

Methodological guidelines and recommendations for efficient and rational governance of patient registries

Editors:
Metka Zaletel, Marcel Kralj

Methodological guidelines and recommendations for efficient and rational governance of patient registries

Project Title

Cross-border Patient Registries Initiative

Project Acronym

PARENT

PARENT Coordinator

Marija Magajne, Matic Meglič

Work Package Leader

Metka Zaletel

Deliverable 7

Version 1.0

Website

patientregistries.eu

Wiki version of the document available at:

parent-wiki.nijz.si



October 2015

Methodological guidelines and recommendations for efficient and rational governance of patient registries

Editors

Metka Zaletel, Marcel Kralj

Publisher

National Institute of Public Health, Trubarjeva 2, 1000 Ljubljana, Slovenia

Email: info@nijz.si

www.nijz.si

Ljubljana, 2015

ISBN 978-961-6911-75-7 (pdf)

Copyright © Authors and National Institute of Public Health, Slovenia

Electronic Publication

This publication arises from the Cross-border Patient Registries Initiative, a Joint Action which has received funding from the European Union through the Consumers, Health, Agriculture and Food Executive Agency of the European Commission, in the framework of the Health Programme 2008-2013. The European Commission is not responsible for the content of this publication. The sole responsibility for the publication lies with the authors, and the Executive Agency is not responsible for any use that may be made of the information contained herein.

CIP – Cataloguing in Publication

National and University Library, Ljubljana, Slovenia

614.253.8(083.86)(0.034.2)

METHODOLOGICAL guidelines and recommendations for efficient and rational governance of patient registries [Electronic source] / editors Metka Zaletel, Marcel Kralj. - El. book. - Ljubljana : National Institute of Public Health, 2015

ISBN 978-961-6911-75-7 (pdf)

1. Zaletel, Metka

281581568

Acknowledgements

We would like to begin by thanking all the authors for their hard work and enthusiasm they have brought to the topic Methodological guidelines and recommendations for efficient and rational governance of patient registries.

In addition, we would like to thank the reviewers: Gerold Labek, Kyriakos Souliotis and Sarunas Tarasevicius, as well as teams in partner institutions of PARENT Joint Action.

For the constructive comments we would like to thank Barbara Foley, Jacoline Bouvy, Mika Gissler, Saskia de Groot, Päivi Hämäläinen, Mirella Marlow and Erwin Eisinger. Our sincere thanks go also to all PARENT Joint Action workshops participants for their support and valuable insights.

Finally, we would like to thank the European Commission and participating Member States who have made PARENT Joint Action possible.

Table of Contents

1	INTRODUCTION	12
2	PATIENT REGISTRIES.....	14
2.1	Definition of a patient registry	15
2.2	Types of Patient Registries	17
2.2.1	Disease or Condition Registries	18
2.2.2	Product Registries.....	24
2.2.3	Health Services Registries.....	28
2.3	Diversity in Use of Patient Registries.....	31
2.4	Overview of European Registries	32
2.4.1	Member State level registries overview.....	33
2.5	Key issues arising within registries	37
3	INTEROPERABILITY	42
3.1	Introduction.....	43
3.1.1	Contexts.....	44
3.1.2	The PARENT Framework.....	48
3.2	Registry interoperability guidelines.....	48
3.2.1	General	48
3.2.2	The political (stakeholder) context.....	48
3.2.3	Legal interoperability level	51
3.2.4	Organizational/process interoperability level.....	52
3.2.5	Semantic interoperability level.....	53
3.2.6	Technical interoperability level	55
4	QUALITY DIMENSIONS OF REGISTRIES.....	58
4.1	Governance	59
4.1.1	Procedures and methods for registry operation.....	60
4.1.2	Education and training	60
4.1.3	Resource planning and financial sustainability	61
4.1.4	Interoperability as a quality dimension.....	61
4.1.5	Self-assessment	62
4.2	Data quality	62
4.2.1	Data quality dimensions and its assessment.....	62
4.2.2	Mode of data collection and impact on data quality	65
4.2.3	Improving data quality	65
4.3	Information quality.....	67
4.4	Confidentiality, security, privacy, ethical issues, secondary use of information	67
4.4.1	Privacy impact assessment (PIA) – a method to assess privacy.....	68
5	GENERAL REQUIREMENTS FOR CROSS-BORDER USE OF PATIENT REGISTRIES	71
5.1	Political context	72

5.2 Organisational aspects – Registries’ operations and procedures	73
5.3 Legal and ethical aspects.....	73
5.4 Semantic aspects	76
5.5 Technical aspects – Guaranteeing shareable data	77
5.6 Effective and sustainable solutions for cross-border registry collaboration	77
5.7 Health Data Resources in Europe: Mapping national strategies.....	77
6 CREATING A REGISTRY.....	83
6.1 Planning a registry	84
6.1.1 Defining the Purpose, Objectives and Outputs of the Registry.....	85
6.1.2 Data Considerations	88
6.1.3 Overview of the Current State and the Importance of Interoperability	90
6.1.4 Considering Legal Aspects and Confidentiality.....	92
6.1.5 Eliciting Expert Opinion & Generating an Advisory Board	93
6.1.6 Defining the Scope of the Registry & Building a Registry Development Team	94
6.1.7 Performing Stakeholder Engagement and Analysis	94
6.1.8 Re-defining the Scope of the Registry	96
6.1.9 Governance, Oversight and Registry Teams.....	97
6.1.10 Resource requirements	101
6.1.11 Funding Strategy.....	102
6.1.12 Risks and feasibility.....	102
6.1.13 Developing an Implementation Plan	102
6.2 Registry research design.....	107
6.2.1 The population covered by a registry	108
6.2.2 Anticipated size and duration.....	108
6.2.3 Registry dataset	109
6.2.4 Data collection procedure	110
6.2.5 Research-based registries - additional points to consider	111
6.3 Registry dataset	118
6.3.1 General principles for building a registry dataset	119
6.3.2 International coding systems, terminologies and common data sets	122
6.4 Data sources for registries.....	127
6.4.1 Definition of Primary and Secondary Data Sources	128
6.4.2 Identification of Available Sources	129
6.5 The role of information system methodologies and techniques in the phase of patient registry creation	131
6.5.1 Why modelling?	132
6.5.2 The role of IS expert (system analyst, process modeller, health informatics expert etc.)	133
6.5.3 Short description of frequently used modelling techniques and notations.....	135
6.5.4 UML	135
6.5.5 Process Modelling	138

6.5.6 Data modelling (using E-R diagram)	147
6.5.7 Entity-Attribute-Value Data Model in Medical databases	153
6.5.8 Temporal modelling in medical databases.....	155
6.5.9 Knowledge management using archetypes	158
6.5.10 The importance of Terminologies and Code lists	167
7 PATIENT REGISTRY INFORMATION SYSTEM DEVELOPMENT AND IMPLEMENTATION	170
7.1 Computer based Patient Registry Information System	171
7.2 Development of Registry Information System	172
7.2.1 Important role of users.....	173
7.2.2 Software testing	173
7.2.3 Training.....	174
7.3 Different options to obtain the Registry system	175
8 RUNNING A REGISTRY	177
8.1 Sequential Processes	178
8.1.1 Collecting data	178
8.1.2 Data Linkage	180
8.1.3 Controlling and Cleaning the Data.....	182
8.1.4 Storing Data	183
8.1.5 Analysis of Registry Data	184
8.1.6 Data Dissemination.....	189
8.2 Overarching Processes	190
8.2.1 Data Quality Assurance	190
8.2.2 Data Quality Assessment.....	191
8.2.3 Evaluation and Improvement of Registry Service	191
8.2.4 Governance	192
8.2.5 Auditing	193
8.2.6 Continuous Development.....	193
8.2.7 Information System Management	194
9 CHANGING AND STOPPING REGISTRIES.....	198
9.1 Changing an existing registry.....	199
9.2 Time to stop? - Stopping a registry.....	202
10 RE-USE OF REGISTRY DATA	205
10.1 Background.....	206
10.1.1 Definition of re-use.....	206
10.1.2 Re-use in the context of patient registries	206
10.2 Why to re-use?	207
10.3 Is re-use possible?	207
10.4 Re-use of data.....	208
10.4.1 Re-use of clinical data in registries	208
10.4.2 Re-use of spatial data	208

10.5 Types of re-use of registry data	208
10.5.1 Internal re-use	208
10.5.2 International comparison (same purpose, different context).....	208
10.5.3 Cross-registry comparison (correlation between diseases)	209
10.5.4 Comparison with information outside the health domain (e.g. environmental, economic, social etc. data).....	209
10.6 Re-use of aggregations vs. re-use of elementary data	209
10.7 Definition of Possible Types of Data.....	210
10.7.1 Aggregated Data (Indicator Compilation)	210
10.7.2 Anonymised Data	211
10.7.3 Pseudo-anonymised Data.....	212
10.7.4 Personal Data.....	213
10.8 Cross-border Use for Public Health	213
10.8.1 Cross-country Data Comparison, Surveillance	213
10.8.2 Outbreak Alerting and Communication	214
10.8.3 Bioterrorism Threat	214
10.8.4 Identification of Best and Cost-effective Practices.....	214
10.8.5 Referral to Services, Establishing New Services	214
10.8.6 Public Health Research	214
10.9 Cross-border Use for Research Purposes	215
10.9.1 Issues	215
10.9.2 Risk Factor Studies.....	215
10.9.3 Genetic Research.....	215
10.9.4 Clinical and Therapeutic Research.....	215
10.9.5 Some additional information.....	215
10.10 Compatibility, comparability and interoperability	216
10.10.1 Data compatibility	216
10.10.2 Comparability	217
10.10.3 Interoperability.....	218
10.11 Interoperability Standards and Approaches for Data Exchange	218
10.11.1 General Concept.....	218
10.11.2 eHealth standards	219
10.11.3 Coding schemes, terminologies.....	220
10.11.4 Ontologies and data structures	222
10.11.5 Mobile health delivery, personalized medicine, and social media applications.....	222
10.12 Problem with populations	223
10.12.1 Definition of population	223
10.12.2 Inclusion and exclusion criteria	223
10.12.3 Mobility.....	223
10.12.4 Socio-demographic, genetic factors.....	223

10.13	Examples of legal frameworks for data protection and data sharing	224
10.13.1	Policy on data submission, access, and use of data within TESSy.....	224
10.13.2	European Commission’s proposal for a General Data Protection Regulation	224
10.13.3	European Data Protection Board, General Data Protection Regulation	225
10.13.4	HIPAA Privacy and Security Rules for Public Health Data Exchange	225
11	APPENDICES	227

List of Abbreviations

ADR	Adverse drug reaction
AE	Adverse event
AHRQ	Agency for Healthcare Research and Quality
AIDS	Acquired immune deficiency syndrome
ATC	Anatomical Therapeutic Chemical Classification System
BPD	Business Process Diagram
BPM	Business Process Management
BPM	Business Process Modelling
BPMN	Business Process Model and Notation
BRIDG	Biomedical Research Integrated Domain Group
CBHD	Cross Border Healthcare Directive
C-DISC	Clinical Data Interchange Standards Consortium
CEF	Connecting Europe Facilities
CIHI	Canadian Institute for Health Information
COPD	Chronic Obstructive Pulmonary Disease
CRF	Case Report Form
CTS2	Common Terminology Services 2
DB	Database
DocDat	Directory of Clinical Databases
DQ	Data Quality
DRG	Diagnosis-related group
EAACI	Academy of Allergy and Clinical Immunology
EAR	European Arthroplasty Register
ECHAlliance	European Connected Health Alliance
ECIS	European Cancer Information System
EDA	Event-driven architecture
EDC	Electronic data capture
EEA	European economic area
EFORT	European Federation on National Associations of Orthopaedics and Traumatology
eHGI	eHealth Governance Initiative
EHR	Electronic health record
EHS	European Hernia Society
EIF	European Interoperability Framework
EMA	European medicine agency
ENCR	European Network of Cancer Registries
EPAAC	European Partnership for Action Against Cancer
EPC	Event-driven process chain
EPIRARE	European Platform for Rare Disease Registries
epSOS	Smart Open Services for European patients
EU	European Union
EUBIROD	European Best Information through Regional Outcomes in Diabetes
EUCERD	European Union Committee of Experts on Rare Diseases

EuraHS	European Registry for Abdominal Wall Hernias
EUREMS	European Register for Multiple Sclerosis
FDA	U.S. Food and Drug Administration
GP	General practitioner
GVP	Good pharmacovigilance practice
HIQA	Health Information and Quality Authority of Ireland
HL7	Health Level Seven
HTA	Health technology assessment
HTML	Hypertext Markup Language
i2b2	Informatics for Integrating Biology and the Bedside
ICD	International Classification of Diseases
ICD10	International Classification of Diseases, 10th revision
IEEE	Institute of Electrical and Electronics Engineers
IHE	Integrating the Healthcare Enterprise
IHTSDO	International Health Terminology Standards Development Organisation
IS	Information System
ISO	International Organization for Standardization
ISPOR	International Society for Pharmacoeconomics and Outcomes Research
IT	Information technology
LIBE Committee	The Committee on Civil Liberties, Justice and Home Affairs
LOINC	Logical Observation Identifiers Names and Codes
MDR	Metadata repository
mHealth	Mobile health
MS	Member States (EU)
NCCP	National Cancer Control Programme
NGO	Non-governmental organization
OECD	Organisation for Economic Co-operation and Development
OSSE	Open Source-Registersystem für Seltene Erkrankungen in der EU
OWL	Web Ontology Language
PARENT	PAtient REgistries iNiTiative
PIA	Privacy impact assessment
PR	Patient Registry
RD	Rare disease
RoPR	Registry of Patient Registries
RoR	Registry of registries
SHAR	Swedish Hip Arthroplasty Register
SNOMED-CT	Systematized Nomenclature of Medicine - Clinical Terms
SOA	Services oriented architecture
SW	Software
UML	Unified Modelling Language
US	United States
WHO	World Health Organization
WP	Work package
XML	Extensible Markup Language

1 INTRODUCTION

Marija Magajne, Matic Meglič

Policy makers, researchers, healthcare professionals and many others are daily facing the challenge of how to prioritise their activities and actions in different areas of their responsibilities. In the health area, their decisions are often influenced by the reality of the needs of an aging population on the one hand and restricted resources on the other. They can only make as good informed decisions as the quality and availability of data they use in the process.

Patient registries, which were designed for patients with sharing characteristics (diagnosis, implanted device, therapy, risk of developing a disease), have for decades served as an important source of the data needed to assess clinical performance, provide health technology assessment or assess policy implications on a local, regional, national and in some cases international level. As a result, hundreds of registries have been set up, ranging from paper based spreadsheets in a physician's office to international rare disease initiatives coupling clinical and genetic data as well as bio-banks. In the last fifteen years information technology has given us an opportunity to greatly redesign the way we make informed decisions about individual patients as well as entire populations by, among other things, enabling clinicians to collect, share, compare and analyse large amounts of patient data.

Where we still fall short is in harnessing information and new knowledge from the wealth of data across registries – be it from one country to another or between/across registries with overlapping characteristics or patient pools. Researchers, HTA organisations and policy creators are wasting valuable time acquiring data from different sources and painstakingly pairing them in order to extract new knowledge. Also, setting up a new patient registry sets the holder on a high risk journey where a number of decisions need to be made about methodologies, processes, technologies and governance of the registry with little available guidance.

To provide guidance and tools on an EU level to solve the above issues is likely to be the largest near-term opportunity towards data and information driven public health decision making, policy creation and research.

The PARENT JA team is proud to present the Guidelines, which were created to provide practical and 'hands on' advice to set up and manage patient registries as well as to enable secondary use of data for public health policy and research. We hope the Guidelines will make life easier for those setting up new registries or redesigning already functioning registries and those exchanging data across registries.

Getting to this point has been a challenging journey but we have made it. It is a result of the commitment and passion of a number of experts from across the EU as well as continuous support

from numerous EU bodies and projects, and the US Agency for Healthcare Research and Quality – all of whom generously contributed their knowledge and insights into the topic.

While the Guidelines are a first step towards greater interoperability of patient registries, a number of exciting and complex challenges still lie ahead, requiring continuous efforts to ensure that we utilise the full value of patient registries.

May the Guidelines serve you well.

2 PATIENT REGISTRIES

*Marko Brkić, Borna Pleše, Vanja Pajić, Mladen Kostešić, Ranko Stevanović,
Tamara Poljičanin, Ivan Pristaš, Metka Zaletel, Marcel Kralj*

Patient registries collect, analyze and disseminate data and information on a group of people defined by a particular disease, condition, exposure or health-related service.

Key principles:

- Registries serve a predetermined scientific, clinical or/and public health (policy) purpose - the improvement of patient care and healthcare planning as well as social, economic and quality of life outcomes and other health indicators.
- According to how their populations are defined, they can focus on a disease/condition, medical product or health service.
- European registry landscape is a collection of divergent registries often built for a single purpose and with a limited user profile operating under different legal frameworks and with little standardization in interoperability and governance rules.
- European registries face the issues of:
 - unstable funding,
 - legal ambiguity,
 - unclear stakeholder roles,
 - predominantly paper-based data collection,
 - lack of awareness of existing standards and standard processes,
 - compromised data quality,
 - lack of registry transparency and openness that support data access for research purposes
 - insufficient data dissemination

2.1 Definition of a patient registry

In the field of health, several definitions of the term registry or register¹ have been provided. In 1949, Bellows (6) defined register as “system of recording frequently used in the general field of public health which serves as a device for the administration of programs concerned with the long-term care, follow-up or observation of individual cases.” In 1974, the WHO (5) defined a register as a “file of documents containing uniform information about individual persons, collected in a systematic and comprehensive way, in order to serve a predetermined purpose.” Another definition was provided by Solomon et al. (8) who defined a registry as a “database of identifiable persons containing a clearly defined set of health and demographic data collected for a specific public health purpose.” A slightly different definition of a registry is proposed by ISPOR (3), which describes a registry as a “prospective observational study of subjects with certain shared characteristics, which collects ongoing and supporting data over time on well-defined outcomes of interest for analysis and reporting.” A more specific definition is provided by the US National Committee on Vital and Health Statistics (1), which defines a registry as “an organized system for the collection, storage, retrieval, analysis, and dissemination of information on individual persons who have either a particular disease, a condition (e.g., a risk factor) that predisposes (them) to the occurrence of a health-related event, or prior exposure to substances (or circumstances) known or suspected to cause adverse health effects.” Despite variations in definition, it is clear that a registry involves a long-term, systematic and organized process of collecting data, which is driven by specific, predefined aims.

Nowadays the term “patient registry” is often used in the health domain. The use of the term “patient” in combination with “registry” (i.e. patient registry) is mainly used to distinguish the focus of the dataset on health information (9). The AHRQ (2) provides the definition of the patient registry, which is “an organized system that uses observational study methods to collect uniform data (clinical and other) to evaluate specified outcomes for a population defined by a particular disease, condition, or exposure, and that serves one or more predetermined scientific, clinical, or policy purposes”.

For the purpose of these guidelines, patient registry is defined as...

