and – due to lack of use of common methodology – valuable data collections often remain unavailable for broad research and public health purposes. Processes and legal agreements for data sharing across registries and Member States are seldom established. Time-consuming search and identification of suitable data sources constitutes another important barrier.

Although there are some best practices in particular areas, the need for a generalised methodology across diseases and medical cases, such as that proposed by PARENT Joint Action, is of paramount importance in ensuring long term use and added value of this growing amount of medical data. Patient registry holders and a number of EU stakeholders have already expressed the need for EU level sharing of registry-related knowledge and best practices, as well as for tools and services improving quality of data and data availability.

Agreeing on a set of guidelines and tools for making patient registries interoperable across the EU is an essential step towards efficient secondary use of data.

What is PARENT Joint Action doing?

PARENT is a joint effort by Member States and the European Commission as a direct response to the objective set in Article 14(2)b of the Directive. PARENT aims to improve secondary use of data from patient registries in a cross-border setting for both public health and research needs.
Based on analyses performed, these objectives can only be achieved by:

(1) improving use of primary data sources for feeding data into patient registries[2];
(2) improving data quality and interoperability of new and existing patient registries;
(3) mapping of patient registries in the EU with the purpose of supporting search and identification of available data sources; and exchanging information about national best practices and lessons learnt on patient registries;
(4) supporting data sharing between and across registries; and data reporting to authorities and relevant bodies;
(5) providing support services for registry holders at EU level.

JA PARENT is now finalising several deliverables. Among them should be mentioned the online tool "Registry of Registries", literature review, and the report on proposals for future steps and actions supporting the implementation of the Directive on patients' rights in cross-border healthcare. One of the key deliverables of PARENT is the Methodological guidelines for efficient and rational governance of interoperable patient registries (the *Guidelines*) and related implementation tools and services. The knowledge and platforms developed within other projects, Joint Actions and other initiatives (like the Patient Summary Guidelines) have served as the reference base for this work.

[2] Where PARENT recognizes the results and ongoing work of several projects focusing on improving interoperability of Electronic Health Records (EHR) as primary sources of data (i.e. SemanticHealthNet, EHR4CR, TRANSFoRM, EMIF, SALUS);

Topics for discussion by the eHealth Network

Taking into account the experiences of Member States in designing and implementing patient registries at national level; experiences in applying interoperability and knowledge of how Member States are using technologies to respond to the health needs of the population, **the eHealth Network Members are invited to**:

- discuss the potential usability of the Guidelines in a cross-border setting and on a national level;
- discuss the follow up by the eHealth Network (update of the Guidelines and report on implementation according to the eHN Multi-Annual Work Programme 2015-2018);
- comment on the Guidelines presented in the Summary;
- discuss and possibly endorse the Patient Registries' Summary Guidelines or Comprehensive Guidelines at the eHN meeting (May 2015).

Experts (registry holders, eHealth, etc.) from Member States will be invited to participate in a workshop (proposed dates in February 2015) where the Guidelines and supporting tools will be presented and discussed in detail.

APPENDIX I

Summary of the Methodological Guidelines and recommendations for efficient and rational governance of patient registries

WORK PACKAGE	WP5
DOCUMENT NAME	Summary of Methodological guidelines and recommendations for efficient and rational governance of patient registries
DOCUMENT VERSION	1.0
DATE	03/11/2014

Project Title	**Cross-border Patient Registries Initiative**
Project Acronym	**PARENT**
PARENT Coordinator	Matic Meglič, NIJZ (till 09.10.2014) Marija Magajne, NIJZ (since 10.10.2014)
Work Package Leader	Metka Zaletel, NIJZ

Work Package	5
Deliverable	
Title	Methodological guidelines and recommendations for efficient and rational governance of patient registries
Version	1.0
Dissemination level	Confidential

Authors of the Guidelines

Metka Zaletel, Matic Meglič, Vesna Lešnik Štefotič, Marcel Kralj, Dmitri Wall, Alan Irvine, Eoin O'Brien, Marko Brkić, Borna Pleše, Vanja Pajić, Mladen Kostešić, Ranko Stevanović, Tamara Poljičanin, Ivan Pristaš, Óscar Zurriaga, Carmen López Briones, Miguel A. Martínez-Beneito, Clara Cavero-Carbonell, Rubén Amorós, Juan M. Signes, Alberto Amador, Carlos Sáez, Montserrat Robles, Juan M. García-Gómez, Carmen Navarro-Sánchez, María J., Sánchez-Pérez, Joan L. Vives-Corrons, María M. Mañú, Laura Olaya, György Surján, Károly Fogarassy, Haralampos Karanikas, Ioannis Skalkidis Persephone Doupi, Arto Vuori, Hanna Liukkunen, Antti Tuomi-Nikula

Authors of the Summary

Metka Zaletel, Vanja Pajić, Dr Dmitri Wall, Óscar Zurriaga, Vesna Lešnik Štefotič, Marcel Kralj, Persephone Doupi, Arto Vuori, Haralampos Karanikas, Marko Brkić

Table of contents

1 INTRODUCTION

As a result of the diversity of EU Member States' specific needs and legislation, the complexity of the health domain and variable progress in IT implementations, there is currently limited standardisation across registries and other eHealth tools. The PARENT Joint Action (www.patientregistries.eu) aims to support the development of comparable and interoperable patient registries, thus enabling secondary data usage for public health and research purposes (e.g. HTA) in a cross-border setting.