... an organized system that collects, analyses, and disseminates the data and information on a group of people defined by a particular disease, condition, exposure, or health-related service, and that serves a predetermined scientific, clinical or/and public health (policy) purposes.

¹ The terms “register” and “registry” are often used interchangeably. However some authors differentiate between these two terms taking the position that the “registry” is the organisation and process that supports one or a number of individual “registers” (4). In this paper we are using the term “registry”, the only exceptions are registries’ official names.

References

1. Available at: Frequently Asked Questions about Medical and Public Health Registries. The National Committee on Vital and Health Statistics <http://ncvhs.hhs.gov/9701138b.htm>.
2. Gliklich RE, Dreyer NA, eds. Registries for evaluating patient outcomes: A User's Guide. 3rd ed. 2014.
3. Polygenis D, ed. ISPOR Taxonomy of Patient Registries: Classification, Characteristics and Terms. Lawrenceville, NJ; 2013.
4. 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
5. Eileen M. Brooke, (WHO). The current and future use of registers in health information systems. Geneva, World Health Organization. 1974. Available from: <https://extranet.who.int/iris/restricted/handle/10665/36936>
6. Bellows, Marjorie T. Public Health Reports, Vol. 64, No. 36, pp. 1148-1158. 1949. Available from: <http://www.jstor.org/discover/10.2307/4587080?sid=21105208469701&uid=70&uid=4&uid=3739008&uid=2&uid=2129>
7. Solomon, D. J., R C Henry, J G Hogan, G H Van Amburg, and J Taylor. Evaluation and implementation of public health registries. Public Health Rep. 1991 Mar-Apr; 106(2): 142–150. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1580226/pdf/pubhealthrep00191-0040.pdf>
8. Workman TA. Engaging Patients in Information Sharing and Data Collection: The Role of Patient - Powered Registries and Research Networks. AHRQ Community Forum White Paper. AHRQ Publication No. 13-EHC124-EF. Rockville, MD: Agency for Healthcare Research and Quality; September 2013. Available from: <http://www.ncbi.nlm.nih.gov/books/NBK164513/pdf/TOC.pdf>

2.2 Types of Patient Registries

Registries should be designed and evaluated with respect to their intended purpose(s) which can be broadly described in terms of patient outcomes. Some of the major general purposes for establishing and running a patient registry are to **describe the natural history of disease**, to **determine clinical and/or cost-effectiveness**, to **assess safety or harm**, and to **measure quality of care**, as well as to serve **public health surveillance and disease control**. In broad terms, patient registries should contribute to the improvement of patient care and healthcare planning as well as social, economical and quality of life outcomes and other health indicators (e.g. access to healthcare, health status, subjective and objective quality, health financing etc.). By following patients in terms of time and location, medium- and long-term outcomes can be observed. A fine differentiation of the types, sub-types, main and secondary purposes of each patient registry is essential. For example, a diabetes registry compared to a surgical procedure registry shares many common datasets that are achieving completely different purposes. Although the logic of tracking how patients progress over time and factors that contribute to outcomes apply to both, there is a clear difference between the two registries used as an example here, as eligibility is characterised by a diagnosis and by an intervention, respectively.

The majority of patient registries can be divided into three general categories with multiple subcategories and combinations. These categories include observational studies in which the patient has a particular disease or condition, has had exposure to a product or service, or various combinations of these.

The multitude of possible combinations of categories and subcategories can sometimes lead to overlaps in many registries and difficulties in determining the taxonomic position of a particular registry (an example is a registry for treated drug addicts primarily a disease, product or service registry, or a mixture of equally important purposes: disease surveillance, outcomes, natural history of disease, national intervention programmes evaluation). Furthermore, in some countries a very clearly defined chronic disease registry (such as a cancer registry) very often serves many secondary purposes, some of which could eventually become its primary purposes. Therefore, in order to establish an appropriate data exchange (sharing) framework for secondary data use in particular (i.e. research questions), an extensive in-depth context analysis of each registry's content unit (data set, data element with properties and classes, value domains and property) should be performed. Such analysis would enable a correct interpretation of the results and a transparent disclosure of methodological restrictions.

Related to this, one of the most important quality indicators of patient registries is the amount and frequency of registry-related scientific publishing (meta-analysis and/or systematic review-like approach). See subchapter [4.3](#), [5](#) and [8.1.6](#).

With the help of information gathered through literature review, as well as with the insights gathered through the construction of the questionnaire and subsequent survey of registries for the RoR pilot, and with concern for the above stated complexity of taxonomy of registries, a multi-level classification of patient registries is offered² ([Table 2.1](#)).

Registries are classified according to how their populations are defined. For example, product registries include patients who have been exposed to biopharmaceutical products, medical devices or

² This classification is by no means definite or indisputable but subject to change and modification.

diagnostic/therapeutic equipment. Health services registries consist of patients who have had a common procedure, clinical encounter, or hospitalization. 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 a disability (1).

Table 2.1: Patient registry classification

Category	Diseases and conditions	Products	Services, events
Object type	chronic, acute communicable, rare diseases, disabilities, cause of death	medicines, devices, equipment	diagnostic, curative, preventive, discharges, births, abortions
Purposes / objectives (primary and secondary)	disease surveillance, control, natural course of disease	post-market surveillance	intervention evaluation, quality of care
	health outcomes (objective, patient reported)		
	effectiveness (clinical, comparative, financial)		
	safety and harm (HTA, vigilance)		
	intervention (planning, guidelines, reminders)		
Coverage (geographical and organizational)	health care unit (GP, hospital)		
	local (counties, districts, insurers, professional associations, NGOs)		
	national (MS, non-MS)		
	international (regional, EU, European region, global)		
Population definition	population (geographically based) ³		
	population based (exposition dependent) ⁴		
Observation unit	patient (user, client, insured party)		
	person with a characteristic of observation	person related device, equipment item	person related event (birth, death, service)

2.2.1 Disease or Condition Registries

The main inclusion criterion which disease or condition registries use 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, chronic condition such as disability) or for a more limited amount of time (e.g. short-term infectious disease). The disease registry could be hospital/clinic-based or population based. The

³ A “population registry” is a registry that intends to cover all residents in a given geographic area within a given time period. The coverage of the specific registry may, however, be incomplete, but it is nevertheless a population registry if the aim is to include all the individuals in the target population. A population is defined by geographical boundaries, but usually only residents (or citizens) within a given time period are included in the definition (38).

⁴ The term “population-based registry” should be used when all persons with a given trait, exposure or event, are intended to be included in the registry. If the registry includes everyone in the population (even the oldest), it becomes a population registry. Intention rather than performance defines the terms. A population-based disease registry aims at including everyone with the disease in the population, be it self-reported, clinically diagnosed or detected at screening. Population and population-based registries may be further classified as of good or bad quality depending on coverage or other characteristics (39).

former is used for a specific disease irrespective of the location of the case. Alternatively, a population based registry is used to compile information on specified diseases by region, community, and state in which they are diagnosed. The aims of disease or condition registries are most often primarily descriptive, such as describing the typical clinical features of individuals with a disease, variations in phenotype, and the clinical progression of the disease over time (i.e. natural course of the disease). The value of disease registries is increasingly recognized as they are able to provide historically comparable data and long-term evaluation, potentially serving as an addition to randomized clinical trials, and thus providing insights about real-sites outcomes that could not be addressed in the limited controlled studies. These registries become even more important to regulators (and other parties involved) when the disease cases are rare or require highly specialised health intervention. Here registries may be the only means by which data can be obtained.

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 (1).

As an example of an EU project/initiative concerning improving disease registries in terms of defining purposes, legal context, semantic and technical aspects, **EUBIROD** (“European Best Information through Regional Outcomes in Diabetes”) (www.eubirod.eu) is mentioned here. The project aims at sharing knowledge about prevention, treatment and patient care. Although there is a large amount of data and reports available, the information on diabetes in Europe is scattered and under-utilized. For this reason, the objective of the EUBIROD project was to improve information supplied to the public and formulate appropriate strategies, policies and actions and targeting appropriate sustainable coordination, in the area of health information, collection of data and information, comparability issues, exchange of data and information within and between Member States, continuing development of databases, analyses, and wider dissemination of information, and in fact to build a common European infrastructure for standardized information exchange in diabetes care. The main outcome of the project is a permanent and sustainable online standardised exchange of data and knowledge between EU countries (40). Production of information is primarily enabled through the use of a common dataset⁵, automatically achieving results that can later be harmonised to produce global indicators. Overall, EUBIROD can serve as a good example and model to be re-used for other chronic diseases as well (2).

There is also the European Academy of Allergy and Clinical Immunology (**EAACI**) (www.eaaci.org) as the next example of making efforts in improving disease registries. It is an association of clinicians, researchers and allied health professionals, dedicated to improving the health of people affected by allergic diseases.

The EAACI project also has goals to help standardization of data collection on allergic diseases, diagnosis and treatment and ultimately improve allergic disease and allergen exposure management. EAACI overall project objective is to provide a platform for the establishment of allergic disease registries across EU country borders to develop suitable monitoring tools for use in both clinical practice and research. The initial A-reg project is focused on two national allergic disease registries that are planned to grow into a pan-European registry, namely anaphylaxis and drug allergy. The main advantage of starting a registry in several European reference centres at the same time is that the

⁵ EUBIROD Deliverable D5.1: Common dataset. Available at: www.eubirod.eu/documents/downloads/D5_1_Common_Dataset.pdf

same methodology (data collection, software use, data analysis and ethics) ensures direct comparability (see chapter [4.2.1](#), [5](#) and [10.10.2](#)) from the start.

Since PARENT's main aim is to support EU MS in developing comparable and coherent patient registries, EAACI recognized this effort and has joined forces with the PARENT project as an official Partner Organization (3, 4).

Regarding cancer, EU policymaking institutions of the EU identified cancer control as a major public health priority, and consequently many EU projects/initiative were started. During 2009-2013, the European Partnership for Action Against Cancer (**EPAAC**) (www.epaac.eu) was established. It conceived a framework for identifying and sharing information, capacity and expertise in cancer prevention and control, in order to avoid scattered actions and the duplication of efforts. The main objective was to assist countries in developing National Cancer Control Programmes (NCCPs), but also includes goals on health promotion and prevention regarding cancer, screening and early diagnosis, research support, and mapping the existence of various data and information sources for cancer in Europe as well as checking the availability and quality of these data (5). Given the importance of cancer registries, much effort has been made to monitor and improve the quality, type and coverage of the information they gather. The European Network of Cancer Registries (**ENCR**) has the goal to enhance comparability of cancer incidence data, promote cancer registration in the European region, and foster the use of cancer information for research and planning. Today, more than 200 cancer registries are active under ENCR in Europe. Data collection systems in the EU reflect the specific organisation of national health systems, and barriers in data access persist. The move from the national to the European scale is still difficult as not all indicators are comparable across the EU. Registries presently provide most epidemiologic data on cancer, yet they are underfunded, mostly understaffed, struggling with national and European laws on protection data, or launched without proper planning (6).

In the area of cancer control, information and data are precious resources for researchers, health professionals and policymakers alike. Potential advantages in the cross-border exchange of cancer data are numerous, but achieving this goal is by no means simple. Cancer registries, being the main repository of data, vary widely in terms of geographical coverage and data quality.

The EPAAC project gathered insights about these issues and has given attention to the need to create an integrated and comprehensive European Cancer Information System (**ECIS**). The main tasks of an ECIS should not imply collection of new data, but rather reorganisation and better coordination of existing activities. Five main types of tasks which should be carried out under ECIS, have been identified: data management (each dataset flowing into ECIS must be organised according to a unique and coherent structure); data quality control (continuous improvement of quality and data standardisation as the only way for obtaining reliable data; datasets organisation (a user-friendly pathway should be implemented to structurally connect different datasets) (such cancer incidence and risk factors distribution across populations); data analysis (a plan of analysis for the main outcomes should be systematically and periodically laid down); information sharing (the ECIS would be a key epidemiologic infrastructure for the European Research Area and the results should be disseminated through general and specialised publications, press, leaflets, and web-based tools) (6).

When discussing disease or condition registries, rare disease registries are given a special overview, due to their specificity. By EU definition (7), a disease or disorder is defined as rare when it affects fewer than 5 individuals in every 10.000 citizens. Yet, because there are so many different RDs – between 6.000 and 8.000 – taken together they affect a significant proportion of the population.

Between 30 and 40 million people in the EU, many of whom are children, suffer from rare diseases. Most rare diseases have genetic origins while others are the result of infections, allergies and environmental causes. They are usually chronically debilitating or even life-threatening.

Just to list a few examples, there are registries for: Niemann-Pick disease (8), Fabry disease (9) and organic acidurias and urea cycle defects (10). A common aim of rare disease registries is to contribute to a better understanding of the natural course/history of rare diseases, through pooling cases of rare diseases, and studying their outcomes. Additional objectives of rare disease registries are to connect affected patients⁶, families and clinicians, and to support research on various (genetic, molecular, physiological) bases of rare diseases.

In the case of rare disease registries and due to low individual prevalence and the scarcity of information, knowledge and experience related to each rare disease, research is often conducted on the widest geographic scope possible (i.e. multi-nationally and/or across the continent), as the benefits of international collaboration, sharing efficiencies and maximization of limited resources should be most obvious here. Also, when resources are combined, identifying standards (i.e. common data elements) becomes more important to allow data to be compared and shared across registries.

Considering the specific nature of rare disease registries another thing may come to mind – creating a single global registry for each disease (or a certain group of diseases). That however is not always feasible, for a multitude of practical reasons and, most importantly, a single global registry would not always be in the best interests of researchers.

At EU level, much is being done to increase research, funding, and public awareness of RD (rare diseases).

To aid the EC with the preparation and implementation of Community activities in the field of rare diseases, The European Union Committee of Experts on Rare Diseases (**EUCERD**) (www.eucerd.eu) was established.

The EUCERD issued Recommendations on national/regional RD patient registration and data collection, which summarize the guiding principles that future actions on RD registration will rely upon and which harmonisation and standardisation procedures should be based across national and regional registries in Europe.⁷

Through project initiatives EUCERD is performing additional tasks, divided into five main areas:

- 1) The implementation of plans and strategies for rare diseases at national level;
- 2) The standardisation of rare disease nomenclature at international level;
- 3) Mapping the provision of specialised social services and integration of rare diseases into mainstream social policies and services;
- 4) The leveraging of the value of EU networking for improving the quality of care for rare diseases;
- 5) The integration of RD initiatives across thematic areas and across Member States (12).

⁶ EURORDIS (www.eurordis.org), as a non-governmental patient-driven alliance of patient organisations, is also bridging the gap between patients, addresses their needs and is active in promoting health policies and services and research policies and actions related to RD.

⁷ List of core recommendations is available at:

www.eucerd.eu/wp-content/uploads/2013/06/EUCERD_Recommendations_RDRegistryDataCollection_adopted.pdf

Orphanet (www.orpha.net/consor/cgi-bin/index.php) is another initiative related to RD, and is considered here a good practice. It is a reference portal and database for information on rare diseases and orphan drugs, run by a large consortium of European partners, with an aim to help improve the diagnosis, care and treatment of patients with rare diseases. Some of Orphanet's services include: an inventory of RD and its classification; an encyclopaedia of RD; a list of European RD registries (13). One of the benefits of the listed services is assistance in identification of potential data sources and collaborators.

EPIRARE (www.epirare.eu) (The European Platform for Rare Disease Registries) project is another important action in the RD field on the EU level. Its wide-ranging mission includes several areas such as: to provide RD methods and guides for EU researchers and policy makers, while also aimed at agreeing on a common RD data set, disease-specific data collection and data validation, simultaneously addressing legal and ethical issues associated with the registration of RD projects. In order to accomplish these objectives EPIRARE has, among other things, conceived a central website platform which would share information and resources (data repository function and predefined output production), and hence increase the sustainability, networking and interoperability of registries, promote the use of standards and of registry quality procedures (common data set and quality assurance function) and provide an effective way of disseminating the results.⁸ EPIRARE has produced guidelines for data sources and quality⁹ and by working on the existing registries it is attempting to formulate the core data elements, which then might be shared in a useful manner within the registry platform. The types of datasets being studied are: a minimum set of common data elements to be collected by all registries (necessary to interlink registries and to selectively extract basic data), other purpose-specific sets of common data elements (selected depending on the predefined outputs to be achieved by the platform), and project-specific sets of data elements (agreed by registries collaborating in ad hoc studies and/or in research on specific diseases).¹⁰

The PARENT project and its Registry of Registries (RoR) component is, although not RD specific, sharing several common goals with the EPIRARE project and is also implementing a cross border platform to support research in various ways.

Another example in the rare disease category is The European Register for Multiple Sclerosis (**EUReMS**) (www.eurems.eu). European MS Platform (EMSP) is developing this tool to assess, compare and enhance the status of people with MS throughout the EU, enabling better data for better outcomes. With regard to persons with multiple sclerosis, across European countries there is currently a widely recognized lack of data at EU and national level on treatment and care for people with multiple sclerosis. A comprehensive approach to data collection in MS is hence needed in addressing these issues, building on existing systems for MS data collection, but bridging their gaps and limitations by incorporating additional purposes, standardized methodological strategies and ensuring its validity across different European populations. Such an effort to provide a cross-border infrastructure for data collection, analysis, interpretation and dissemination of results in the MS field should be built on existing national/regional data collections, registries or cohorts and using the expertise of clinicians, researchers and patient organizations involved. As is the case with other similar cross-border

⁸ EPIRARE Deliverable D5: Delivering a European Platform for Rare Disease Registries. Available at:

www.epirare.eu/_download/del/D5_DevelopingaEuropeanPlatformforRareDiseaseRegistries%20FINAL.pdf

⁹ EPIRARE Deliverable D4: Guidelines for data sources and quality for RD Registries in Europe. Available at:

www.epirare.eu/_download/del/D4_GuidelinesfordatasourcesandqualityforRDRegistriesinEurope.pdf

¹⁰ EPIRARE Deliverable D9.3: Common Data Set and disease-, treatment and other specific modules. Available at: www.epirare.eu/_download/del/D9.3_ProposalforCDE_FINAL.pdf

initiatives, interoperability (presented and discussed in chapter 3) is the key enabler here and the prerequisite for completing such objectives.

The **OSSE** project (*Open Source-Registersystem für Seltene Erkrankungen in der EU* / Open Source Registry System for Rare Diseases in the EU) (www.unimedizin-mainz.de/imbei/informatik/ag-verbundforschung/osse.html), funded by the German Federal Ministry of Health, provides a reusable software for RD registries. The aim of the project is to provide patient organizations, physicians, scientists and other parties with open-source software for the creation of patient registries. As a result, the national registry landscape would be improved to comply with European principles regarding minimum data sets, data quality etc. (summarized in the EUCERD Recommendation on RD registries mentioned above) along with achieving necessary interoperability to allow a federation of registries on a national and international level (e.g. distributed searches designed to comply with data protection requirements and preserve data sovereignty).

OSSE's backbone is a registry toolkit that enables scientists to build a registry for a specific rare disease even without special IT knowledge. A registry editor allows for the definition of forms for longitudinal and basic medical data and of the corresponding data schema, including an ID management/pseudonymization service. ID Registry fields (including, inter alia, data type, ranges, measurement units and value sets) are defined within the metadata repository (MDR) which is another integral part of the OSSE architecture, providing semantic interoperability and data quality. It is envisaged that all harmonized data sets for rare diseases would be available through the MDR. Also, each user of the OSSE registry toolkit should register with a registry of registries (RoR). Exchanging data among (national and regional) rare disease registries on the OSSE architecture is achieved taking into account data ownership and privacy aspects, through a search function with specified search queries based on the existing MDR. Depending on the search exposé which contains the description of the research question along with contact information, the data owner decides if and what to reply. Also, the OSSE architecture is not restricted to a single registry software solution but also enables integration of registries built on different software.

Another initiative which aims to develop a global registry or registries for a certain rare disease or diseases is **TREAT-NMD** (www.treat-nmd.eu). It is a network for the neuromuscular field that provides an infrastructure to ensure that the most promising new therapies reach patients as quickly as possible. When a clinical trial is being planned, it is very important that patients suitable for that trial can be found and contacted quickly and the best way of ensuring this can happen is to make sure that patients' details are all collected together in a single database or "registry". That registry then contains all the information that researchers will need, including each patient's particular genetic defect and other key information about their disease.

The TREAT-NMD network is creating this kind of registry in countries across Europe and is also linking with other national registry efforts worldwide. The national registries all feed into a single global registry which combines the information from each of the national registries (with a pre-agreed internationally mandatory dataset), and this ensures that patients who register in their national registry can be contacted if their profile fits a clinical trial. In addition, these registries can help researchers to answer questions such as how common the individual diseases are across the world and will support other activities to improve patient care, such as the assessment of care standards in different countries. The network has also, issued a registries tool kit as a useful concise guide for creating a registry (be it general or NMD-specific)¹¹. Some benefits of the TREAT-NMD registries include

¹¹Guide is available at: www.treat-nmd.eu/downloads/file/registries_toolkit/UK_SMA_registry_protocol.pdf

(15): one single entry point for access to patient data worldwide; registries contain accurate, verified genetic diagnosis together with key clinical data items including medication use and ambulation status; patient data are updated at least once a year; it is a powerful feasibility tool as it can filter patients by precise mutation, age, ambulation status, medication type and location; and finally it is a powerful recruitment tool since patients have consented to being contacted about trials for which they may be eligible.

2.2.2 Product Registries

Once a drug or device passes the stage where it is approved for use by a regulatory authority (depending on the national state legislation) the user base becomes much bigger and from a more diverse population than the one in the stage of clinical trials, when the population is narrowly defined and only a small segment of the overall population. To address a need for quality assessment during this important post approval phase is where using a registry for identifying and enhanced understanding of product safety (acute as well as chronic use) should, as one of the available tools, come into consideration. Registries that aim to **assess safety or harm** associated with the use of various products (drugs) or devices need to 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.

It is important to note that medical devices are significantly different from pharmaceuticals in the manner in which AEs and product problems present themselves, in the aetiology of their occurrence, and in the regulation governing the defining and reporting of these occurrences, as well as post approval study requirements.¹²

Also, compared with drugs, device technologies change more rapidly over a shorter time span, requiring device registries to adapt accordingly to the changes. In addition, healthcare providers may have different levels of experience with the device, which then may influence patient outcomes (especially with devices considered as implants). Medical device registries should attempt to classify all parts of a device with as much identifying information as possible. All of the abovementioned special characteristics of medical devices should be thus taken into consideration when developing a device registry.

Product registries include patients who have been exposed to biopharmaceutical products, medical devices or diagnostic/therapeutic equipment (1).

Device registries can be designed for a variety of purposes, such as providing helpful information on the long-term effectiveness of devices and their safety, combined with keeping track of the impact of factors such as type of surgical technique, surgeon, hospital, and patient characteristics. Proper analysis from medical device registries, with control for the most relevant confounding variables, can often provide important information for decision making by clinicians, patients and policymakers.¹³

Post marketing vigilance of medical devices and drugs is needed as too much is unknown about the safety of the product when it is approved, and spontaneous AE reporting is a traditional (and legally

¹²Other sources provide more information about defining and reporting of device-related AEs and product problems, and about post marketing studies (including those involving registries), such as: Baim DS MR, Kereiakes DJ, et al. Postmarket surveillance for drug-eluting coronary stents: a comprehensive approach. *Circulation* 2006; (113):891–7.

¹³ (AHRQ) Registries for Evaluating Patient Outcomes: A User's Guide, 3Ed, Volume 2. In: Guide, editor, 2012.

binding) method through which this need is addressed. In comparison with spontaneous reporting of AEs, safety/harm registries provide certain advantages, which are here considered. There are two main characteristics of these registries that are extremely important. Firstly, we know from other science fields that any choice data architecture that demands an active and non-systematic effort by the clinician to report an adverse event is inferior (in terms of under-reporting, rather than the quality of reporting) to a systematic follow-up of those events. Secondly, and related to this, in a non-systematic reporting of adverse events we usually do not know the denominator (the exposed population) and are therefore not able to provide any epidemiological measures of disease occurrence. In a structured safety/harm registry with a defined population we can calculate the incidence of adverse events and these registries are becoming increasingly more common in the area of medical products and medical devices¹⁴.

Thus, depending on the need to comply with a post-marketing requirement or out of a desire to complement spontaneous AE reporting, the proposed product and disease registries should also be considered as a resource. The registries could be used for examining unresolved safety issues and/or as a tool for proactive risk assessment in the post approval stage. Once again, the advantage of registries is that their observational method and non-restrictive design may allow for surveillance of a diverse patient population that can include sensitive subgroups and other groups not typically included in initial clinical trials (such as children or patients with multiple co-morbidities). In contrast to clinical trials, registry populations are generally more representative of the population actually using a product or undergoing a procedure. To list just a few advantages that those registry features provide: data collection may lead to insights about provider prescribing, and also any follow-up duration can take a long time to identify the consequences of long-term use (16).

Legislation on the EU level regarding pharmacovigilance for medicines marketed within the EU is provided for in: Regulation (EC) No 726/2004¹⁵ with respect to centrally authorised medicinal products and in Directive 2001/83/EC¹⁶ with respect to nationally authorised medicinal products (including those authorised through the mutual recognition and decentralised systems). There is also a central European medicine agency (EMA) (www.ema.europa.eu/ema), which could be roughly compared to the U.S. Food and Drug Administration (FDA), although not centralized and with a lesser level of authority. EMA has issued Guideline on good pharmacovigilance practices¹⁷ (GVP) in order to facilitate the performance of pharmacovigilance activities. Finally, EMA is responsible for the management of EudraVigilance (eudravigilance.ema.europa.eu/human/index.asp) – an EU data processing network and management system for reporting and evaluating suspected adverse reactions during the development and after the market approval of medicinal products in the European Economic area (EEA).

The current system for medical devices is defined by European Medical Device Directive 93/42/EC¹⁸, which sets and describes harmonized standards¹⁹ for device manufacturing, labelling, and expected performance and safety profiles to be met. Any medical device placed on the European market must comply with the relevant legislation where there are three types of medical devices outlined: general medical devices, active implantable medical devices, and In-vitro diagnostic medical devices.

¹⁴ The term 'medical device' covers all products, except medicines, used in healthcare for the diagnosis, prevention, monitoring or treatment of illness or disability

¹⁵ Available at: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2004:136:0001:0033:en:PDF>

¹⁶ Available at: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2001:311:0067:0128:en:PDF>

¹⁷ Available at:

www.ema.europa.eu/ema/index.jsp?curl=pages/regulation/document_listing/document_listing_000345.jsp

¹⁸ <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CONSLEG:1993L0042:20071011:en:PDF>

¹⁹ e.g. ISO 14971 – Risk management for medical devices

Also, the EC has issued guidelines which aim at promoting a common approach by manufacturers and Notified Bodies²⁰ involved.²¹ EC Directives also describe the basic standards for manufacturing quality-control systems and responsibilities for AE reporting.

Guidance documents also contain templates for data collection and reports, including “clinical evaluation reports,” which are intended to provide an outline of the technology underlying a specific device and current clinical data supporting its use, ideally in reference to established standards or similar devices. In practice, each country variously interprets the requirements for quality assurance and AE reporting.

Competent Authorities, which oversee NBs in each MS, submit AE and recall data to the European Databank on Medical Devices (EUDAMED)²² (ec.europa.eu/idabc/en/document/2256/5637.html), a central database run by the EC. Since the database is non-public the basis for device approval and any post marketing commitments are largely unknown and EU-wide adverse event data are not accessible, though some MSs post market surveillance events in a non-systematic manner.

There are numerous product/device registries in the EU, differing in objectives, scope, field of medical expertise etc.

EU-ADR (www.euadr-project.org) was an EC funded project (FP7 program) with an objective to design, develop and validate a computerized system that exploits data from electronic healthcare records and biomedical databases for the early detection of adverse drug reactions (ADRs). In this project, electronic health records (EHRs) comprising demographics, drug use and clinical data of over 30 million patients from several European countries were available. EHR databases also form the foundation of the project, insofar as they supply the patient data on top of which the system is built. The EU-ADR system then intended to generate signals (drug-event pairs of pharmacovigilance interest) through the use of data mining, and epidemiological, computational and text mining techniques. Finally, an ultimate objective of the project was to demonstrate that an earlier detection of ADRs is possible through using EHRs (16).

EuraHS (European Registry for Abdominal Wall Hernias) (www.eurahs.eu) is a registry which observes hernia operations and not patients. Its mission is to develop and provide for all members of the EHS (European Hernia Society): an international platform for registration and outcome measurement; an online platform for reporting early or late mesh complications (as a survey of implant materials); a set of definitions and classifications for use in clinical research on abdominal wall hernias; a uniform method of presenting outcome results in clinical studies of its repair. It is also trying to convince existing European hernia databases to join the EuraHS, in order to collect their data on the same Internet platform, and to fulfil the goal of the registry as being a good instrument to acquire data for post marketing surveillance, increasing quality and quantity of outcome reports in hernia device-related surgeries (17).

EAR (The European Arthroplasty Register) (www.ear.efort.org) is a major activity of the European Federation on National Associations of Orthopaedics and Traumatology (EFORT). It is organized as a

²⁰ Private, for-profit third party bodies that are devices certified for marketing approval

²¹ MEDDEV 2.12-1 Rev8 “Guidelines on a Medical Devices Vigilance System” and MEDDEV 2.12-2 Rev2 “Post-Market Clinical Follow-up (PMCF) Studies”. Available at: http://ec.europa.eu/health/medical-devices/documents/guidelines/index_en.htm

²² Evaluation of EUDAMED from 2012 is available at: http://ec.europa.eu/health/medical-devices/files/pdfdocs/eudamed_evaluation_en.pdf

scientific non-profit making society located in Austria. It was founded in 2001 as a voluntary network of national arthroplasty registries. The main aims of the EAR project are: to support national orthopaedic societies to establish national arthroplasty registries based on the EU-level standardization and harmonization of processes, to conduct basic research on comprehensive patient registries, to give support for scientific activities related to arthroplasty registries and cooperation with other stakeholders using arthroplasty registry data. Currently, 30 projects in 26 countries in Europe, Israel and Saudi Arabia are linked to the EAR-network. The projects are in different stages of development. The work is organized at national level as a cooperation of orthopaedic societies and national public health authorities. EAR's main focus of activities is on outcome research and methodological research in the context of arthroplasty registries. Arthroplasty Registries are considered as a powerful instrument to assess the performance of arthroplasty procedures, and a major source for scientific discussion. EAR supports the development of Arthroplasty Registries and registry documentation, and aims to enhance the comparability of reports by standardization. EAR also produces minimal datasets which are included in all national arthroplasty registries upon which EAR's evaluations are based (18).

As already discussed above, a product/device registry may possess great potential and effectiveness in areas of post market surveillance, adverse effect reporting, assessing safety and harm but also in improving quality of care, depending on the registry's objectives.

The **Swedish Hip Arthroplasty Register (SHAR)** (www.shpr.se/en) is presented as an illustrative example of registry effectiveness. The registry started in 1979, a web-based reporting system has been in place since 1999, and since 2002 it has measured patient reported variables.²³ In 2005 the registry also started collecting data about partial arthroplasty. The registry has excellent coverage (patient coverage 98% and hospital coverage 100% in year 2009) (19). The registry is governmentally funded, and no device-manufacturing industry funding is present (although the registry sells data to industry, without identifiers). The Swedish legal context enables undisturbed data collection. The data are collected after surgery and reported to the SHAR through the internet. In accordance with the Swedish Data Act, all patients are informed about the registry and are free to give up their participation in the registration at any point. Analyses of registry data focus primarily on re-operations, short-term complications revisions (surgeries to replace devices) and patient-reported outcomes. Revision rates for hip implants in Sweden declined substantially over the years, which is largely due to the registry's success in detecting devices for hip replacement surgery which have longer survival rates. Judging by the registry's success, for instance in comparison with other countries such as the U.S., the survival of hip replacement implants among Medicare patients in the United States (1997-2005) and patients aged 65 and older in Sweden, the failure rate is about three times higher in the U.S. (20).

From its original focus on the devices themselves, the registry has moved on to analyse the whole process surrounding hip implant surgery to find predictors of good and poor outcomes (21). Also, beyond the registry's quality improvement purpose, the data in SHAR have been used for research, including several doctoral dissertations and a stream of publications on outcomes associated with different prostheses and surgical techniques; age, ethnic, and socioeconomic predictors of outcomes of hip replacement surgery; the occurrence of rare adverse events; and patient-reported outcomes. The creation of the Nordic Arthroplasty Register Association that pools data from the Swedish, Norwegian, Danish, and Finnish registries is creating additional research opportunities both because of larger numbers and because the countries have different user profiles (22).

²³ The patients complete a question form about satisfaction and health-related quality of life, serving as a baseline for comparisons one, six, and ten years after surgery.

2.2.3 Health Services Registries

Another type of patient exposure that can be used to define registries is exposure to a health care service. The focus of health service registries is on providing information used in the management of health services. They are based on service generated data derived from health facilities and patient-provider interactions. Health care services that may be used to define inclusion in a registry include individual clinical encounters, such as office visits or hospitalizations, procedures, or full episodes of care.

Health care service registries are sometimes used 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” (23).

Hospital discharge data are a specific type of health service registry data. They are widely available and very useful for monitoring the quality of health services. This source almost always includes individual records that capture different dimensions of the interactions between the health service and the individual (measuring or paying costs, basic statistics of procedures and diagnosis etc.). Hospital discharge data have been used in quality-of-care research and, recently, as an input for effective coverage assessment. As a result, it is advisable to regularly assess the quality of health service data and to help ensure some basic standardization, to the extent possible, to better serve national and regional interests. Regular monitoring also helps to better understand the aggregate capacity of a health system to provide care (productivity of care) (24).

Quality improvement registries (QI registries) seek to use systematic data collection and other tools to monitor and improve quality of care at the local, regional, or national level, as well as to broadcast clinical research. QI registries generally fall into two categories: registries of patients exposed to particular health services (e.g., procedure registry, hospitalization registry) around a relatively short period of time (i.e., an event); and those with a disease/condition tracked over time through multiple health services (1).

QI may be used for various purposes, such as: to monitor trends in the use of certain 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.

These registries may identify disparities in the availability of care, identify and investigate sub-optimum practice and processes, as well as demonstrate potential improvement opportunities. The steadily increasing costs of health care (for OECD countries annual health expenditure averaging almost 5% growth rate over the period 2000-2009) (25) imply the need to justify health care interventions and plans with accurate cost/benefit measures and by showing the impact of interventions on relevant outcomes. Without a valid system for monitoring outcomes within institutions there is little space for management to be aware of how their services truly compare with services elsewhere or with pre-determined quality standards. Since a registry can continuously record data, it has the potential to identify unnecessary or inappropriate variations of healthcare quality and incite its improvement by creating a feedback loop which can pinpoint areas of poor quality (25). Longitudinal data also provide the needed understanding in order to act as an early warning system if quality declines.

Health services registries consist of patients who have had a common procedure, clinical encounter, or hospitalization (1).

Registries can drive quality in a variety of ways, be it indirectly – through stimulating competition, or directly – through evaluating adherence with best practices and through affecting healthcare policy (pricing and regulation). In order to improve measurement of quality health care indicators, one should fully exploit the potential of (national) registries, particularly through the implementation of unique patient identifiers, secondary diagnostic coding and present-on-admission flags.²⁴

As creating and maintaining a registry of this type may require (considerable) resources, narrower focus of quality registries should thus be concentrated on conditions and procedures where outcomes are thought to vary and where improvements in quality of care have the greatest capacity to improve quality of life, and reduce costs (27) (i.e. monitoring renal transplantation outcomes, as poor outcome from this procedure forces patients to revert to haemodialysis, with subsequent consequences of lower quality of life means much higher costs to society). Finally, the ultimate purpose of data from quality of care registries should be to inform clinical practice, policy development and resource allocation.

Although these registries share common objectives in improving quality and can prove a powerful tool in improving health care value, their usefulness on carrying out the objectives varies depending on the registry's stakeholders (research or health policy oriented), scope, quality of data and finally utilization of registry information by policy makers.

The EC is supporting such initiatives, and one such is example is **HoNCAB** (honcab.eu) – a pilot network of hospitals related to payment for care for cross-border patients, with the main objectives to determine patients' rights in terms of access to cross-border health assistance and entitlement to reimbursement of such treatment, to ensure access and provision of safe, high-quality, efficient and quantitatively adequate healthcare abroad, to support collaboration between MSs regarding healthcare, and finally to obtain a better understanding of the financial and organisational requirements that may arise as a result of a patient receiving healthcare outside the MS affiliation. The network of hospitals features a functioning organisational structure and established means of communication, supported by a web-based database to collect and exchange information, all with the aim to share between MS practical experiences, problems and solutions related to cross-border care.

The benefits of quality of care registries are apparent. For example, the registry set up by the Danish Lung Cancer Group through feedback of indicators of high-quality care derived from registry data to those delivering care has been largely responsible for improvement in 30-day, 1-year and 2-year survival rates for people with lung cancer of 1.6%, 8% and 10%, respectively (28).

There is an international momentum gathering to develop new clinical registries as quality-improvement measures.

At EU MS level the biggest effort in developing quality of care registries is in Sweden, where a system of **national quality registers** (www.kvalitetsregister.se) has been established (since the 1970s). The

²⁴ OECD Health Ministerial Meeting. Improving Value in Health Care: Measuring Quality. Forum on Quality of Care, Paris, 7-8 October 2010. Available at: www.oecd.org/health/ministerial/46098506.pdf

system is recently on the rise, going hand in hand with a number of initiatives at both the national and local Swedish government levels and suggesting that governance of health care services is guided by an emerging performance paradigm. Although the traditional objectives of distributive justice and cost control are still valid, they have been complemented by objectives concerning efficiency and value for money spent on health care services (29).