Two parallel actions are underway to achieve the long term general objective of the PARENT JA in: first, a comprehensive set of recommendations and specific guidelines to support MS in methodology, development, implementation, governance and improvements of national and local patient registries are being drafted. Second, activity plans, business model proposals and policy proposals to ensure sustainability of cross-border collaboration on usage of patient registry data and proposals for future activities to support implementation of a cross-border health care directive are being developed.

The main results of the PARENT JA - Work Package 5 are the "Methodological guidelines and recommendations for efficient and rational governance of patient registries" (hereafter referred to as Guidelines). The target group of the Guidelines is registry holders, researchers, healthcare providers, developers, and in general, competent authorities responsible for registries etc. who are intending either to set up a new patient registry or to modernize an existing patient registry. One of the purposes is also to upgrade the registries that do not fulfil the minimum requirements for operating in the cross-border setting.

The process of the Guidelines preparation has been complex, with many different topics. Work Package 5 involved approximately 40 authors drafting the chapters. The first step was the drafting of the structure of the Guidelines; secondly, PARENT JA partners, Associated Project Group and stakeholder representatives reviewed the structure and discussed it at a workshop held in December 2013. In May 2014, the structure was presented at eHealth Governance Initiative representative meeting. The first draft of the Guidelines was then prepared and a second workshop was organised to discuss open issues in June 2014. At present, the Guidelines are in the editing phase and will be finished by the end of the year. It should be noted that the work[3] of the Agency for Healthcare Research and Quality (AHRQ) has been of considerable help in informing the authors and in structuring this Guidelines.

The Guidelines will be supported by a WIKI[4] tool to provide all users with quick and effective access to the information they need for their current activities.

This document serves as a summary of the Guidelines and the presentation of the topics covered. The structure of the document is the same as the structure of the first level of the Guidelines (chapter levels).

[3] Especially the publication "Registries for Evaluating Patient Outcomes: A User's Guide" [1].

[4] A wiki is a web application which allows people to add, modify, or delete content in collaboration with others.

2 PATIENT REGISTRIES - OVERVIEW

2.1 Definition of a patient registry and types of registries

The terms "register" and "registry" are often used interchangeably and terminology in this field can therefore be confusing [2]. However, a registry is the organisation and process that supports a register and should be distinguished from the register itself. One registry may support a number of individual registers.

For the purpose of PARENT work, a **patient registry is defined** as an organised system that collects, analyses and disseminates data on a group of people defined by a particular disease, condition, exposure, or health-related service, and that serves scientific, clinical and/or public health (policy) purposes (adapted from the AHRQ's "Registries for Evaluating Patient Outcomes: A User's Guide"[1]).

Registries should be designed and evaluated with respect to their intended purpose(s) which can broadly be described in terms of patient outcomes. Some of the main **purposes** for establishing and running a patient registry are to **describe the natural history of a disease**, to **determine clinical effectiveness and/or cost-effectiveness**, to **assess safety or harm**, to **measure quality of care** and to serve **public health surveillance and disease control**. In broad terms, patient registries contribute to the improvement of patient care and healthcare planning, to social, economic and quality of life outcomes and to other health indicators (e.g. access to health-care, health status, subjective and objective quality, healthcare systems financing, health technology assessment, DRG settings etc.)

With the help of information and insights gathered from various activities of PARENT (literature review, questionnaire survey of registries for the Registry of Registries (RoR) pilot), and with concern to the above stated complicacy of taxonomy of registries, a several level **classification of patient registries** is offered in the Guidelines.

Registries are classified according to how their populations are defined. For example, disease or condition registries are defined by patients having the same diagnosis, such as cystic fibrosis or heart failure, or the same group of conditions such as disability. Product registries include the data of the patients who have been exposed to biopharmaceutical products, medical devices or diagnostic/therapeutic equipment. Health services registries consist of patients' data who have had a common procedure, clinical encounter or hospitalisation [1].

- **Disease or condition registries:** the main inclusion criterion for disease or condition registries is the state of a particular disease or condition. That state varies, as the patient may have a lifetime disease (e.g. rare disease such as cystic fibrosis or a chronic condition such as disability) or a short term diseases (e.g. infectious disease). The disease registry could be hospital/clinic-based or population based.
- **Product registries**: aim to assess safety or harm associated with the use of various products must anticipate and assess the need for adverse event (AE) detection, processing, and reporting and registry sponsors are encouraged to discuss plans for AE collection and processing with local health authorities when planning a registry.
- **Health services registries:** may be created to measure and improve the quality of care, defined as "the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge" [3]. This kind of registry usually compares patients' data or sometimes providers based on the treatments or achieved outcomes in relation to performance measures with "gold standards" or some other benchmarks for specific health outcomes (e.g. infection rates). These type of registries may be used for various purposes: to monitor trends in the use of certain clinical

procedures and to evaluate trends in healthcare usage; to examine provider adherence to safety protocols and best practice guidelines; to monitor the impact of prevention efforts and public health awareness campaigns; to survey the quality of care patients receive.

2.2 Overview of European registries

The current European registry landscape can be best described as a collection of divergent registries. Design, development, and maintenance of patient registries revolve around registry platforms (software tools for managing registries' data). This approach leads to the creation of segregated silos, resulting in expensive and inflexible IT systems. Registries are often built for a single purpose, with their own data stores and for limited user profiles. Furthermore, registries have different legislative and governance rules and obligations and are spread across various European countries and types of organisations. As a result, patient registries implement only a subset of the potential registry functions, using and producing only a subset of the registry data, and often not applying existing interoperability approaches (standards, best practices). Thus, these registries manifest themselves as islands of data and governance rules.