Today, Sweden boasts 89 certified national quality registries of various types: interventions or procedures (e.g. hip fracture repair and cardiac surgery); diagnoses and episodes of care (e.g. myocardial infarction and stroke); and chronic disease (e.g. diabetes and leukaemia). National quality registers cover more than 25% of total national health expenditures, about one third of the registries collect patient data on more than 90% of all Swedish patients diagnosed with a given condition or undergoing a particular procedure, and many have been in place long enough to provide unique longitudinal information on patient cohorts.

Thus, in addition to being a comprehensive primary data source for comparative studies, the high percentage of coverage of health services enables Swedish registry data to play an important role in the monitoring and evaluation of health care quality, as well as help in developing nationwide health care policy²⁵, while constantly being a resource for research, one of the registries' common feats.

The vision for quality registries and competence centres is to constitute an overall knowledge system actively used at all levels (health provider, hospital, regional, state) for continuous learning, and evaluation, development, quality improvement and management of all health care services (30).

A national quality registry contains individualized data concerning patient problems, medical interventions and outcomes after treatment; within all healthcare production. It is annually monitored (quality control) and approved for financial support by an executive committee. Funding comes from central state level and is allocated to a few competence centres, where several registries share the costs of staff and systems which it would not be possible for a single registry to fund. The successful development of the Swedish National Quality Registries is explained largely by their decentralized nature. Caregivers that have the greatest use for data also have the main responsibility for developing the system and its contents, and databases are spread out among different clinical departments throughout Sweden. Another potential reason for success could be relatively liberal legal provisions concerning personal data in Sweden, where special permission can be obtained that allows national personal data to be recorded and processed²⁶ (even in universities) (31).

Also, data quality of registries in the national quality list is quite high and as a result sufficient for use in clinical research (32).

Outside the EU, Australia is also trying to establish a national base of clinical quality registries with goals similar to those of Sweden, and with certain advantages (national level policy) and disadvantages (existing registries lack nationwide coverage). Clinical quality registries in Australia are envisioned as indicated in [Figure 2.1](#) (33).

²⁵ A recent assessment of quality in Swedish health care, including the country's register system, made by the Boston Consulting group, found that the registries are improving quality and efficiency in health care. The report from BCG recognized the potential of registries to increase value in health expenditures, and they estimated that investing in registries in the Swedish context would generate a significant cumulative return over the next years because of improvements in quality. Available at: www.bcg.com/documents/file64538.pdf

²⁶ Possible importance of privacy legislation for success of a registry – see subchapter [2.2.2](#) regarding the Swedish Hip Arthroplasty Register.

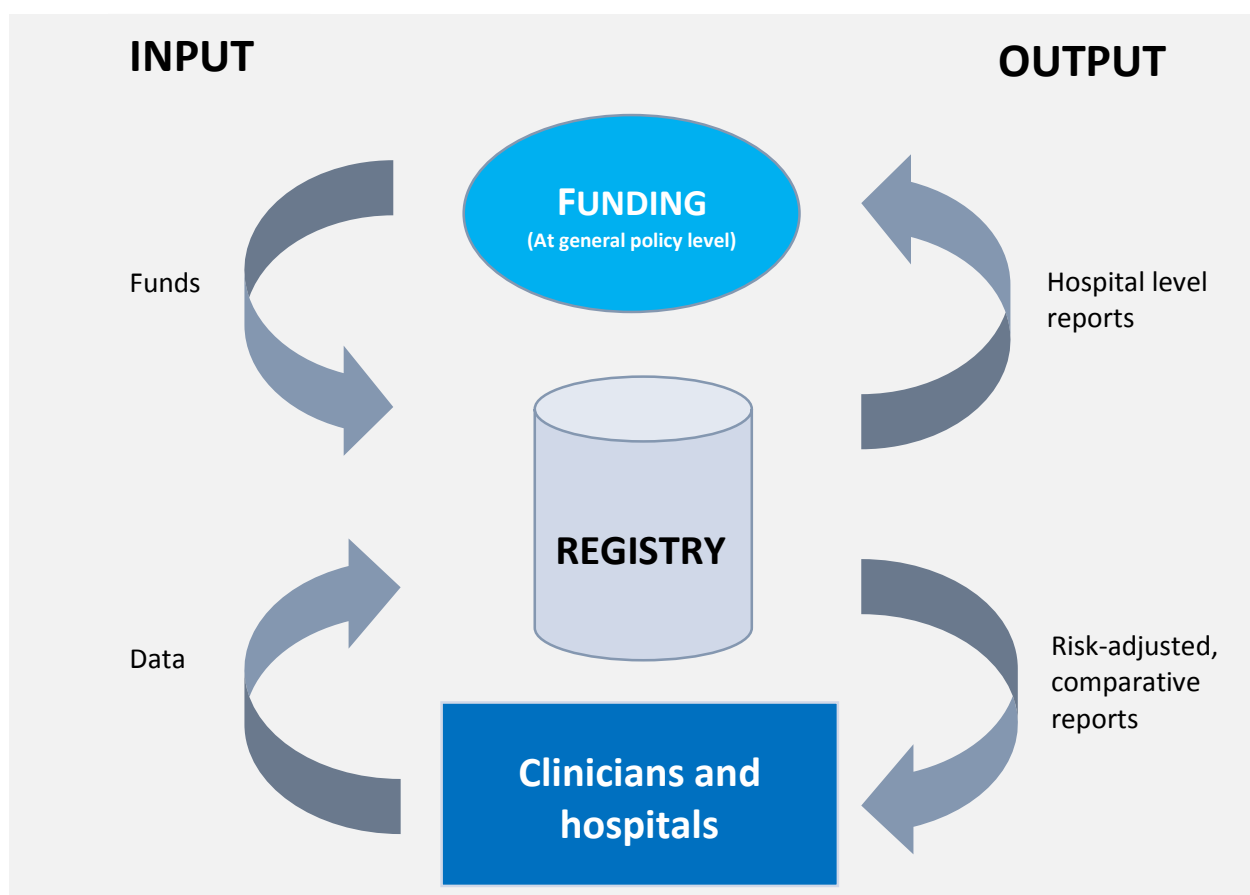


Figure 2.1: Dependencies of clinical quality registries in Australia

To summarize, when considering a clinical quality registry, collection and feedback of data must be based on an effective central governance structure, with strong clinical leadership, and a regulatory framework that provides incentives for quality improvement and dedicated approaches for managing poor performance. Local clinical leaders should ensure that registry outcomes drive quality improvement.

2.3 Diversity in Use of Patient Registries

As illustrated in the previous chapter, a patient registry can be a powerful tool for a number of potential needs: to understand variations in treatment and outcomes, to examine factors that influence prognosis and quality of life, to describe care patterns, to assess effectiveness, to monitor safety and harm, and to measure quality of care. Through functionalities such as feedback of information, registries are also being used to study quality improvement (34).

Registries today vary by organization, condition and type, and have different strengths and limitations accordingly. Different stakeholders perceive and may benefit from the value of registries in different ways. For a clinician, registries can collect data about disease presentation and outcomes on large numbers of patients rapidly, thereby producing a real-world picture of disease, current treatment practices, and outcomes. For an organization of physicians, a registry might provide data that can be used to assess the degree to which clinicians are managing a disease in accordance with evidence-based guidelines, focus attention on specific aspects of a particular disease that might otherwise be overlooked, or provide data for clinicians to compare themselves with their peers (35). From a private payer's perspective, registries can provide detailed information from large numbers of patients on how

procedures, devices, or pharmaceuticals are actually used including data for evaluating their effectiveness in different populations. This information may be useful for determining coverage policies (36). Furthermore, for a drug or device manufacturer, a registry-based study might demonstrate the performance of a product in the real world, develop hypotheses, or identify patient populations that will be useful for product development, clinical trials design, and to identify individuals eligible to participate in research. The use of patient registries varies by priority condition, with cancer and cardiovascular disease having a large number of registries and areas such as developmental delays or dementia, far fewer. Overall, the use of patient registries appears to be active and growing (1).

2.4 Overview of European Registries

The current European registry landscape is often viewed 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 creation of segregated silos, resulting in expensive and inflexible IT systems. Often, registries are 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 different European countries and types of organizations. As a result, patient registries implement only a subset of the registry functions, using and producing only a fraction 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.

However, some efforts are being made to improve the situation. Through performing a literature review numerous such projects have been identified and recognized as best practices (briefly presented in chapter [2.2](#)).

The criteria for recognizing best practices are in accordance with overall PARENT aims, and include projects, organizations, initiatives, registries etc. working on national, regional or international levels in the fields of:

- Recognizing and converging similar sources of data (based on disease, device and/or service) in order to improve surveillance, quality, outcomes, safety and/or effectiveness.
- Tackling different levels of data exchange (individual or aggregated level, metadata) between similar (group of registries) or different sources of data (registries – EHR – insurance databases).
- Addressing healthcare data exchange issues such as standards, interoperability, metadata, platform, common datasets etc.
- Defining needs for efficient health information exchange on different levels (patients, health care providers, researchers, payers, decision makers etc.) and ways to address those.
- Promoting collaboration, reducing redundancies, and improving transparency among patient registry holders.
- Aligning patient registries classification, definitions, taxonomy, purpose, development and governance.
- Adding value through evaluating information produced by secondary use of health data.

The next subchapter presents a list of patient registries and a short descriptive analysis of some of their features.

2.4.1 Member State level registries overview

To date, PARENT WP4 team in collaboration with project partners has compiled a list of registries identified as suitable for being taken into consideration as regional/national/county and/or local level patient registries. The list²⁷ currently contains 1028 registries and is continuously growing as additional information arrives (newly discovered data sources, literature, and information from project partners). It should be noted that the results presented in this chapter below are based on responses from project partners.

Most of the registries from European countries are located in Spain, mainly due to the specific organisational principle of Spanish healthcare registries. The vast majority of patient registries in Spain are county-based, which means that each contains equivalent registries (e.g. Basque Country Cancer Registry, Murcia cancer registry, etc.), while many other countries may collect the same type of data at a national level (e.g. Polish national cancer registry). Other highly represented countries are also characterized by a comparatively high-level of organization of healthcare registries at a national level. These often provide meta-registries or registry lists (UK DocDat, IR HIQA, PT DIS, SE National Quality Registers, etc.) which provide information on a large number of patient registries that are, or have been operating within a certain country. Less prominently featured countries often have a smaller number of active patient registries in total, but may also be underrepresented due to a lower level of international visibility. This may be due to organizational issues, lack of connectedness between registries at a national level and/or lack of other specialized focal organizations at an international level. Although these international organizations often contain comprehensive lists of patient registries, they are often characterized by a specific focus (like Orphanet, which contains the most comprehensive list of data on 641 patient registries, but consists only of rare diseases registries), which is why it is likely that there would be immense benefits from an establishment of a general cross-border meta-registry organized around collecting data on all active patient registries. There are also already several multi-country registries in our list which collect data from several countries at once. These may be either international registries of specific conditions such as coronary events, or specialized international studies collecting patient data.

²⁷ Available at: www.parent-ror.eu/#!state/list_all

Table 2.2: Distribution of identified registries across European countries

Country	N	Country	N
Spain	191	Latvia	17
UK	139	Estonia	16
France	82	Slovenia	15
Portugal	66	Netherlands	14
Ireland	65	Multi-country	13
Germany	41	Czech Republic	11
Hungary	40	Switzerland	10
Austria	38	Malta	9
Italy	38	Cyprus	8
Finland	32	Greece	7
Sweden	29	Romania	6
Croatia	28	Lithuania	4
Poland	24	Serbia	2
Norway	23	Albania	1
Belgium	19	Bulgaria	1
Denmark	19	Georgia	1
Slovakia	18	Turkey	1
		Total	1028

Based on our general classification (primary purpose) we recognized that the majority (64%) of patient registries were disease/condition based, followed by service (26%) and product based patient registries (10%).

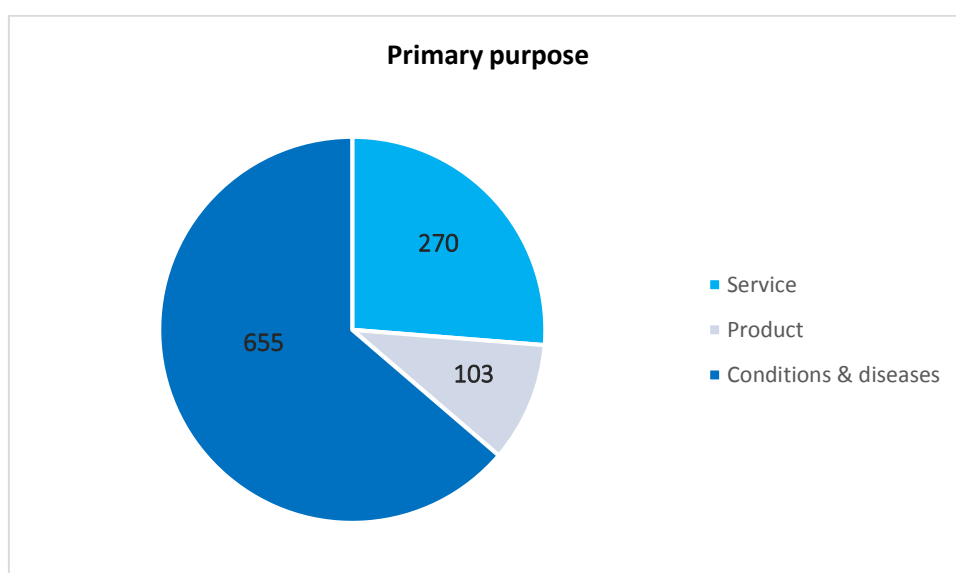


Figure 2.2: Breakdown of all registries based on primary purpose (N=1028)

Further categorizing them into **disease/condition based patient registries** according to entry criteria definition (particular disease or condition), we recognized several sub-categories based on organ system (cardiovascular, neuromuscular etc.) or clinical field (cancer, rare, congenital, occupational) irrespective of body part focus. The largest number of disease/condition based registries in our list falls

under the coronary/vascular subcategory (27%)²⁸, followed by cancer/tumour/haematological (20%), infectious (9%), rheumatic (8%) and pulmonary (7%). Although rare diseases contribute only to 6% of registries in our list, the extended list²⁹ contains 641 rare disease registries in total (to be integrated as a joint activity of PARENT and Orphanet). All other subcategories account for 23% of total disease/condition registries in the list.

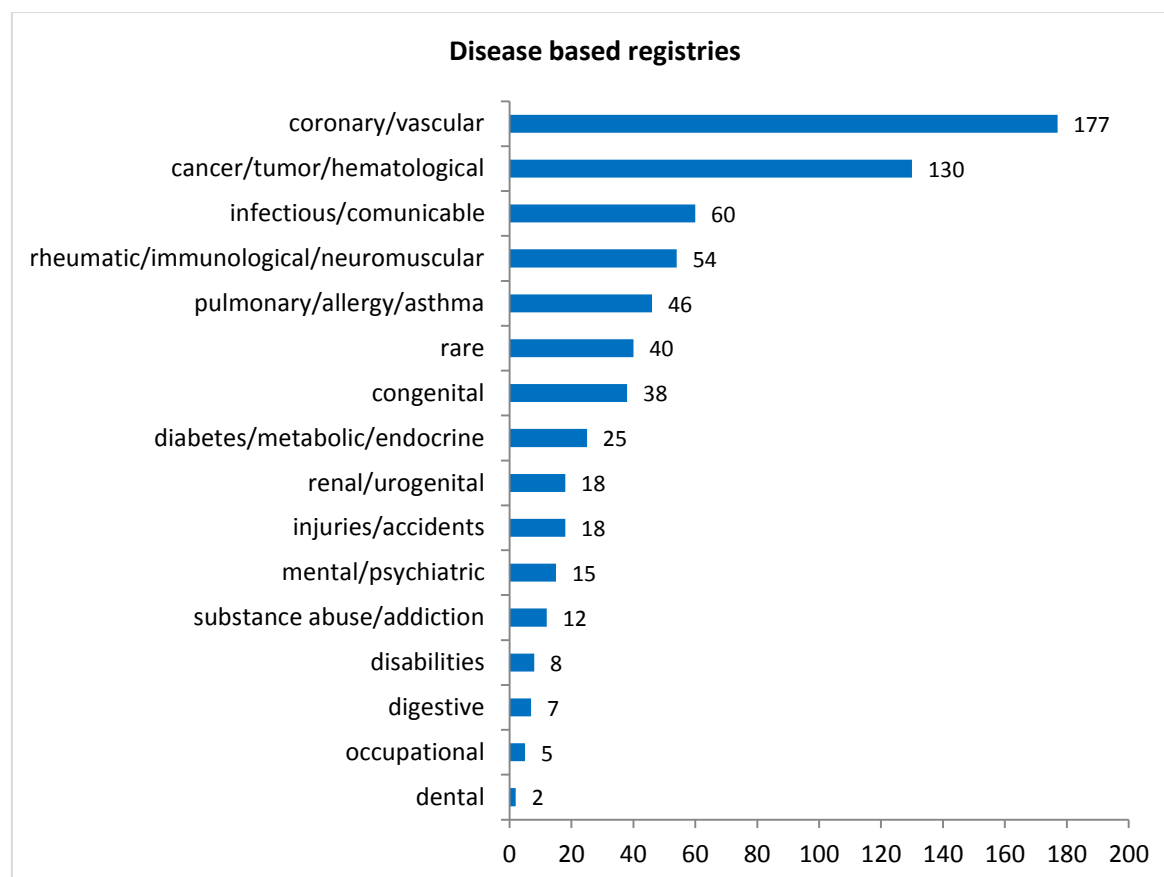


Figure 2.3: Breakdown of Condition and Disease based registries (N=655)

While the number of **product-based patient registries** represented a minority of all registries in our list, two subcategories can be further identified: device registries (most prominently featuring devices such as pacemakers or arthroprosthetics) and pharma registries (registries collecting data on pharmacological products). Less than 20% of product registries belong to the latter, while the much larger proportion of product registries were identified as medical device-based registries.