2.3 Key issues in registry development and governance at EU and national level

The development process of the PARENT Registry of Registries unveiled a variety of key issues concerning the development and governance of registries at EU and at national level. The most important are:

1. **Unstable funding and** therefore **limited sustainability**.
2. Many **legal issues** concerning registry set-up, data protection and re-use of data.
3. Unclear roles of different stakeholders within registry development and governance: data ownership, the role of data holder, data user, etc.
4. **Modes of data collection**: almost half of the EU registries included in RoR are still based on paper-and-pen mode (paper-based questionnaires, paper-based health records and laboratory results).
5. **Lack of awareness of existing standards and standard processes** when building or maintaining a patient registry. These standards are actually wanted by registry holders.
6. Balance **between accuracy and timeliness** is usually skewed in favour of accuracy, resulting in low timeliness. **Comparability** over time and/or space (as another quality component) is often limited due to set-up procedures, specific funding, etc.
7. **Data quality** (including completeness) is often compromised. There is low awareness of existing quality standards and there is also a lack of knowledge about quality assessment.
8. **Registry transparency** and **openness** with the emphasis on data access for research purposes
9. **Insufficient data dissemination**

3 INTEROPERABILITY

As already noted, patient registries in the EU, spread across different countries and types of organisations, manifest themselves as islands of data and governance rules. The EU eHealth Action Plan 2012-2020 [4], which defines the overall operative context, focuses among others on challenges of interoperability of eHealth services. This also applies directly to interoperability of patient registries. Patient registries are key healthcare information repositories, therefore interoperability of their stakeholders and users is crucial part for execution of the eHealth Action Plan. By applying existing interoperability approaches (standards, best practices) and ultimately achieving interoperability, it is possible to prevent potentially valuable data from becoming trapped in information silos and facilitate more accurate representation of concepts and cross-border comparison of data.

When discussing patient registries with interoperability as a goal, stakeholders need to review, verify and agree on requirements for four interdependent interoperability levels, as recognized by the EIF [5][6]. The interoperability levels "classify interoperability concerns according to who/what is concerned and cover, within a given political context, legal (1), organisational (2), semantic (3) and technical (4) interoperability", where each interoperability level has its own objective and eliminating any level from the process can result in inadequate solutions.

The PARENT JA Guidelines contain patient registry interoperability recommendations intended to help patient registry stakeholders such as registry holders and researchers in grasping their interoperability environment. The Guidelines include clarifications of the interoperability levels, with guidance on achieving interoperability through a top-down stepwise approach to achieving interoperability levels starting with identifying the political context and then moving down to legal, organisational, semantic and finally technical interoperability.
Moreover, PARENT Joint Action proposes the PARENT Framework. This is an objective-based interoperability framework, with a function to provide a shared infrastructure for development of common interoperability support and functionalities to all stakeholders and projects joined around PARENT objectives in order to achieve EU-wide registry interoperability. It should support the full interoperability range and functions as its integrator. It is envisaged as an always active mechanism governed by key stakeholders on the political context level, providing support and service repositories for all interoperability levels. On the strategic level it is intended to provide a means to unify, standardise and deliver functions needed by all participating and potential stakeholders and to gather and disseminate information and knowledge that can generally speed up interoperability development among the target group. On the operational level it is intended to develop functions for the support of interoperability harmonisation, project deployment and integration of project outcomes in the Framework.

4 GENERAL REQUIREMENTS FOR CROSS-BORDER USE OF PATIENT REGISTRIES

Cross-border use of registries can take several different forms as the mapping work of PARENT has demonstrated, among others registry networks (e.g. the International Association of Cancer Registries, the Nordic Arthroplasty Register Association NARA), international clinical studies (GRACE – Global Registry of Acute Coronary Events) and international registries e.g. IBIR - International Breast Implant Registry). There are several strong drivers in using registry data across borders, such as the needs of studying differences between countries in morbidity, health system-level interventions' effectiveness and utilization of procedures used; the advantages of large international datasets vs. national ones in timely detection of rare or previously unknown effects; gathering and promoting information on best practices worldwide.

Independent of the motive driving cross-border registries utilization, the success of the endeavour will always rely on the degree of achievement of certain prerequisites, the implementation of which starts at the local (regional and/or national level). The purpose of fulfilling these prerequisites is the achievement of interoperability in the broadest understanding of the term, i.e. on legal, organizational, semantic and technical levels [6]. The focus of this chapter is primarily on the requirements imposed by legal and organizational interoperability aspects, and to a lesser extent on semantic and technical interoperability issues; these are in turn addressed in more detail in Chapters 3 and 5.5. An exception is the topic of metadata, which we briefly discuss here. Further analysis of organizational interoperability aspects will constitute part of the business models analysis of PARENT, an independent deliverable which is currently work in progress.

The most important European law affecting patient registries' operations is the Data Protection Directive (95/46/EC) that regulates the collection, processing and distribution of personal data. Registry holders should always be aware of the basic notions and effective norms of Data Protection, as securing privacy of the research subjects is a fundamental task when establishing and maintaining a patient registry. Currently the implementations and interpretations of the Data Protection Directive vary between Member States. Additionally the roles of Ethical Committees and data protection authorities vary a lot. The legislative process toward the new harmonizing Data Protection Framework is still unfinished. Moreover, the European Union Directives and Regulations considering Medical Devises, Pharmacovigilance, Clinical Trials and Cross-Border Health Care induce new information needs that will increase demand for patient registry data. Registry holders should actively follow the ongoing overhauls of the aforementioned laws.

By and large a patient registry can be established using either of two legal instruments; by explicit consent of the data subject, or based on law. Current practices among the EU Member States registry holders' surveyed by PARENT appear to be almost equally divided between the two models. The final content of the forthcoming Data Protection Regulation will play a decisive role in the choices available for registry establishment and operations in the future.