²⁸ Mainly due to integration of European Society of Cardiology (ESC) meta registry data – dynamic portal (www.esc-crt.org/workstream/Pages/dynamic-portal.aspx Accessed: 9th June 2014)

²⁹ Orphanet list of rare diseases registries, January 2014 (www.orpha.net/orphacom/cahiers/docs/GB/Registries.pdf)

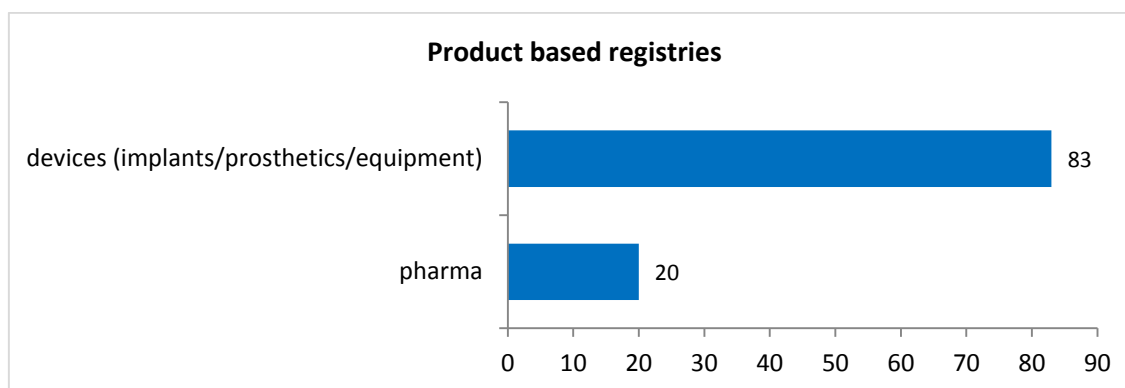


Figure 2.4: Breakdown of Product based registries (N=103)

Apart from condition/disease and product registries, our review of patient registries yielded a third category of registries which we categorize as **service-based patient registries**. This group is the most heterogeneous of all and consists of registries whose primary definition and focus is ostensibly based upon healthcare services. The largest identified subcategory contains registries evaluating preventative services, quality of care, and health monitoring. It accounts for exactly a third of all service-based registries and includes population, permanent sample and vulnerable groups' registries and registries used for evaluating preventative screening programs or monitoring population health. The second biggest subgroup contains various specific medical procedures registries (24%) which monitor specialized surgical procedures, therapeutic or diagnostic services or emergency interventions. All other observed service based registries accounted for less than a half of this group and were subcategorized as registries of transplant procedures and/or donors (blood, bone marrow, organ etc.), various obstetric and gynaecological services registries (births, abortions, medically assisted fertilization), immunization, causes of deaths registries, hospital discharges registries and registries for health/social insurance purposes.

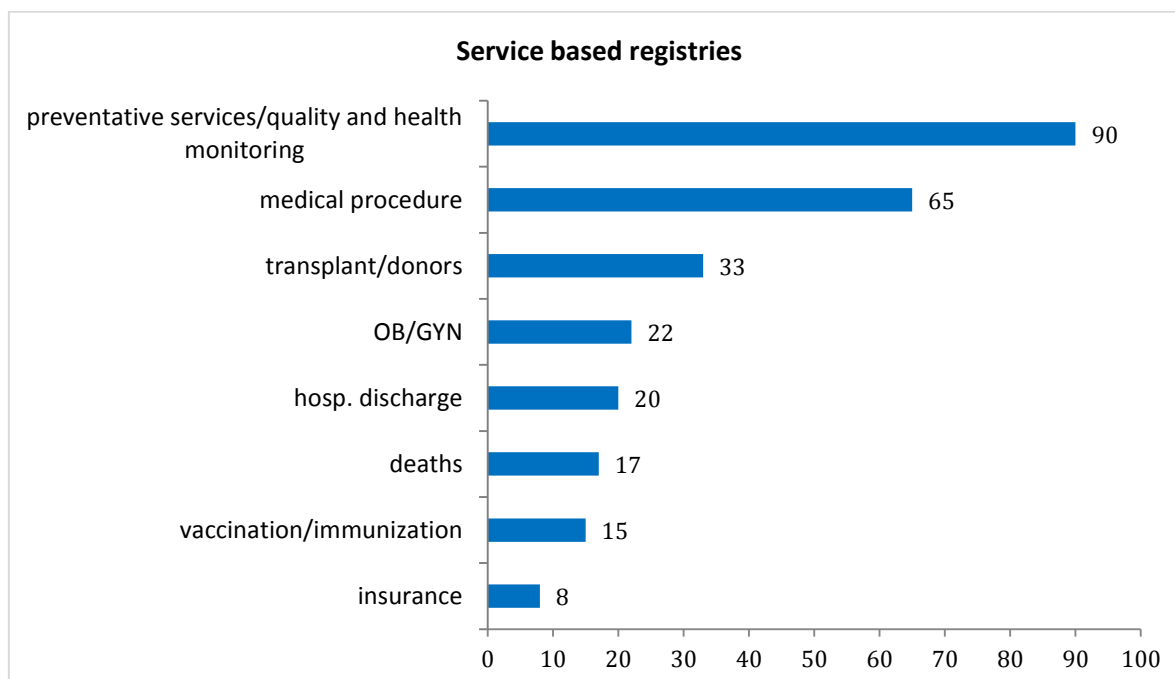


Figure 2.5: Breakdown of Service based registries (N=270)

2.5 Key issues arising within registries

Within this subchapter, we will briefly look into the most important and emerging issues arising from registries. It should be noted that issues presented here are based on the questionnaire survey of registry holders (n=131; registry list available in Appendix A) performed by the PARENT project team. Therefore, the results do not reflect a regulatory perspective, from which other issues within registries may be recognized.

Possible solutions and proposals are listed in subsequent chapters. Emerging issues at national and regional level differ from issues at EU level, in line with differences in setting-up and running a registry. At the same time, it is necessary to point out that the majority of EU-level registries are based on secondary data sources.

Besides, there are different views to key issues regarding the role of the reader: major concerns of registry holders might differ from major concerns of data users. The issues mentioned below are trying to address both sides, but again different views cause different perspective of the same issue. Therefore, one can recommend to browse the list below and refer to later chapters of the guidelines where these challenges are further elaborated.

Not all of the issues listed below are relevant for national or EU registries, but when setting-up the registry, all of them should be considered.

1. The most important issues among EU registries are **unstable funding and therefore limited sustainability**. At this point, the differences among national (or regional) registries and EU-level registries are important and influence the sustainability: Funding of national registries by national authorities might not be stable; as mentioned in 'PARENT - Deliverable 5: Registry analysis and Report', only half of the registries are currently funded by national government authority, about 16 % have "no specific funding". EU registries are funded either by an umbrella organisation or by a certain project, which again introduces instability and limited sustainability.

Table 2.3: Funding source

Funding source (question 7)	Initial registry funding - set-up		Current registry funding	
	N	%	N	%
National government authority	58	36%	76	52%
No specific funding	27	17%	24	16%
Regional Authority	18	11%	18	12%
University/Research Institute	14	9%	1	1%
Foundation	12	8%	8	5%
EU commission agency	10	6%	3	2%
Hospital	10	6%	5	3%
Industry	9	6%	5	3%
Patient Association	2	1%	6	4%
Total	160*	100%	146**	100%

* Multiple choice question

** The difference in numbers of initial and current registry funding registries is due to missing answers.

2. There are many **legal issues** concerning registry set-up, data protection and re-use. Legal backgrounds in Member States differ greatly. At this stage, the preparation of the new regulation on data protection should be pointed out as it might influence the future of the majority of patient registries in EU. Much more on these issues is described in chapter 5.
3. Within the phase of development (or setting-up) the registry and also later on, **the roles of different stakeholders** are very important and, in many cases, not very clear. There are different possible roles: data owners, data holders, data users, etc. Much more on these issues is described in chapter 6.
4. **Modes of data collection:** almost half of the EU registries are still based on paper-and-pen mode (paper based questionnaires, paper based health records and laboratory results). The situation is burdensome for data providers and causes lower data quality. One should point out that paper-and-pen data collection mode is nowadays not desirable since it is costly, time consuming and does not allow any control of the data filled in. To read more on this topic, see chapter 6.1.2.1.3 and 8.1.1.1.

Table 2.4: Sources of data for a registry

Sources of data (question 17)	N	%
Paper based questionnaires	67	22%
Electronic health care records	56	18%
Online questionnaires	53	17%
Paper based health records	44	14%
Paper based laboratory results	34	11%
Electronic laboratory results	26	8%
Directly from clinical examinations	17	5%
Interviews	14	5%
Total	311*	100%

* Multiple choice question

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. More information on standards is available in chapter 3.2.5.1 and 10.11.
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 on quality assessment. On the other hand, only 20 % of registry holders would like to have a common quality control tool (see 'PARENT - Deliverable 5: Registry analysis and Report'). More on registry data quality is described in chapter 4.
8. Registry transparency and openness with the emphasis on **data access for research purposes**: a majority of registries are closed to researchers from other institutions than the data holder. There should be protocols enabling users/researchers to access the data under certain conditions (see project Data Without Boundaries: www.dwbproject.org/).
9. **Insufficient data dissemination**: a minority of registries actually disseminate their aggregated data on the websites allowing users to get easy access to the first results. The commendable exceptions are cancer registries with wide dissemination (see <http://eu-cancer.iarc.fr/EUCAN/Default.aspx>, <http://eu-cancer.iarc.fr/EUREG/Default.aspx>, [38](http://eu-

</div>
<div data-bbox=)

cancer.iarc.fr/EUREG/Default.aspx). These registries have established standards that should be followed by other registries. On the other hand, it should be pointed out that dissemination standards differs from country to country; in general, the data in all forms are much easier accessible in, for example, Nordic countries, UK, France,... than in some other countries. To read more on data dissemination, see chapter 8.1.6.

References

1. Gliklich RE, Dreyer NA, eds. Registries for Evaluating Patient Outcomes: A User's Guide, 3Ed, Volume 1. In: Guide, editor, 2012.
2. <http://www.eubirod.eu>
3. <http://www.eaaci.org/activities/task-forces/1904.html>
4. <http://www.eaaci.org>
5. Available at: <http://www.epaac.eu>
6. Martin-Moreno JM, Albrecht Tit, Krnel Radoš S, eds. Boosting Innovation and Cooperation in European Cancer Control. Ljubljana, 2013. Available at: http://www.epaac.eu/images/OF_Ljubljana/Cancer_book_web_version.pdf
7. http://europa.eu/rapid/press-release_MEMO-14-141_en.htm
8. Patterson MC, Mengel E, Wijburg FA, Muller A, Schwierin B, Devon H, Vanier MT, Pineda M. Disease and patient characteristics in NP-C patients: findings from an international disease registry. Orphanet J Rare Dis. 2013 Jan 16;8:12 <http://www.ojrd.com/content/pdf/1750-1172-8-12.pdf>
9. https://www.lsdregistry.net/fabryregistry/hcp/understd/freg_hc_u_aboutreg.asp
10. <https://www.eimd-registry.org/>
11. Council Recommendation of 8 June 2009 on an action in the field of rare diseases (2009/C 151/02) <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2009:151:0007:0010:EN:PDF>
12. Aymé S., Rodwell C., eds., „2013 Report on the State of the Art of Rare Disease Activities in Europe“, July 2013. Available at: <http://www.eucerd.eu/upload/file/Reports/2013ReportStateofArtRDAactivitiesIII.pdf>
13. http://www.orpha.net/consor/cgi-bin/Education_AboutOrphanet.php?lng=EN Accessed: 30.5.2014.
14. <http://www.eurems.eu>
15. <http://www.treat-nmd.eu/industry/trial-sites-and-patients/patient-registries>
16. <http://www.euadr-project.org/drupal/files/pdf/FinalPublishableSummary.pdf>
17. Muysoms F, Campanelli G, Champault GG, DeBeaux C, Dietz UA, Jeekel J, Klinge U, Köckerling F, Mandala V, Montgomery A, Morales Conde S, Puppe F, Simmermacher RKJ, Śmietański M, Miserez M. EuraHS: the development of an international online platform for registration and outcome measurement of ventral abdominal wall hernia repair. Hernia. 2012 Jun; 16(3): 239–250.
18. <http://www.ear.effort.org>
19. Leonardsson O, Garellick G, Kärrholm J, Åkesson K, Rogmark C. Changes in implant choice and surgical technique for hemiarthroplasty: 21,346 procedures from the Swedish Hip Arthroplasty Register 2005–2009. Acta Orthop. 2012 February; 83(1): 7–13.
20. Kurtz SM et al. “Future Clinical and Economic Impact of Revision Total Hip and Knee Arthroplasty.” Journal of Bone and Joint Surgery 2007; 89(Suppl 3): 144-151.
21. Swedish Hip Arthroplasty Register. Annual Report 2009. Available at: <http://www.shpr.se/Libraries/Documents/AnnualReport-2009-EN.sflb.ashx>
22. Gray BH. Registries as a Knowledge-Development Tool: The Experience of Sweden and England. Urban Institute: 2013. Available at: <http://www.urban.org/UploadedPDF/412860-Registries-as-a-Knowledge-Development-Tool-The-Experience-of-Sweden-and-England.pdf>
23. (IOM) <http://www.iom.edu/Global/News%20Announcements/Crossing-the-Quality-Chasm-The-IOM-Health-Care-Quality-Initiative.aspx> Accessed: 27.5.2014.

24. Health Information Systems Knowledge Hub <http://www.uq.edu.au/hishub/docs/WP05/HISHUB-WP5-Full-12-WEB12Oct12.pdf> Accessed: 15.7.2015.
25. <http://www.oecd.org/els/health-systems/health-expenditure.htm> Accessed: 27.5.2014.
26. http://www.heart.org/idc/groups/heart-public/@wcm/@adv/documents/downloadable/ucm_438049.pdf Accessed: 27.5.2014.
27. Eyenet Sweden. Handbook for Establishing Quality Registries. Sweden: Eyenet Sweden, 2005.
28. Grosvenor Management Consulting. Australian Clinical Quality Registries Project - Final Report. 2010. Available at: http://www.safetyandquality.gov.au/wp-content/uploads/2012/02/27_External-Evaluation-Report-by-Grosvenor-Management-Consulting-REPORT-PDF-2935-KB.pdf
29. Jakobsen E, Palshof T, Osterlind K, Pilegaard H. Data from a national lung cancer registry contributes to improve outcome and quality of surgery: Danish results. *Eur J Cardiothorac Surg* 2009; 35: 348-352.
30. Anell A, Glenngård AH, Merkur S. Sweden: Health system review. *Health Systems in Transition*, 2012, 14(5):1–159.
31. Öien RF, Ovhed I. The Swedish Quality Registries and Primary Health Care. *ImPrim Report*, 2013. Available at: <http://www.ltbleskinge.se/download/18.588b0a5513a52b7563450e/WP3+Swed+Quality+Registries+ImPrim.pdf>
32. Labek G, Janda W, Agreiter M, Schuh R, Böhler N. Organisation, data evaluation, interpretation and effect of arthroplasty register data on the outcome in terms of revision rate in total hip arthroplasty. *Int Orthop*. 2011; 35(2): 157–163.
33. Swedish Association of Local Authorities and Regions (SALAR). National Healthcare Quality Registers in Sweden 2005, Stockholm, 2005. Available at: <http://webbutik.skl.se/bilder/artiklar/pdf/7164-096-7.pdf>
34. Adapted from: Swedish Agency for Growth Policy Analysis. Measurements for Improved Quality in Healthcare – Australia 2013. Available at: <http://www.tillvaxtanalys.se/en/home/publications.html?state=view&skip=30&sv.url=12.6288e13b13a4f43c5882525>
35. Labresh KA, Gliklich R, Liljestrand J, et al. Using “Get With The Guidelines” to improve cardiovascular secondary prevention. *Jt Comm J Qual Patient Safety*. 2003 Oct;29(10):539–50
36. Kennedy L, Craig AM. Global registries for measuring pharmaco-economic and quality-of-life outcomes: focus on design and data collection, analysis and interpretation. *Pharmacoeconomics*. 2004;22(9):551–68.
37. Dhruva SS, Phurrough SE, Salive ME, et al. CMS’s landmark decision on CT colonography – examining the relevant data. *N Engl J Med*. 2009;360(26):2699–2701.
38. Porta M (ed), *A dictionary of Epidemiology*. New York. Oxford University Press. 5th edition 2008
39. Olsen J, Basso O and Sørensen HT. What is a population-based registry?, *Scand J Public Health* 1999; 27:78
40. http://www.eubirod.eu/documents/downloads/BIRO_Monograph.pdf

3 INTEROPERABILITY

*Marko Brkić, Borna Pleše, Vanja Pajić, Mladen Kostešić,
Ranko Stevanović, Tamara Poljičanin, Ivan Pristaš*

Whether you govern, develop, operate or use registries, being interoperable with your registry stakeholders and peers can significantly improve your performance and resolve your challenges. But most of all, interoperability can help you to become a part of an integrated resource, relieve you of chores beyond your personal and professional interest and consolidate your status, integrity and autonomy.

Key principles:

- Get acquainted with the big interoperability picture (a network of stakeholders, users, services and registries) and how registries fit in it.
- Understand four levels of interoperability (legal/formal, organizational, semantic and technical).
- Apply interoperability principles to all aspects of registry including establishment, development, operation, use and governance keeping in mind that the user is a key interoperability factor.
- The PARENT framework is an objective-based interoperability framework, in other words its function is to provide a shared infrastructure for development of common interoperability support and functionalities to all stakeholders and projects joined around the PARENT objective.

3.1 Introduction

As a multi-stakeholder project and effort PARENT is a model environment in which interoperability is the key prerequisite for successful accomplishment of project objectives (including objectives and outcomes for each particular stakeholder), since all PARENT issues are, in essence, interoperability issues.

Interoperability, in the broadest sense, stands for “ability to operate with others”, thus can be applied to any situation where two or more entities achieve their goals or purpose by successfully interchanging services.

This principle can and should be applied to all aspects of registry establishment, development, operation, use and governance. This principle is also crucial for efficient cooperation with other national and EU registries and stakeholders and incorporating your registry in the connected European environment.

Since both registry and interoperability basic principles are by rule generic (bear no specifics regarding patient registries), these guidelines should be viewed as a brief introduction to the topic, necessary to understand, participate in and achieve the PARENT-specific goals as described in the rest of this document.

For the purpose of this chapter, the term *registry* includes the formal, current, verifiable, undisputable structured list of patient-related, medical or public records and the organizational and technical mechanism required to adequately maintain its function and records and provide related services.

Interoperability maximises the utility of patient registries and provides new opportunities for research, reduces administrative workload, provides accelerated communication and more efficient collection of data from multiple systems, enables automated data sharing and meaningful comparisons of data between registries. This finally results in the better effectiveness of registry information. The benefits of interoperability are plentiful, and it is thus recommended for a patient registry that it takes interoperability issues into account from the very start, during the registry creation phase.

These guidelines will provide essentials for assessing and building interoperability capabilities, and should be seen as a living reference which will be enhanced and supplemented through development of the PARENT framework.

For a deeper insight and advanced concepts behind these guidelines please consult and study the given references.

3.1.1 Contexts

3.1.1.1 EU context

Being an EU effort, the best way for PARENT to both support overall EU objectives and efficiently and effectively achieve goals is through compliance with EU interoperability context.

Interoperability is a key prerequisite for timely achievement of EU-level strategic plans by efficient and coordinated joint operative action of all stakeholders across all Member States.

This is reflected throughout EU strategic and operational documents and activities. Becoming acquainted with them can help to build an initial interoperability capacity and adequately position a role and goals in the context.

eHealth Action Plan 2012-2020 (1) defines the overall operative context for the Plan period. It is facing significant challenges of interoperability in eHealth, some of which have been detected by eHGI (2), and many are directly reflected as patient registry interoperability issues.

Success of EU-level action plans rely on interoperability of all involved parties. An excellent high level interoperability context overview is given in the eHealth EIF study (3).

For an insight into a service environment that demonstrates cross-border interoperability between electronic health record systems in Europe please review epSOS³⁰ project website.

Among the significant projects focusing on particular interoperability level issues are SemanticHealthNet³¹ and EHR4CR³².

For a practical context working model and coordination outcomes refer to CALLIOPE³³ EU thematic network site.

Patient registries are key healthcare information repositories, therefore interoperability of their stakeholders and users is crucial for eHealth Action Plan execution.

3.1.1.2 Registry interoperability context

In order to contextualize registry interoperability the following figures illustrate the intended approach. [Figure 3.1](#) is shown for easier comparison of small, but significant differences between a simplified traditional registry context and the interoperability development context ([Figure 3.2](#)).

³⁰ epSOS - Smart Open Services for European patients: www.epsos.eu

³¹ SemanticHealthNet, a scalable and sustainable pan-European organisational and governance process for the semantic interoperability of clinical and biomedical knowledge: <http://semantichealthnet.eu>

³² Electronic Health Records for Clinical Research, adaptable, reusable and scalable solutions for reusing data from Electronic Health Record systems for Clinical Research; <http://ehr4cr.eu>

³³ EU-funded Thematic Network "CALLIOPE - Creating a European coordination network for eHealth interoperability implementation: www.calliope-network.eu

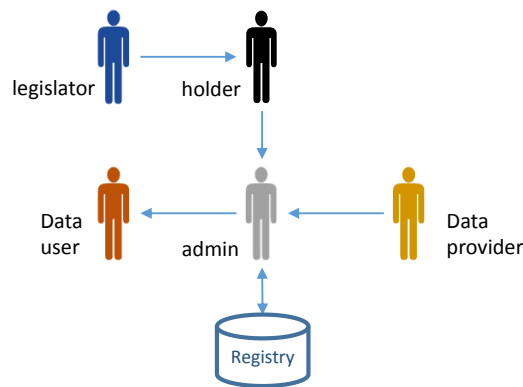


Figure 3.1: Generic single registry context

In the generic single registry context (no interoperability) all registry rules, administration and service provision are determined by the registry holder according to current legislation and business decisions. Dedicated administration (human actor) performs four-way interoperability actions when needed, as a part of a “business-as-usual” job description. It is a closed system where each interaction is prescribed by the administration and, regardless of the IT solutions being used, requires human intervention in every use service.

When we want to optimize the relationship and enable parties to interoperate we need to establish a new, unprecedented joint business environment. If this is not mandated through legislation, the first prerequisite is a formal agreement/commitment of all stakeholders to jointly develop and use new functionalities (services). In eHealth EIF terms this formal mandate or formal agreement is called *political context*.

Once the political context is established, stakeholders need to make sure that their mutual interoperability is adequate to efficiently achieve the functionality determined on the political level, or what is needed to build the required capacity.

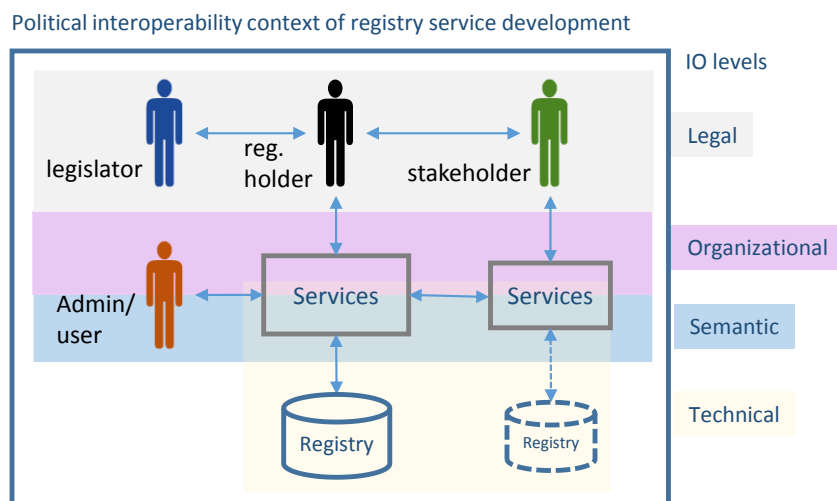


Figure 3.2: Interoperability development context

Competent and appointed stakeholder representatives need to review, verify and agree on requirements on four interdependent levels: legal (formal), organizational, semantic and technical

level and adequately adjust their business systems and services. Elimination of any level from the process can result in inadequate solutions.

Once a number of stakeholders reaches higher levels of interoperability we are going to be able to delegate more functions to a fully interoperable registry service context, as presented in [Figure 3.3](#):

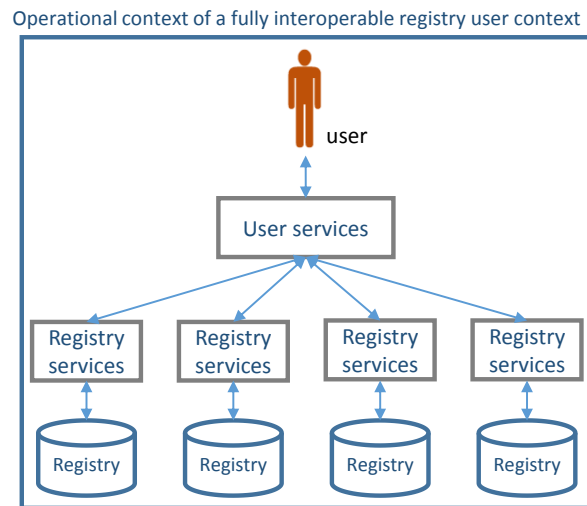


Figure 3.3: Advanced registry interoperability context

[Figure 3.4](#) shows a functional PARENT framework (interoperability) context. The *user* now represents all stakeholder roles in previous contexts, while user services provide the interoperability environment. In this context registry holders ensure interoperability by developing and maintaining own registry services and using standardized and shared user services to interoperate either with human users or other technical systems and registries.

3.1.1.3 Generic use case context

Keeping in mind the previous contexts and disclaimers, a generic use case “*perform registry service*” covers key registry interoperability situations.

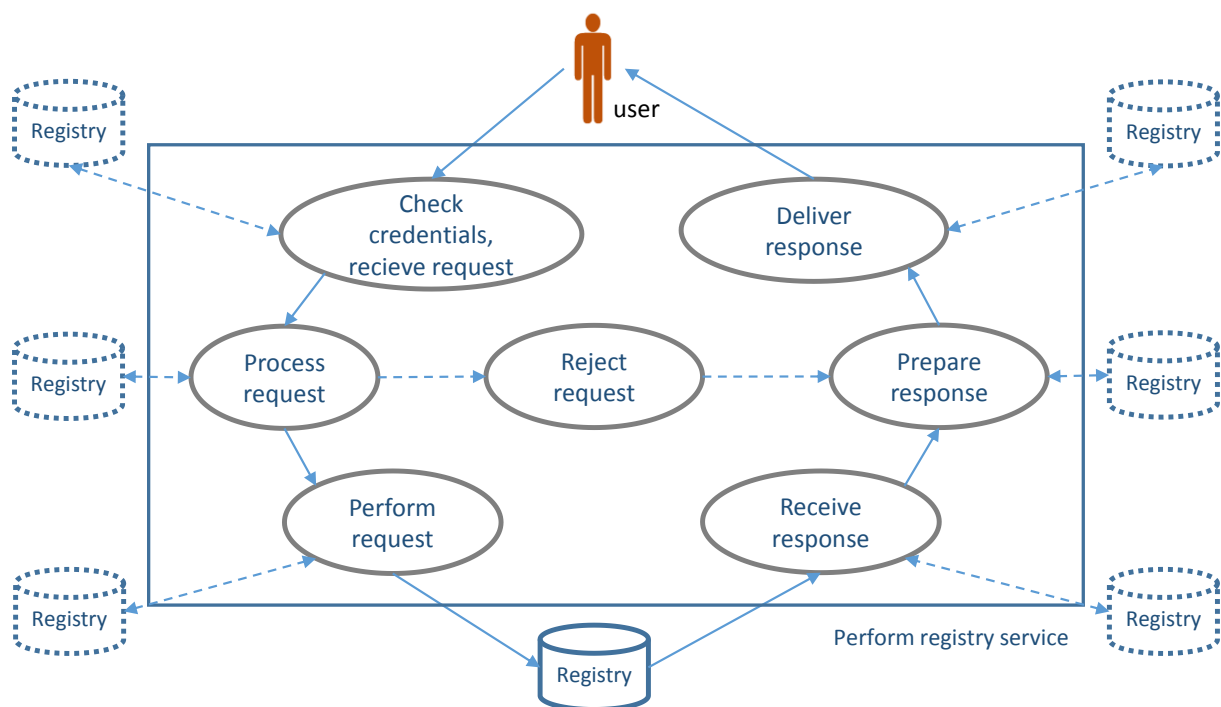


Figure 3.4: PARENT framework (interoperability) context

In this use case the actor *user* should also be viewed in the broadest sense. It can be anything, a human individual or a machine in any role: from a registry holder, physician, patient, government official to another service performed by a hospital or completely independent system.

If needed, each shown task can be also viewed as a separate interoperability use case and performed by any number of stakeholders that can provide the best result. Each task can also connect to one or more registries when the needed interoperability conditions are met.

Each task can and should be viewed from all interoperability level angles.

Each task can be performed either by a human or a machine. This enables modular development and can help a gradual interoperability capacity development for complex traditional registries. For example, at one point a nurse can perform all registry service tasks by hand and paper, while later one step, several or all steps can be automated until the nurse takes the position of a user and is relieved of administrative tasks.

From an interoperability point of view, the service or any particular task in this context could be delegated and performed by different independent stakeholders and systems located around the EU and supported by registries located somewhere else.

The value of this use case context is that it contains all key elements needed to perform create, read, write and delete actions (the core of any registry operation) through services. This can help to map a particular registry situation to this generic model as a starting point in interoperability development that can be either the initiator's task or a task which the initiator participates in. It can also be a tool for conceptualizing a new registry that has all the interoperability prerequisites. The PARENT framework will provide tools to help achieve this.

3.1.2 The PARENT Framework

At the initial stages of interoperability development most of the burden lies on individual stakeholders, but one of the most important interoperability possibilities is resource and capability sharing. Having a standardized system that could systematically build, incorporate and manage resource and capability sharing would relieve stakeholders and users of the interoperability and standardization overhead (the need to re-build and re-learn to be interoperable) and allow them to focus on their core business.

The PARENT framework is an objective-based interoperability framework, in other words its function is to provide a shared infrastructure for development of common interoperability support and functionalities to all stakeholders and projects joined around the PARENT objective. It should support the full interoperability range and functions as its integrator. It is a constantly active mechanism governed by key stakeholders on the political context level, providing support and service repositories for all interoperability levels.

At a strategic level it is intended to provide means to unify, standardize and deliver functions needed by all participating and potential stakeholders, as well as to gather and disseminate information and knowledge that can generally speed up interoperability development among the target group.

At an operational level it is intended to develop functions to support interoperability harmonization, project deployment and integration of project outcomes in the framework.

The PARENT framework's development and functionality will follow stakeholder requirements and priorities. Therefore it directly depends on the level of participation and involvement of participating stakeholders.

3.2 Registry interoperability guidelines

3.2.1 General

These guidelines primarily focus on providing initial orientation interoperability recommendations intended to aid patient registry stakeholders in grasping their interoperability environment, potential and issues, building own interoperability capacity and participating proactively in PARENT activities.

The guidelines contain two viewpoints:

1. the stakeholder viewpoint focuses on interoperability issues which might be directly faced, and
2. the PARENT framework viewpoint, showing how the framework is envisioned to provide interoperability environment to PARENT participants, stakeholders and users.

The guidelines structure follows the interoperability structure as described in the eHealth EIF document.

3.2.2 The political (stakeholder) context

The political context, once well defined, is a simple and powerful overview tool. It can be compared to a letter of intent, defining a common goal, participants and their responsibilities in a multi-stakeholder development initiative.

The political context defines general initiative goals and measures. All later outcomes of the interoperability process on each level should be checked and verified against the context.

The political context can be defined in various ways:

- it can be dictated by public legislation, strategies and planning,
- it can be defined through project or taskforce initiation documents,
- it is proposed by an initiative leader, or
- define it with partners or alone.

Whenever it is required or necessary to participate/lead in one of the listed possibilities it is necessary to assess the position in the context: who are the entities to interoperate with and how does each relationship reflect on the environment and operation?

The context simplifies problem analysis and solution drafting, since the relationship overview helps to detect requirements, differences and interdependencies that define what changes need to be made (or proposed) to achieve interoperability. It is recommended that a relationship overview be created and maintained even for simple situations (2-3 stakeholders). Often it will be quickly found that important stakeholders were initially omitted from the picture or that there are some valuable relationships that were not apparent in the beginning.

3.2.2.1 Context stakeholders

Key interoperability stakeholders are entities whose participation is required to achieve a goal, since it can be done only if mutually interoperable. It is advisable to include all known issue stakeholders in the context, even if at a certain point they are not considered to be essential for the cause in point. Awareness and early inclusion of the full context can help in anticipating or orchestrating situations that might prove critical for success or solution to a broader issue that might arise later. Continuously review any stakeholder list, propose and consult with them to avoid the most common mistake: to omit inclusion of indispensable stakeholders. This often happens with end users.

3.2.2.2 Context maintenance

Before moving on to harmonization by issue's interoperability levels it is important to have at least all key stakeholders conclusively agree on the mutual purpose, commitment, responsibilities and a well-defined scope of the joint undertaking. Lacking a commitment of a key stakeholder means there is a high probability of failure and loss of time and resource investment. In that case it is important to either consider reducing the scope to a level at which all key stakeholders can be on board, or postpone all further activities.

Interoperable development is an iterative process, allowing continuous adjustment and correction at all levels. If an issue emerges that challenges the political context (whether in outer environment, on a political level or on lower levels), stakeholders must jointly review the issue and decide how to handle it. This can result in a context revision that requires revision of all lower interoperability level developments.

3.2.2.3 PARENT framework context

When defining or participating in a political context it is useful to compare it to the PARENT framework context, since it is a prototype of all patient registry contexts. All (current) general stakeholder groups are presented, and generic group names can be replaced with stakeholder names within the required context. It is most probable that a given context does not contain all presented groups, so these can simply be excluded.

The context can actually be used and overlapped for different initiatives and projects. It is very probable that parts will be able to be reused with minor changes or with no changes at all. This is very helpful if there are separate project teams on different projects, as they can easily synchronize and spot possible synergies (reducing and joining development efforts and costs).

A core business stack of information can also be continuously maintained, helping to be ready to join future interoperable projects. In that case each project can start using a common and updated template.

This context also describes stakeholder generic roles in an interoperable system. This might enable anticipation of possible interoperability issues and preparation for them at early project consideration phases.

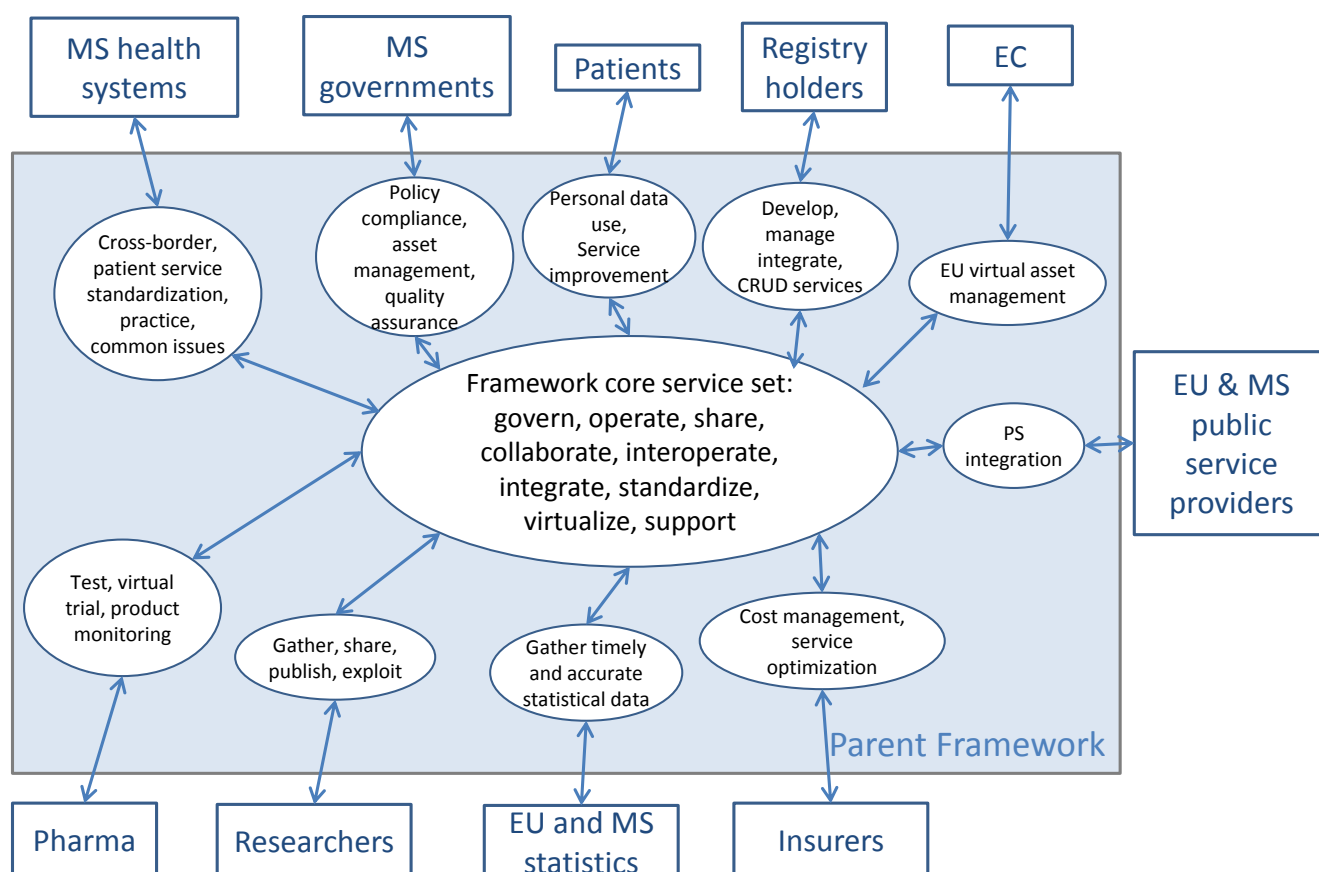


Figure 3.5: PARENT framework

On the operation side it is possible to start building an interoperable service set that fits respective description and that could be universally used in many or all contexts. When the PARENT framework

prerequisites are met the services operation and maintenance to the framework might also be delegated and it might be possible to allow other stakeholders in the group to use them.

The central service set (envisioned to be incrementally provided by the PARENT framework) is a set of services required in all interoperability projects (person-driven or automated). For project risk management purposes actual political contexts should define stakeholder responsibilities for all of them.

3.2.3 Legal interoperability level

Patient registries need to pay special attention to legal issues, since they contain very sensitive personal data, are subject to frequent updates, and support multipoint and multi-stakeholder data exchange.