The generalized interpretation has become that even encrypted and pseudonymous data are personal data. That is why it is pivotal to understand the basic notions regarding personal data in order to understand the areas where Data Protection Rules are applicable. Registry holders and data processors should always be able to differentiate clearly the notions of pseudonymous data, encrypted data, anonymised data and aggregated data (*possible reference to PARENT glossary*).

It is likely that the upcoming European Data Protection Framework will require more transparency and accountability from patient registry holders. Generally it is advisable to be open about the registration purposes and give clear information to maintain public trust and credibility of patient registries. This involves ethical and well-structured informed consent practices, as well as maintaining clear and open descriptions of the registry and its metadata online.

Metadata is "structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information source". It is meant to describe the phenomenon it concerns, and also document its changes over time. Good quality metadata is vital for data utilization. To make datasets comparable and useful for other users and between registries, metadata should be standardized according to validated and widely used classifications. Another aspect of standardization is recording metadata elements in the register's information model. That is, to make standardization as complete as possible, it must also cover data architecture and programming details.

When establishing and maintaining a registry, it is pivotal to identify the relevant stakeholders and generate a co-operation structure within them. The key stakeholders from the registry holders'

perspective are usually health care professionals, patients, drug and medical devices industry, ICT-suppliers, policy makers, researchers and other registries. If taken further, the opening of detailed metadata in standardized format would ease the registries' multi-stakeholder cooperation as well, particularly in the cross-border setting. The first step in opening registry metadata could include basic information about the data, such as description, owner, information content, target group, update intervals, dependencies from other data etc. This kind of increase of visibility would benefit patient registries and lead to new ideas and innovations.

As open data has recently gained importance also on state administration level (e.g. the British government's "opening up government" initiative and the Finnish Ministry of Finances' "open data programme"), open data and the possibilities it may yield must be carefully considered in the patient registry environment. Firstly, a line must be drawn between the data which can be opened given the technical, and, above all, data security restrictions, and the data which cannot be opened (such as patient registries' microdata).

Researchers' access to classified registry data has generally been quite complicated and time consuming starting with locating appropriate data, preparing research applications and on to requesting permissions and negotiating data transmissions. New solutions of more straightforward application processes and remote access to data are being developed. Open data is an overarching idea which stretches to cover parts of classified data in the form of metadata. Openly publishing the content information of limited access systems would boost the efficiency of scientific research, enhance the quality of results, increase transparency and help create new research ideas.

The creation, maintenance and development of registries is dependent on the positioning (or lack thereof) of health data resources on national level strategic prioritization concerning scientific data resources and research infrastructures. PARENT is analysing in a parallel activity national strategies and initiatives concerning Health Data and the ways in which they impact patient registry work.

5 CREATING A REGISTRY

5.1 Planning a registry

This document assumes that a registry is designed to fulfil a need that can be met through the scientific analysis of predefined data, collected in a real-world setting. While this data might ultimately be utilised to answer other questions, it is essential that registry establishment is an organised, well governed and purposeful scientific process rather than a purposeless exercise in data collection. This will ensure the creation of a resource that maximises resource allocation, efficiency and has well-defined, valuable outputs that can be measured, so that the quality and success of the registry can be verified.

This section evolves a logical, sequential process, components might be best addressed in tandem or may need to be revisited in an iterative fashion as further information becomes available. Addressing each section will add value to the registry and increase the likelihood of developing a successful registry.

During each phase of planning, PARENT Joint Action advises considering how it may fit into the bigger picture, not just of the registry that is being created, but also with respect to the local, regional, national and international environment in which it is created. As the digital world becomes more connected, the role of registries becomes progressively more valuable. This will only happen if they are developed in a manner that is cognisant of the importance of interoperability.

There is a wealth of experience to be gained from regulatory authorities, other registry groups and registry experts whose contributions could not only be helpful in the construction of a successful registry, but also critical to its implementation. PARENT JA endorses in particular the creation of a resource such as PARENT's Registry of Registries or the AHRQ's Registry of Patient Registries (RoPR), which help to connect registries while raising the standard of registries considerably [7][8]. There is a strong recommendation on ensuring that any registry created forthwith joins such initiatives.

5.2 Registry study design

The process of developing a study design[5] for a registry includes different elements and factors that registry holder needs to consider. This chapter describes those elements and covers various important aspects that must be taken into account when establishing a registry study design.

During a registry study design phase, a registry holder must first to develop the research questions and/or hypotheses. When defining them, it essential that they are accurate, understandable and focused enough for a specific registry. The next step is the identification of key exposure and outcome variables. When identifying the key exposures and outcomes, it is important to note that more outcomes sometimes need to be selected (as a result of multiple questions of interest), and exposure often includes a collection of different information, such as dose, duration of exposure, route of exposure, and adherence [1][10]. It is useful to take into account independent risk factors for the outcomes, and confounding variables as well.

The research questions and the key exposures and outcomes of the registry directly influence the study design decisions. One of these is the selection of a registry study model. The chapter describes several study models that are more commonly applied in registries, namely cohort study, case-control study, nested case-control study, case-cohort study and case series.

Selecting patients for a registry is another important consideration when developing the registry study design. It is crucial to define the target population accurately, since it is a key factor in determining the registry sample. The registry holder should specify criteria that define what types of patients/cases are going to be included in a registry (so-called inclusion criteria), and criteria that disqualify subjects from inclusion in the registry (i.e. exclusion criteria). Concepts such as a sampling frame, sampling, the non-coverage issue and representativeness are introduced here in detail, because these are the elements that affect the registry study results.

An estimation of anticipated registry size is a relevant part of the planning process. If a registry will include only a sample of the target population, it is recommended to estimate prematurely how many cases the registry is planning to include. If the registry is too small, it may have insufficient analytical power and it may not ensure adequate exploration of the objectives. On the other hand, a registry that is too large may waste time, resources and money. The chapter lists some points that are must be taken into account when estimating registry size and provides some tools for sample size calculation.

The duration of the registry study (enrolment and follow-up) should also be specified when developing a registry. The duration of a study depends on what type of study it is, what the specific procedures in the registry are and what objectives have to be met. Some registries collect data at only one time point and others collect data for the lifetime of the patient. A registry may be open-ended or it may have a fixed end point when enough data to achieve the study objectives is expected to have accrued [11]. If we neglect the funding as the biggest factor for registry duration and sustainability, the factors that

[5] A study design is a specific plan or protocol for conducting the study, which allows the investigator to translate the conceptual hypothesis and research question into an operational one [9].

the registry planner should consider when estimating registry duration include the induction period for desired outcomes, sufficient follow-up time for the exposure, registry design (e.g. data collection method, sample size, complexity of the data being collected), anticipated accrual of enrolled subjects and deadlines for the dissemination of results [1][12].

The chapter ends with the presentation of a case study: developing registry study design for the Slovenian Arthroplasty Registry.

5.3 Data elements for a registry

The process of selecting and building data elements is one of the most important and challenging tasks that often determines the final success of the registry. If the registry does not collect data that would fulfil its intended purpose and goals, it can turn out to be useless. On the other hand, if the registry sets too complex data collection, inducing higher costs and burden, this may jeopardise its sustainability. Hence, a careful approach is required and many aspects need to be taken into consideration when building a dataset for a registry. The aim of this chapter is to highlight these aspects and describe the process of building a registry dataset.

The selection of data elements for a registry starts with the identification of the data domains which are collections of data elements that relate to a common topic. Data domains that are commonly used in registries include the patient domain, provider domain, exposure domain, outcome domain covariate/confounder domain and administrative domain [1][11][13]. The chapter provides detailed descriptions of these domains.

The general aspects and principles of building a registry dataset are:

- A minimalist approach to building a dataset
- Explicit definitions
- Use of data standards
- Consideration of data quality for data elements
- The privacy aspect
- The burden and costs of data collection
- Availability of data sources for data elements
- Deciding on the minimum data set
- Testing the dataset
- Well-documented data elements
- Accessibility, usability and visibility of the dataset
- Consideration of the impact of changes when modifying data elements

The chapter emphasises the importance of using standard classifications and terminologies, where a comprehensive list of existing classifications and terminologies is provided. PARENT common data set concept is proposed, which is the intersection and collection of existing and relevant common datasets that have been developed in various EU organisations/activities/actions.

The chapter ends with the description of the data dictionary (including guidance for its structure), and presentation of the case study: preparing data elements for the Slovenian Arthroplasty Registry.

5.4 Data sources for registries

The data sources used for registries can be classified into:

- Primary data sources are the ones collected from the individuals to create (or supplement) the patient registry.
- Secondary data sources are the sources that were established or collected previously for other purposes. Examples of these sources are EHRs, medical charts, various databases (e.g. hospital administration database, census database).

Primary data sources are in most cases costly, time consuming, but on the other hand can provide data of higher quality with regard to the dimensions of completeness, validity and reliability. When collecting data by a questionnaire or other research instrument, we also create a burden for data providers (patients, clinicians, etc.). The burden needs to be taken into account when planning a survey/data collection. Examples of primary data sources are patient reported data and clinician reported data.

Secondary data sources are on the other hand less costly and easier to gather – provided that there is a sufficient legal background. Nevertheless, there are some considerations that must be taken into account concerning usage of secondary data sources. Examples of secondary data sources are EHRs, human resource and financial databases, population databases and censuses, other health registries.

5.5 The role of information system methodologies and techniques in the phase of Patient registry creation

In the phase of patient registry development (designing the patient registry content and functions), it is very useful to take advantage of information system development methodologies, techniques and tools.

In this section on the role of information system (IS) methodologies and techniques, the following items are described:

- how and why different modelling techniques (from the field of IS design) can be applied in patient registry creation;
- how important it is to involve an IS expert (or other person with experience in IS modelling techniques) in the patient registry creation and to clearly understand the role of such expert;
- techniques for eliciting requirements/knowledge for patient registry;
- the importance of standard terminologies and code lists.

The main purpose of this section is to briefly introduce some of the most widely used IS design techniques and diagramming notations to the reader. After reading this chapter, the reader will:

- understand why modelling techniques are useful in patient registry creation;
- be familiar with some commonly used modelling notations[6] and terminology, and be able to read a model;
- understand the role of the IS expert (or other person with experience in IS modelling techniques) in patient registry creation;
- understand the importance of using standard terminologies and code lists if they exist.

For more information on this subject and the techniques described, readers will be encouraged to explore the links provided to free tutorials and additional reading.

[6] Notation = standardized way of presenting models (usually real world issues, like processes, things etc.)

6 PATIENT REGISTRY INFORMATION SYSTEM DEVELOPMENT AND IMPLEMENTATION

Patient registries are dealing with data and information, collecting it, looking for it, storing it and analysing it. It is therefore imperative that **information in patient registries is designed and managed in the most effective way possible** in order to ensure high quality and reliable outcomes **using information technology (IT)**.