The first step after agreeing on the project political context is to review it against the legal frameworks of each key stakeholder and the project as a whole. Besides compliance with official legislation (an important issue in cross-border projects due to legislation differences between participant countries) each stakeholder might be affected by particular legal restrictions or obligations (compliance to professional or sector rules, valid contract with other parties, constituent or owner-related issues, etc.).

If any of these presents an obstacle to the project, parties must propose either a reduced project scope compliant with the legal framework or feasible enabling measures or decisions to be approved at the political context level. Otherwise it would be wise to recommend postponement of further activity until these issues are solved or conditions are met or to terminate initiative activities.

An example could be that a registry holder's country legislation forbids cross-border exchange of specific data and a research organization from another country is interested to use the data. If both stakeholders want to achieve this they can work together to define legally acceptable options.

If all key stakeholders can agree on an acceptable initial legal frame that enables project continuation, project harmonization can continue to the next interoperability level.

Registry holders should pay particular attention both to domestic legislation and EU regulations on exchange of registry data. Registries with no previous experience in exchanging data with any but traditional stakeholders, or registries where a part of the registry processes and services are off-line should pay particular attention and start early. We strongly recommend a thorough review of the entire legal context of registries and implications of intended changes.

Special attention should also be given in cases where a registry receives part of its content from other registries, in which case their holders must be included in the political context.

In some cases registry holders must also review legal situations where the intended interoperability solution might affect in any way the content delivered to users (for example where part of the delivered content now comes from other sources).

We encourage registry stakeholders to consult national personal and data exchange agencies, as well as healthcare and social security institutions, as they should have an overview of latest developments in this area.

On the EU level the following should be taken into account:

- Cross Border Healthcare Directive (CBHD)³⁴
- Personal data protection regulation proposal³⁵
- LIBE committee proposal³⁶

3.2.3.1 PARENT framework formal implications

The PARENT framework provides a general communication and harmonization platform for participation of stakeholder legal representatives on general legal review, recommendation and legislation change initiatives. At the beginning the scope will be limited to isolated cases and a task force model.

The PARENT framework will enable tracking of the legal environment in general and through single projects, both as an interactive reference resource and application support. The framework itself might initiate or develop its own rules that would ensure optimization of the framework.

3.2.4 Organizational/process interoperability level

Based on the previously agreed project political context and the legal frame, stakeholders need to review in detail operational responsibilities, roles, outcomes, service and data exchanges. Stakeholders should precisely define these elements for all required project processes, without overlaps or gaps. This includes governance, quality control, and other issues pertinent to smooth, traceable, controllable, uninterruptable and conflict-free process execution, as well as risk management measures.

Organizational and process interoperability is the most complex issue to tackle to achieve smooth interoperability, thus it is hard to generalize or describe in a guidelines format. It can be taken as a rule that each stakeholder has a specific organizational approach, when complexity grows with the level of IT influence on the organization and processes.

Patient registries are used throughout the healthcare process, reflect on numerous national services and report to different EU level institutions. This makes this process extremely sensitive and need thorough analysis of each process step where data creation, recording, usage or exchange takes place. This is the level where previously hidden project or legal issues might emerge. There might also be some unresolvable operational issues (e.g. inability to agree who should be responsible for a process). These should be documented and returned to the appropriate higher level for review and final decision.

³⁴ CBHD: http://europa.eu/legislation_summaries/information_society/data_protection/l14012_en.htm

³⁵ <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52012PC0011>

³⁶ <http://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX:52012PC0010>

An example could be a project where a patient's treatment is carried out in a number of hospitals in different Member States, so different medical records, practices and insurance schemes need to be harmonized.

A pragmatic approach on this interoperability level would be to avoid all cases that would involve complex organizational interventions in existing stakeholder business environments. It is better to develop a business case consisting of an integral set of services that can perform manageable, automatic and traceable functions. The case should reflect joint stakeholder effort to minimise the need for human intervention in the process and reduce cross-system data exchange and transformation.

A more long-term approach would be a full business implementation of SOA³⁷ with EDA³⁸, also known as SOA 2.0.

The business level is also the level where an IT business strategy and financial implications should be agreed on. It is recommended to maximize and optimize the use of existing business and IT resources. It is up to stakeholders to determine which stakeholder will perform which function without jeopardizing functionality and sustainability of the whole system.

In general, use of a standardized business modelling is recommended, and BPMN encouraged, to facilitate articulation of the solution and its communication to the IT level.

When all key stakeholders agree on the acceptable organizational model and business process specification that enables project continuation, project harmonization can continue to the next interoperability level.

3.2.4.1 PARENT framework organizational and process implications

Although at the beginning the framework is not intended to go beyond resource provision and development support, it is actually a potential provider for all shareable organizational and process models and building blocks resulting from successful interoperability projects by stakeholders.

PARENT has developed and provides the following organizational support at framework level (aiding stakeholder organizational leaders in issue harmonization, project execution and deployment, and local application):

- A governance model and services,
- a collaboration model and services,
- a roles and responsibilities model and services,
- a quality assurance model and services.

3.2.5 Semantic interoperability level

Patient registries, particularly in EU cross-border data, multi-lingual and information exchange and sharing, need a careful semantic consideration, analysis and harmonization.

³⁷ Services oriented architecture

³⁸ Event-driven architecture

After agreeing on clear legal and business project definitions they should both pass the stakeholder semantic review and result in full agreement and implementation (where required). The main focus is on these general semantic review areas:

- Processes that generate or transform data exchanged between stakeholders,
- The data that are exchanged,
- Roles and identifiers of stakeholders present in the process,
- Information and instructions for general users given or exchanged on all system access points,
- Information system metadata, data structures and ontologies.

This does not exclude any other aspect of semantic review and acceptance.

3.2.5.1 Standards, models and tools

Since the semantic interoperability is a highly structured, rule and standard-rich segment governing terminology, knowledge, standard interpretation and document interpretation, identifiers, etc. all agreements should aim to be compliant with standards or practices dominantly accepted for a particular area, particularly if determined at EU level.

Naturally, the interoperability process requires an initial assessment of stakeholders' current compliance with semantic standards, models and tools, so users should be able to be acquainted and ready to exchange such information about their system with others and be aware of acceptable alternatives to be able to adjust.

Here is an overview list of key semantic standards, tools and approaches for future reference. Their implementation and use closely depends on particular circumstances.

Metadata

- ISO/CEN Metadata standard 11179
- Dublin Core Metadata

Data structure/exchange

- OpenEHR
- HL7 RIM CDA, C-CDA
- HL7 FHIR
- I2b2
- ISO/CEN 13606
- IHE
- Clinical information modelling initiative

Terminologies

- CTS2 standard
- IHTSDO SNOMED-CT
- ICD10
- LOINC
- ATC
- ICPC-2
- ICF
- ICHI

- DRG

Ontologies

- OWL

Pharma and research

- C-DISC
- BRIDG

Semantic approach

- Archetypes
- Templates

3.2.5.2 PARENT framework organizational and process implications

PARENT should develop and provide the following organizational support at framework level for participation of stakeholder semantic experts. The main areas:

- data harmonization, unification and standards,
- ontologies,
- data integration and reuse rules,
- multilingual support,
- archetypes,
- PARENT dictionary,
- data quality.

3.2.6 Technical interoperability level

This level of interoperability should be reviewed after all previous levels are fully harmonized and defined, since together they represent a detailed system specification. The most important part of the specification comes, from the IT point of view, from the organizational/process interoperability level.

Depending on the agreed business process and responsibilities, there are numerous possibilities regarding use, interconnection and sharing of existing stakeholder IT systems, using cloud capacities, building shared infrastructure, or EU modelled infrastructure, such as Connecting Europe Facilities (CEF), etc.

Before engaging in IT interoperability harmonization assessment of IT system standards in the context of agreements reached on previous interoperability levels is essential. The following should be reviewed:

- The database solution,
- The business application solutions,
- Web technologies used and supported,
- Web portal and interface used,
- Communications protocols supported.

In interoperability projects it will probably be discovered that Patient registries currently operate on highly diverse IT infrastructures, and it would be unreasonable to expect their major modification in

the near future, due to complexity, sensitivity and risk of such action. That's why, in general, project technical interoperability efforts should focus more on solutions which rely on web technology based service and data exchange between existing IT systems wherever possible, in a way that uses existing systems without major modifications.

In each particular interoperability case stakeholders should review and decide on the most convenient suite of standards and protocols that best match their existing systems. All new development should adopt EU initiative models and standards as much as possible. A good reference point is the epSOS project and other references given in the Introduction.

XML is a well-established and universally accepted data exchange format that should be adopted whenever feasible, particularly in mixed health and public stakeholder environments.

HL7 is a data-communication protocol and format for the exchange, integration, sharing, and retrieval of electronic health information that supports clinical practice and the management, delivery and evaluation of health services. As it is especially developed for health systems it should be reviewed as a possible choice in dominantly health-oriented cases.

Interoperability frameworks, such as eHealth EIF and its more general counterpart and predecessor EIF Annex II (4) give models for IT implementation of interoperable solutions. The development and implementation of new IT systems, as well as for more advanced cases, should be founded on these models.

3.2.6.1 PARENT framework technical implications

PARENT framework fully implements the eHealth EIF and EIF Annex II in the SOA 2.0 environment, including the service model and adequate development and operational structure.

The technical implementation of its componentized model (fully compatible with the EIF service model) should provide continuous quick development, sharing and reuse of PARENT service IT solutions, reducing efforts in technical interoperability harmonization and development. Each component is a set of dedicated non-redundant services that comprise the whole framework service portfolio.

Incorporation of every interoperability project into the PARENT framework will augment the PARENT stakeholder service portfolio and reduce the need to develop new IT solutions for each new interoperable business case.

Actual PARENT framework IT environment will implement EU regulations, guidelines project results, key standards and technologies and take into account actual technical status, ensuring user and framework interoperability with systems and projects out of PARENT boundaries.

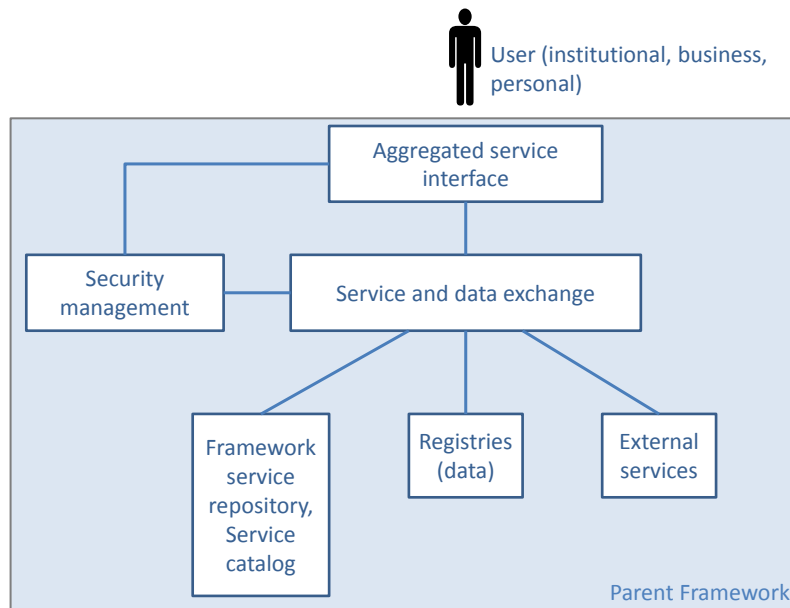


Figure 3.6: PARENT framework technical infrastructure model

References

1. eHealth Action Plan 2012-2020 – Innovative healthcare for the 21st century. Available at: http://ec.europa.eu/information_society/newsroom/cf/itemdetail.cfm?item_id=9156
2. eHealth Governance Initiative Discussion paper on semantic and technical interoperability, page 2-3. Available at: <http://www.ehgi.eu/Download/eHealth%20Network%20%20Paper%20-%20eHGI%20Discussion%20Paper%20Semantic%20and%20Technical%20Interoperability-2012-10-22.pdf>
3. eHealth European Interoperability Framework. Available at: <https://ec.europa.eu/digital-agenda/en/news/ehealth-interoperability-framework-study-0>
4. European Interoperability Framework. Available at: http://ec.europa.eu/isa/documents/isa_annex_ii_eif_en.pdf

4 QUALITY DIMENSIONS OF REGISTRIES

*Marko Brkić, Borna Pleše, Vanja Pajić, Ivan Pristaš,
Ranko Stevanović, Mladen Kostešić, Tamara Poljičanin,
Dmitri Wall, Barbara Foley, Metka Zaletel, Marcel Kralj*

The primary dimension of registries' quality is the quality of the data. Data quality is influenced by a number of other identifiable registry features. Four basic categories of factors influencing registry's quality are:

- Governance, as an organizational foundation of patient registries, is mostly concerned with guidance and decision making. Adequate governance model makes sure to address issues such as overall direction and operations (procedures and processes), communication, scientific content, ethics, safety, data access, transparency, publications, change management and registry life-span planning.
- Data quality is assured by defined requirements/standards for data collection and management. Data quality is also to be assessed against a list of dimensions which can be defined and measured.
- Information quality is an output of a data collection process. It is measured by the amount and impact of scientific publications based on registry data.
- Quality is also influenced by features like confidentiality, security, privacy and ethical issues. These influence a registry's interoperability capability and information dissemination. Meeting ethical and legal requirements concerning privacy influences registry's interoperability capability and information dissemination. Privacy component of the registry is measured by privacy impact assessments (PIAs).

Integrally addressing advices indicated within stated categories during registry planning and creation but also while running a registry, should ensure high level of registry performance.

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 (1)”. In that light, other quality dimensions can also be (and should be) assessed. This way data quality remains the primary dimension within registry quality evaluation, but acknowledging that it is influenced by other identifiable registry features. Based on such a rather holistic view and through conducting a literature review we have identified numerous “quality influencing factors” and categorized them into four groups, which are not to be viewed separately. These are: 1) governance; 2) data quality; 3) information; 4) ethical issues, security and privacy (Figure 4.1). It is useful to consider these categories while planning and evaluating registries, since they should, rounded all together, provide a rough estimate basis for assessing registry performance.³⁹

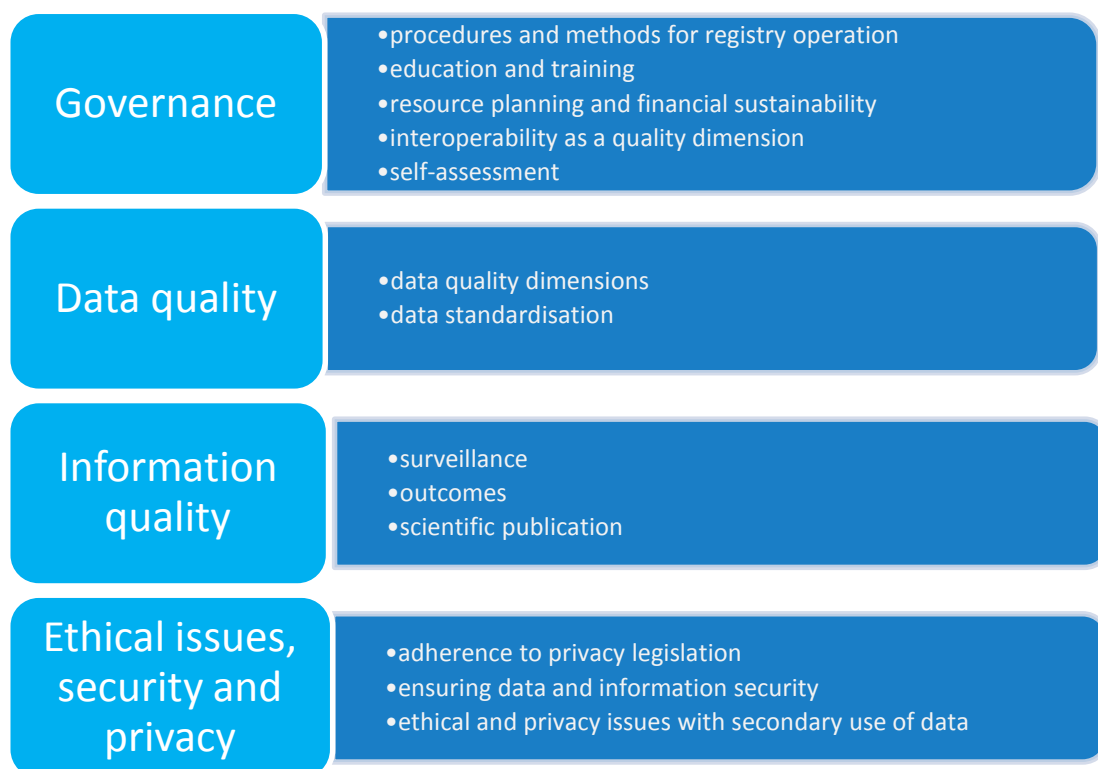


Figure 4.1: Quality dimensions of registries

4.1 Governance

Governance and management are the organisational foundations of patient registries, by:

- providing the framework to ensure that the registry achieves objectives set on its establishment
- driving the registry’s functioning in terms of securing resources (financial, human, technical), measuring performance and ensuring sustainability
- influencing data quality and registry outputs regarding dissemination of information
- complying with legal pre-requisites

³⁹ Several other quality frameworks have been defined by different users’ groups. At this point, one should note the European Statistical System’s Quality Assessment Framework, available at <http://ec.europa.eu/eurostat/documents/64157/4392716/ESS-QAF-V1-2final.pdf/bbf5970c-1adf-46c8-afc3-58ce177a0646>.

Applying proper governance principles should ensure that robust operational procedures and processes are in place, clearly communicated, and easy to access for everyone involved in data collection. Besides basic managerial and operative functions, the goal of apt governance should also be transparency to stakeholders in operations, decision making, and finally in reporting of results.

Governance is thus mostly concerned with guidance and decision making, which include the topics of registry concept, funding and dissemination of information. A governance plan is important at a registry's onset as it substantially determines future functioning. Therefore, the plan for registry governance and oversight should clearly address issues such as overall direction and operations, scientific content, ethics, safety, data access, publications, and change management. It is also helpful to plan for the entire lifespan of a registry, including how and when the registry will end and any plans for transition at that time (2). Specific elements of the governance quality dimension are presented below.

4.1.1 Procedures and methods for registry operation

In order to justify the holding of personal health data, the establishment of a registry first requires a clearly stated purpose. The stated purpose should contain a brief description of why the registry is established and what its intended use is (e.g. program administration, service delivery or research). This purpose should also be subject to review and change should the objectives and aims of the registry change. A statement of purpose should also contain information such as: full (legal) name of the registry, contact details, name of responsible registry holder, year started, overall function, objectives, list of data providers, legal basis for establishment, legislation and standards (privacy, national, international) that the registry must adhere to (3).

Determining the appropriate scope of the registry, data set and target population, along with a study plan or protocol is fundamental to proper data collection and to the future quality functioning of the registry. At a registry's outset proper documentation managing should be upheld, meaning that the goals of the registry, its design, target population, all procedures related to data (methods and procedures for data collection, clearly defined data elements and items, data management, appropriate data analysis and reporting practice procedures), and how human subjects will be protected (privacy legislation) should be documented.