In Chapter 6, the basics of patient registry (PR) information system development and implementation is described by presenting a typical software development lifecycle and emphasizing the importance of the (end) user in this process. PARENT JA address the different possibilities to obtain patient registry software (SW): in-house development, buying/using a PR SW product or outsourcing the development of PR SW. Training in PR software is essential to the proper and efficient use of the application and it is therefore covered as a separate topic.

The registry holder or any other reader will be able to:

- understand the basics of the SW development lifecycle and different SW development models;
- understand the importance of user involvement in SW development;
- develop an awareness of the different ways of obtaining PR SW.

7 RUNNING A REGISTRY

A registry is a complex structure, since it can be organised in many different ways and purposes. Therefore, running a registry is not a simple process, but a multifaceted procedure that requires technical knowledge, scientific aptitude and rigorous execution of a previously established plan.

In this chapter the sequential and the overarching processes to be followed in a patient registry are presented. The sequential processes include several aspects, as described below.

The way of **collecting data** for a registry is a crucial part, because it determines its feasibility. Different ways of collecting data (paper-based, electronic) are described with a definition of the registry case report form and related to the data-entry.

Getting **data from different sources** is almost compulsory nowadays in order to increase completeness and quality of the registry or to check the registry data. For that reason, methods for data linkage, deterministic and probabilistic, are reviewed.

Once data has been obtained, it is needed to control and clean it. **Data control and cleaning** in patient registries involves the process by which erroneous data is removed or fixed and missing data is filled. Three different phases in the cleaning process can be distinguished: screening, diagnosis and editing.
Storing and retrieval of data are among the IT services giving support to registry operations. Although storing data is a technical aspect - mainly relating to the informatics framework - data privacy is a major concern in European countries. Legal aspects, especially those related to cross-border use, are becoming important issue (security, access permission, anonymisation of stored personal data).
Registry **data analysis** is a highly relevant issue for a registry, taking into consideration the fact that the registry outcomes will depend on the results of the analysis. Ideally, a detailed data analysis plan should be established beforehand, but flexibility is needed to deal with situations that registry planners could not originally foresee. The data analysis plan needs to be executed in accordance with the

characteristics of the registry data, the appropriate statistical methods (descriptive tools, making inference, survival analysis) and other analytical considerations relating to potential sources of bias (selection bias, non-response bias, information bias and recall bias), confounding by indication and handling missing data.

Finally, the process of **data and information dissemination** has to be taken into consideration, focussing on all the interested public and stakeholders. Data should be disseminated in different ways, depending on the addressee of the data. Choosing the best way to do so will have an effect on the success of the registry.

The overarching processes concern the following factors:

The **assurance and assessment of data quality**, that lies on its dimensions: completeness, reliability, validity, timeliness, comparability, relevance and availability. All of these have to be assessed in order to improve the data quality. The mechanisms for doing so are reviewed.

The **evaluation and improvement** of a registry service is addressed, to assess the criteria and internal quality validation in registries. Methods and indicators are related to the type of registry.

Patient registry **governance** comprises the systems and procedures by which a registry is directed and managed. It refers to guidance and high-level decision making, including concept, funding, execution, and dissemination of information. The governance of the organisational structure (steering committee, scientific advisory board) and the establishment of the governance plan and responsibilities, duties, roles of the people in charge of the registry are also essential for a good quality service.

When running a registry, an audit (examination or review that establishes the extent to which a condition, process or performance conforms to the predetermined standards or criteria) is required to confirm that all the processes are well executed, according to the plan.

Running a registry also takes a good deal of IT. Some technical problems regarding the information system management, have to be taken into consideration, for instance: operations, service requests and incidents, problems, continuity, security services and process controls.

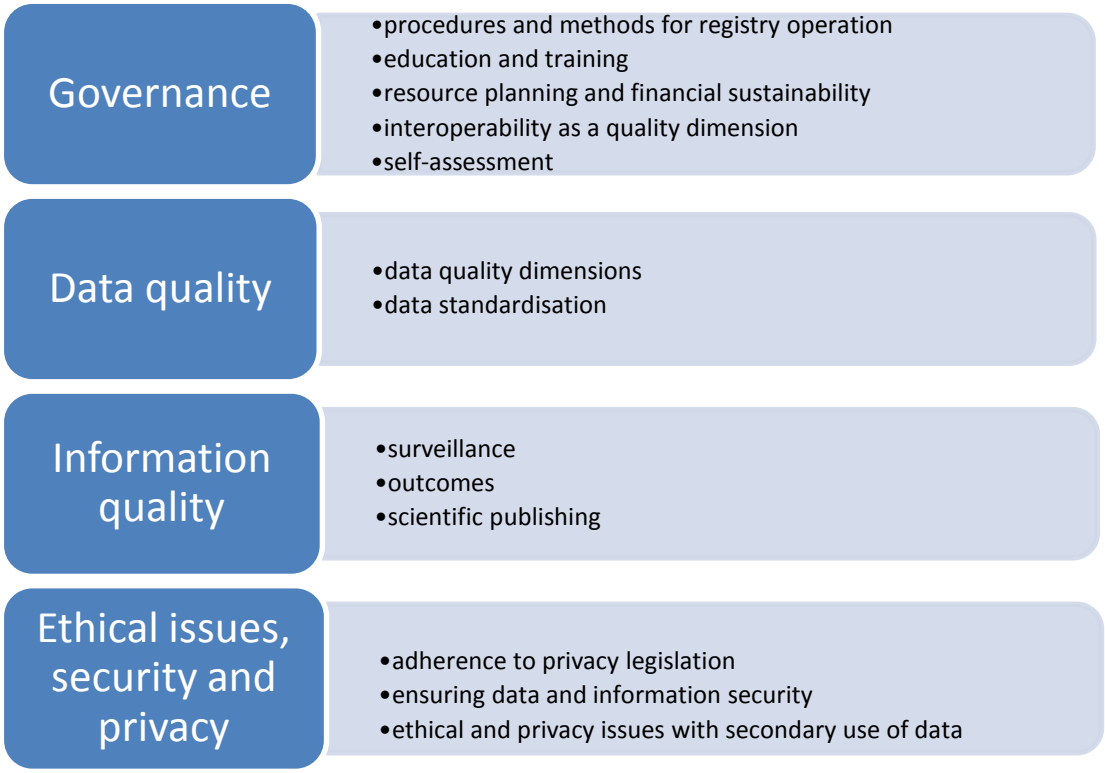
The chapter ends with two case studies: 1) REMRENAL, a regional renal registry, and 2) ENERCA (*European Network for Rare and Congenital Anaemias*) project registry.

8 QUALITY DIMENSIONS OF REGISTRIES

In broader terms, quality can be defined as “the standard of something as measured against other things of a similar kind; the degree of excellence of something” [14]. In that light, other quality dimensions apart from data quality can also (and should) be assessed. In this way, data quality remains the primary dimension within registry quality evaluation, but it is acknowledged that it is influenced by other identifiable registry features. Based on this rather holistic view and by conducting a literature review, PARENT JA has identified numerous “quality influencing factors” and categorised them into four groups, which should not be viewed separately. These are: 1) governance; 2) data quality; 3) information; 4) ethical issues, security and privacy (Figure 1).

It is useful to consider these categories when planning and evaluating registries, since they should, rounded all together, provide a rough estimate basis for assessing registry performance.

Figure 1 Quality dimensions of registries



In this chapter, each dimension is explored further and described in more detail. For the purpose of this summary, the data quality dimension is exposed as one of the most important quality dimensions since use of registries greatly depends on the quality of the data that they contain.

Data quality can be defined as the totality of features and characteristics of a data set that bear on its ability to satisfy the needs that result from the intended use of the data [15]. Requirements for data collection and quality assurance should be defined during the registry creation phase, and following the "collect once, use many times" rule of data collection and management, it is paramount that the data is of sufficient quality, as information and subsequent use for multiple potential purposes are all derived from that initial data.

8.1 Data quality dimensions and their assessment

In addition to a full understanding of study design and methodology, analysis of registry events and outputs requires an assessment of data quality. Determining the quality of data is possible by means of data assessment against a list of dimensions. These data quality dimensions can be defined as a “set of data quality attributes that represent a single aspect or construct of data quality” [16]. Through review of the literature, a large number of distinct data quality attributes that might determine usability have been identified. In attempting to describe data quality, it emerged that most of the data quality dimensions overlap and have different interpretations, often with ambiguous definitions or no

definitions at all, while the two most frequently cited were data ''accuracy'' and ''completeness''. A more detailed overview[7] of frequently cited dimensions is provided in Table 2.

Table 2 - Data quality dimensions

Data quality dimension	Description
accessibility	ease of access; awareness of data users of what data is being collected and knowing where it is located
accuracy	how closely the data correctly captures what it was designed to capture
coherence	internal consistency of the data; by the use of standard data concepts, classifications and target populations
comparability	consistency of data between organisations and over time allowing comparisons to be made
completeness	data collected matches the data set that was developed to describe a specific entity
interpretability	ease at which the user can understand the data
relevance	data meets the needs of users; relevance for decision makers
reliability	extent to which data is collected consistently over time and by different organisations either manually or electronically
timeliness	data is collected within a reasonable time period and is available within a reasonable timeframe to be used for intended purpose
usability	extent to which data can be accessed and understood
validity	data has been collected in accordance with rules or definitions applicable for that data

In summary, efforts should be made to create various relevant data quality dimension groups[8], depending on the type and objectives of the registry, and to devise methods and indicators for assessing data quality, so that a registry can use those methods to measure and gradually improve its data quality, potentially through a self-assessment tool for data quality.

Again, the importance of data quality should be highlighted as a pre-requisite for ensuring meaningful analyses and registry outputs.

8.2 Mode of data collection and impact on data quality

Considering data quality not in isolation but as part of a complex whole brings out another important and often neglected aspect that can influence data quality: the point where data is collected.
The quality of initial data input from clinicians and health practitioners can vary.

9 CHANGING AND STOPPING REGISTRIES

A registry is a living system that evolves over time. In order to remain or become more useful and successful a registry sometimes needs to be modified. In general, it is important that a registry is

[7] Overview adapted from: International Review of Data Quality. Dublin: HIQA, 2011 [17].
[8] Data quality dimensions specifically tailored to the registry type, or in other words, groups which contain quality dimensions of highest importance for the specified registry type.

flexible and adaptable, with a sense of continuous development. Regular checks and evaluations (e.g. internal or external reviews) of whether any of the registry's components needs to be modified are important factors that affect the sustainable success of a registry.

There are various reasons for a registry to undertake the modification or adaptation process. Unmet registry stakeholder needs, failure to meet certain standards, reducing the burden of the registry team or participants, new regulatory or legal requirements, innovations and changes in medicine and health care (i.e. new products, procedures, and services), innovations in information technology, or changes in the financing of the registry are just a few examples.

Some minor changes to a registry can be implemented more easily and quickly, but modifying a registry can also be a complex task that requires more effort, time and money. It is therefore highly recommended that a registry team wanting to modify a registry considers various elements in order to implement changes successfully and run the transition smoothly. The chapter highlights and addresses these points, including:

- rationale for the registry change
- scope of the registry modification
- planning and assessment of the feasibility of registry modification
- a team for transition
- reconsideration of the legal aspect
- notification protocol for registry stakeholders
- changes in technology, and training
- modification of the data elements,
- data mapping and data migration

Another issue that should certainly not be underestimated is the activity of stopping a registry. Planning for a registry stop is as important as planning a registry launch. When stopping a registry, it is essential that there is a clear decision on stopping, that a registry team establishes communication with data providers and decides what will happen with the registry data, taking into account the legal requirements for data maintenance/storage as well. This chapter introduces these important aspects and describes the process of stopping a registry.

10 RE-USE OF REGISTRY DATA

Re-use of information means cases in which information recorded for a given purpose is used for another one purpose. The fact that all information is purpose-dependent generates serious limitation of re-use, which of course does not mean that information cannot be used for any other purpose but for which it was originally recorded.

Registries are often and preferably realisations of information re-use. It is hard to justify to collect data just for registry purposes that are not relevant or not needed for clinical purposes. The primary reason for storing patient data is clinical need and **registries should be extracts of clinical records**.
To summarise these considerations, we may state the followings:

- Designing and operating registries should serve well defined purposes.
- The normal way of using registry data is to serve the defined purpose
- Re-use of registry data is using data for any other purpose than the originally planned ones

Certain questions emerge in the context of the Guidelines:

Why to re-use? In many health systems, a vast amount of information is collected and poorly utilised in practice. If re-use is possible, it has advantages over separate data collections for different purposes. Re-use is a much more cost-effective and easy way.

Is re-use possible? In spite of certain concerns or limitations, re-use is possible, although care must always be exercised. For example, data, collected originally for health care reimbursement can often be used for quality assessment or capacity planning. But we have to know, that using some data for financial purposes always induces some distortion. Indeed, all observations distort the phenomenon that we want to observe to some extent.

Re-use of clinical data in registries: it is a critical success factor for designing and implementing registries that the administrative burden of health care providers is minimised. Data collection systems should be automated as much as possible. Re-use of clinical data for registry (and other public health) purposes is usually an abstraction process based on some sort of knowledge.

There are several types of re-use of registry data: internal re-use, cross-registry comparison, re-use of aggregations vs. re-use of elementary data, comparison with information outside of the healthcare domain (e.g. environmental, economic, social etc. data), cross-border use for public health, cross-border use for research purposes.

In connection with the types of re-use of registries mentioned above, interoperability standards and approaches for data exchange should be followed with an emphasis on:

- Coding schemes, terminologies
- Standards (e.g. EN/ISO 13606, openEHR, HL7)
- Mapping between classification and coding systems
- Ontologies and data structures

11 REFERENCES

1. Gliklich R, Dreyer N, Leavy M, eds. Registries for Evaluating Patient Outcomes: A User's Guide. Third edition. Two volumes. Available from: http://www.effectivehealthcare.ahrq.gov/registries-guide-3.cfm
2. Newton J, Garner S. Disease Registers in England. A report commissioned by the Department of Health Policy Research Programme in support of the White Paper entitled Saving Lives: Our Healthier Nation. Institute of Health Sciences. University of Oxford. 2002.
3. IOM. Crossing the Quality Chasm: The IOM Health Care Quality Initiative. Available from: http://www.iom.edu/Global/News%20Announcements/Crossing-the-Quality-Chasm-The-IOM-Health-Care-Quality-Initiative.aspx Accessed: 27.5.2014.
4. eHealth Action Plan 2012-2020 – Innovative healthcare for the 21st century. Available from: http://ec.europa.eu/information_society/newsroom/cf/itemdetail.cfm?item_id=9156
5. eHealth European Interoperability Framework. Available form: http://ec.europa.eu/digital-agenda/en/news/ehealth-interoperability-framework-study
6. EIF Annex II. Available from: http://ec.europa.eu/isa/documents/isa_annex_ii_eif_en.pdf
7. PARENT (PAtient REgistries iNiTiative). PARENT Pilot Registry of Registries 2014. Available from: http://www.parent-ror.eu/#!state/list_all
8. Agency for Healthcare Research and Quality. RoPR - Registry of Patient Registries: U.S. Department of Health & Human Services; 2014. Available from: https://patientregistry.ahrq.gov/.
9. WHO. Training manual for community-based initiatives: A practical tool for trainers and trainees. 2006.
10. International Agency for Research on Cancer. Cancer Epidemiology – Principles and Methods.1999.
11. Polygenis D, ed. ISPOR Taxonomy of Patient Registries: Classification, Characteristics and Terms. Lawrenceville, NJ; 2013.
12. Gregg, M.B. Field Epidemiology. New York: Oxford University Press, 2002.
13. Australian commission on safety and quality in health care. Operating Principles and Technical Standards for Australian Clinical Quality Registries. 2008.
14. Oxford dictionary. Available from: http://www.oxforddictionaries.com/definition/english/quality.
15. Arts DGT, de Keizer NF, Scheffer G-J. Defining and Improving Data Quality in Medical Registries: A Literature Review, Case Study, and Generic Framework. Journal of the American Medical Informatics association. 2002; 9(6):600-611.
16. Wang, RY, Strong DM. Beyond accuracy: What data quality means to data consumers. Journal of Management Information Systems. 1996; 12(4):5-33.
17. The Health Information and Quality Authority. International Review of Data Quality. Available from: http://hiqa.ie/press-release/2011-04-28-international-review-data-quality