It is very important for a patient registry to have a complete and detailed manual containing descriptions of protocols, policies, structures and procedures. Documenting registry policies and procedures enables it to become more process dependent than person dependent, potentially enhancing data quality stability and reliability. Document management should be an active process, maintaining and updating documentation through the registry's further operational period. One more feature closely linked to document management is the registry's overall adaptability, as technical, regulatory and ethical frameworks of the registry should be periodically reviewed in order to address possible newly emerging issues.

4.1.2 Education and training

Staff education and training is another important aspect of registry quality. Inadequate registry staff training may cause data quality issues as well as security breaches and/or privacy violation. Sufficient staff qualification and training is thus necessary, and this can be achieved through training and education. All staff should receive training and education relating to their roles and specific job

responsibilities, as well as proper training on the patient registry protocol and procedures, data sources, data collection systems and data definitions (with interpretation), accompanied by formal records of training and education (2).

For instance, if registry governance decides on applying standards, this does not all by itself necessarily lead to enhancing any of the registry quality dimensions. Such action also demands achieving a certain satisfactory level of education and training of registry staff, in order to ensure a straightforward implementation process of standards. For example, the ICD-10 terminology, depending on the purpose of use (cause of death, cancer, discharges, DRG, infectious diseases), requires appropriate levels of education and training fit for and according to purpose of use.

The registry governance should have a training plan through which refresher training is to be provided on an ongoing basis. Training content should also regularly be subjected to review and updates, following potential changes in legislation, and national and international standards (3).

Training is not only important for registry staff, but also for the staff of the healthcare unit which provides data for the registry, in order to increase data quality. Training includes various methods, from providing manuals for proper data collection and data dictionaries to organizing training sessions with data providers (clinicians etc.) as participants where e.g. data extraction guidelines are discussed and practised with patient cases (4).

4.1.3 Resource planning and financial sustainability

Since achieving objectives relies foremost on available resources (human, physical, financial), the managing organisation responsible for the registry should plan and manage its resources to ensure that they are used efficiently and effectively. Resources should be adequate to ensure the sustainability, continual relevance and maximum impact of the data for which the registry holders are responsible. Considering that budgets are limited, careful planning and management of the use of resources is crucial to ensure they are used in the most efficient, useful and effective manner. How resources are used very much influences the quality of the information provided and the future sustainability of the registry. The allocation of resources is therefore also a fundamental factor in the delivery of quality data (3). One of the more promising ways to provide financial sustainability is collaboration amongst all the stakeholders involved in the registry, an approach which can reduce or avoid duplication of efforts, foster improved quality and robustness of data collected, and finally, in a positive way, sustain registries as long-term ventures (5).

4.1.4 Interoperability as a quality dimension

Interoperability can be viewed as a quality dimension under the governance group in regard to the following:

Impact on any particular registry quality dimension cannot be pursued only within the registry holder (e.g. institution), as it is also necessary to influence the business processes and *modus operandi* of other registry stakeholders (data sources, identified data users, health information authorities etc.). Ideally, interoperability should be established through a gradual process of connecting internal processes of the stated stakeholders, therefore transforming the business processes towards convergence and making them mutual and public. Interoperability concepts and issues as well as interoperability as an envisioned common goal for patient registries across Europe is discussed further in chapter 3.

4.1.5 Self-assessment

One of the registry governance roles should be to consider how to ensure overall quality to a level sufficient for the intended purposes; therefore registries must pay careful attention to quality assurance issues. Quality assurance is important for any registry to ensure that appropriate patients are being enrolled and that the data being collected are accurate. Quality assurance activities can help to identify data quality issues resulting from inadequate training, incomplete case identification or sampling, misunderstanding or misapplication of inclusion/exclusion criteria, or misinterpretation of data elements and hence improve the overall quality of registry data (2).

Self-assessment should perform quality control and serve to identify the sources of potential data quality issues and assess them by using indicators on data quality dimensions, developing measurements for evaluation, subsequently used to correct issues and track improvements. Use of quality assessments is also recommended to guide any decision on changing or modifying registry practices and procedures. Self-assessment can be an important registry governance feature, as it constitutes in fact a great self-propelling mechanism that ensures continual quality improvement.

Data quality improvements can be based on regular internal data quality audits including the quality of coding that incorporate clinician input (data source) as well as on external audits and external data quality reports. Self-assessment refers to periodically performing quality control through a data quality assurance programme and subsequent instituting of data quality improvements based on identified quality issues. However, self-assessment is here envisaged also as a governance responsibility, which should concern not only data quality checks but also overall registry functioning.

4.2 Data quality

In addition to a full understanding of study design and methodology, analysis of registry events and outputs requires an assessment of data quality. Requirements for data collection and quality assurance should be defined during the registry creation phase, and following the "collect once, use many" rule of data collection and management, it is paramount that the data are of sufficient quality, as the information and subsequent use for multiple potential purposes are all derived from that initial data. 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 (6). High-quality data are then data that are fit for use by data consumers, data that have sufficient usefulness and usability. This fact leads to viewing data quality as having many attributes, or in other words data quality is presented as a complex multidimensional concept.

4.2.1 Data quality dimensions and its assessment

Determining the quality of data is possible through data assessment against a list of dimensions which can be defined and measured. Data quality dimensions can be defined as a "set of data quality attributes that represent a single aspect or construct of data quality" (7). The dimensions are organized in a data quality framework, which attempts to capture all aspects of data quality that are important to data consumers.

Deciding on a list of quality dimensions is mainly dependent on the patient registry context (nation and/or region specific provisions, legal obligations etc.), type and purpose. When defining a data

quality framework, in order to ensure subsequent appropriate measurements of data quality, the developer should take care to include all the context relevant data quality dimensions.

A large number of distinct data quality attributes that might determine usability through literature review have been identified.⁴⁰ Most of the data quality dimensions were overlapping and had different interpretations, often with ambiguous definitions or completely lacking definitions, while the two most frequently cited were data „accuracy“ and „completeness“.

Trying to list all internationally used data quality dimensions and include their interpretations would prove a futile effort. Thus, the underlying principle for deciding on these dimensions and arranging them into a meaningful whole was providing comprehensive coverage while keeping dimensions organized in a collectively exhaustive way. Mutual exclusiveness was desired but is hardly achievable at the general level of description. Here is proposed a set of six data quality dimensions (Table 4.1 'Data quality dimensions').

Table 4.1: Data quality dimensions

Data quality dimension	Description
1 Accuracy	<p>How well information in or derived from the data reflects the reality it was designed to measure (11). It is usually characterized in terms of error in statistical estimates. It may also be described in terms of the major sources of error that potentially cause inaccuracy (e.g., coverage, sampling, non-response, response) (12).</p> <ul style="list-style-type: none"> ✓ How good are the data? ✓ What is done with the data?
2 Completeness	<p>Extent to which all necessary data that could have registered have actually been registered (6). It is usually described as a measure of the amount of available data from data collection compared to the amount that was expected to be obtained (e.g. coverage) (13).</p> <ul style="list-style-type: none"> ✓ Are all the appropriate data present?
3 Interpretability and Accessibility	<p>Ease with which data may be understood and accessed (11).</p> <p>This includes the ease with which the existence of information can be ascertained, the suitability of the form or medium through which the information can be accessed, whether data are accompanied with appropriate metadata and whether information on their quality is also available (including limitation in use etc.) (12).</p> <ul style="list-style-type: none"> ✓ How readily accessible are the data? ✓ How well documented are the data? ✓ How easy is it to understand the data?
4 Relevance	<p>The degree to which data meet the current and potential needs of users. The purpose is to assess how well data collection can adapt to change and whether it is perceived to be valuable (11).</p> <ul style="list-style-type: none"> ✓ Can user needs be anticipated and planned for? ✓ How valuable are the data?

⁴⁰ For a more detailed summary of the internationally commonly used data quality dimensions refer to a publication from HIQA: International Review of Data Quality. Dublin: HIQA, 2011. Available at: <http://hiqa.ie/press-release/2011-04-28-international-review-data-quality>

5 Timeliness	<p>Refers primarily to how current or up to date the data are at the time of release, by measuring the gap between the end of the reference period to which the data pertain and the date on which the data become available to users (11). It is typically involved in a trade-off against accuracy. The timeliness of information will influence its relevance (12).</p> <ul style="list-style-type: none"> ✓ Are data made available in a reasonable amount of time? ✓ Are key documents released on time?
6 Coherence	<p>Reflects the degree to which it can be successfully brought together with other statistical information within a broad analytic framework and over time. Coherence covers the internal consistency of data collection as well as its comparability both over time and with other data sources (14). The use of standard concepts, classifications and target populations promotes coherence, as does the use of common methodology across surveys. Coherence does not necessarily imply full numerical consistency (12).</p> <ul style="list-style-type: none"> ✓ Does the database use standard definitions for data definitions? ✓ Can common groupings be derived from the data? ✓ Can databases be joined via a common data element? ✓ Are data values being converted correctly? ✓ Are data comparable with themselves over time?

The dimensions provided in the table are applicable to different registry types (and with different objectives), however not all may be equally important.

Assessing quality includes adequate management of each dimension, and additionally failure in one dimension can severely hinder the usefulness of the final registry report (i.e. when considering cancer registries insisting on the dimension of data completeness may ruin the demand for timely reporting). Likewise, each of the dimensions may possess equal importance, but also there may be instances where the relative importance of one dimension exceeds another. As stated previously, the importance of a particular quality dimension depends on the set objectives of the registry, its type, as well as its scope and methodology. Specifically, based on the definition of data quality provided above, the intended use of registry data actually determines the necessary properties and requirements of the data.

For instance, in a registry that is used to calculate incidence rates of diseases, it is essential to include all existing patient cases, therefore the completeness dimension is of critical importance.

Additionally, the need to explore different aspects of data quality is an accepted practice among patient-registries, and should be accentuated when not present.

For example, population-based cancer registries are considered particularly attentive to assessing data quality, as the value of the modern cancer registry and its ability to carry out cancer control activities rely heavily on the underlying quality of its data and the quality control procedures in place (8).

Data quality regarding cancer registries is usually assessed against the following three quality dimensions: comparability, validity, completeness, as well as sometimes timeliness as a fourth one. Factors influencing data quality and methods (both quantitative and qualitative) for measuring data quality within these dimensions have been devised and made available.⁴¹

⁴¹ Reviews of these methods are presented in more detail in (8) and (15).

Data quality dimensions are components that allow the user to quickly identify specific problematic aspects of data. Interrelatedness and overlapping are always necessarily present; the quality dimensions are not specific to quality measuring, and for that to be possible, as exemplified by cancer registries, decisions are needed to identify which methods and indicators are to be used in order to successfully measure registry data quality against dimensions. The data quality assessment programme should thus precisely define a data quality framework, preferably logically grouping what should be measured and how it should be measured and monitored in the data domain, thus making data dimensions more specific by creating data characteristics and criteria, along with a rating method. Such an example of a comprehensive method for assessing data quality is the Data Quality Framework (11), by the Canadian Institute for Health Information (CIHI) issued with the purpose of improving data quality of national health data collections. The Data Quality Framework is based on Statistics Canada guidelines and methods and information quality literature. It is a highly developed hierarchical framework model, with established criteria useful for systemic data quality assessments.

In summary, efforts should be made to create various relevant data quality dimension groups dependent on type and objectives of the registry, and devise methods and indicators for assessing data quality, so that a registry can use those methods to measure and gradually improve data quality.

4.2.2 Mode of data collection and impact on data quality

Considering data quality as part of a complex whole brings out another important and often neglected aspect which can influence data quality – the point where data are collected.⁴²

The quality of initial data input from clinicians and health practitioners can vary. Quite frequently incorrect patients are registered or data items can be inaccurately recorded or not recorded at all.

A sustainable workflow model is an important element of a successful registry, a workflow that can be integrated into the everyday clinical practice of doctors, nurses, pharmacists, and patients (while respecting privacy legislation). Prior to the full launch of a registry, pilot testing can be organized to gather preliminary input from health care workers and others included in the data collection.

A decision should be made on the mode of data collection, as there are a few ways to collect data, where the primary difference is whether it is collected in its conventional paper form or the modern electronic form.⁴³

4.2.3 Improving data quality

Since data quality is critical to any registry, a patient registry should seek to implement and maintain a high standard in all of the quality dimensions identified here of patient registries (governance, data quality, information quality, ethical issues, security and privacy). The governance dimension is crucial here (as discussed in subchapter [4.1.5](#)), as the initiative within an organisation to improve data quality is driven by managerial decisions, setting forth standards and channelling staff efforts. In this light, the Health Information and Quality Authority of Ireland (HIQA) describes “seven essentials for improving data quality”⁴⁴, which it is useful to consider in the context of a patient registry. These essential features are presented in the table below.

⁴² This issue has been briefly discussed in subchapter [4.1.2](#), [6.1.2.1.3](#), [6.2.4](#), [6.4](#) and [9.1](#)

⁴³ Methods of data collecting (paper or electronic) are discussed in subchapter [6.1.2.1.3](#).

⁴⁴ Health Information and Quality Authority. What you should know about Data Quality. Dublin, Ireland: HIQA, 2012.

Table 4.2: Essentials for improving data quality

Leadership & Management	<ul style="list-style-type: none"> •What: involves having in place executive-level responsibility, accountability and leadership. •Why: knowing who does what (e.g. the establishment of a governance committee that will ensure the registry is committed to data quality). Decision-wise, this includes the selection of only essential data elements when datasets are established.
Policies and procedures	<ul style="list-style-type: none"> •What: developing and implementing clear policies and procedures on data quality for staff that are based on legislation and standards. •Why: can help ensure that a high level focus on data quality is translated into good practice amongst all those involved in data collection and handling within the registry.
Standardisation	<ul style="list-style-type: none"> •What: ensuring that data are collected and processed in a standardised fashion (e.g. use of minimal datasets, data dictionaries and the creation of standard templates for data collection), designing the registry with respect to national and international standards. •Why: facilitates data interoperability and making data available. Also can improve consistency and reduce error.
Data quality dimensions	<ul style="list-style-type: none"> •What: set of data quality attributes upon which data can be assessed, aligned with policies, procedures and training. •Why: measuring and monitoring level of data quality within a registry.
Training	<ul style="list-style-type: none"> •What: training of the staff in the requirements and importance of data quality. •Why: ensuring that policies and procedures adopted to generate high quality data are implemented and understood in practice.
Data quality audits	<ul style="list-style-type: none"> •What: independent systematic examination of data (internal or external). •Why: providing feedback to all staff, indicating the areas for improvement, highlighting good practice in order to facilitate learning (e.g. automation of data collection over manual collection where possible will reduce error rate, however, this will not be verified without planned audits of data quality).
Make data available	<ul style="list-style-type: none"> •What: availability of data when and where needed, in accordance with information governance safeguards (security, privacy). •Why: fulfilling the purpose for which the registry was created, increasing quality of registry data through its efficient utilization and dissemination.

4.3 Information quality

Information can be considered an output and the extension of the data collection process. Its quality is measured by the purpose of its use, which in the case of patient registries can be grouped into surveillance (including health statistics), outcomes, and scientific publication.

Scientific publication can be considered as a control for methodological prerequisites including sufficient level of data quality. Therefore, it can serve as an indirect information quality indicator. Levels of measuring can be publication amounts (total, yearly), subject relevance, up-to-date, impact factor, citation index.

Similarly, statistical data from registries focused on surveillance can be used as an indirect quality measure with regards to real-world decision making. Outcome based registries serve the same purpose in terms of indirect quality measurement albeit from a different viewpoint, i.e. using information from patient registries for influencing and improving treatment outcomes. Quality information gained from patient registries leads to informed healthcare management and better decision making.

4.4 Confidentiality, security, privacy, ethical issues, secondary use of information

This quality dimension is concerned with ethical issues and confidentiality and privacy regarding use of personal health information, as well as the need for proper patient registry data security and clear provisions regarding secondary use of information. Although actually concerning data and stemming from the wider dimension of (information) governance, it is here discussed separately as it involves privacy protection, a sensitive and seminal issue when discussing patient registries.⁴⁵

Not meeting ethical and legal requirements concerning privacy renders the patient registry inoperable. Levels of data confidentiality, privacy and security also influence registry interoperability capability as well as information dissemination.

Creating a balance between respecting individual privacy and providing high quality personal health information can, although very important, also be a difficult task faced by patient registries as well as other healthcare related stakeholders. Striving for cross-border interconnecting and interoperability of patient registries is accompanied by emerging security risks concerning privacy, judging by the fact that health information systems present technical challenges to existing privacy protection legal frameworks.

In order to maintain the privacy of participants enrolled in a registry and the data confidentiality, security measures should be implemented. All security measures should be contained in a document that describes in detail the data security risks, policies, and procedures specific to that registry. Physical and technical safeguards should be incorporated in the collection, storage, transmission of and access to data. These include data encryption, restriction of data access, data back-ups, methods (software) for de-identification of local data during potential transmission and storage etc. Also, implementation of safeguards should not be done only once, but should undergo continuous review and revision.

Considering data usage, we can distinguish between two types: 1) primary purpose; 2) secondary use of data.

⁴⁵ Privacy, confidentiality and security are mentioned in more detail in chapters [5](#) and [6.1.4](#).

This classification as primary or secondary is based on the relationship of the data to the registry purpose. Primary data sources include data collected and being kept by the registry holder (custodian) for direct purposes of the registry (i.e., primarily for the registry). The secondary use of health data considers uses for purposes other than those for which it was originally collected. Secondary uses include using information for (further) research, performance monitoring, service planning, audit and quality assurance purposes etc. When thinking about the secondary use of health data, it is necessary to carefully balance between the public interest and individual data subjects. Since secondary use of data may violate patient privacy, precautions should be taken and conditions must be satisfied if proposing to use information for secondary purposes. Clear definitions of the circumstances where data are to be used for secondary reasons should be developed.

Legislative provisions concerning the secondary use of data are typically contained within general privacy or data protection legislation, which can differ depending on the specific MS context.

The important things with secondary data use are that patients should be made aware that their information may be used for this purpose and have the benefits of the practice clearly explained to them. Likewise, consent must be obtained for the collection, use or disclosure of information for purposes outside the direct registry's data outline plan. Efforts should be performed to make data anonymous as well as using data sharing agreements which offer an additional safeguard against inappropriate use of information.

To repeat and to conclude the subchapter, researchers and other data users should disclose clearly how and why personal information is being collected, used, and secured, and should be subject to legally enforceable obligations to ensure that personally identifiable information is used appropriately and securely. In this manner, privacy protection will help not only to ensure research participation, public trust and confidence in medical research, but also prompt cross-border registry cooperation. If registry holders are confident that their information is being appropriately protected and have trust in the system, then they are more likely to share information, which leads to improved safety and quality of care at an individual level.

4.4.1 Privacy impact assessment (PIA) – a method to assess privacy

A privacy impact assessment (PIAs) is a tool, process or method to identify, assess, mitigate or avoid privacy risks (9). PIAs are used internationally and across all sectors but are particularly useful to healthcare providers to identify potential risks around the collection and use of sensitive personal health information. PIAs can help respond to the new privacy challenges in the design of cross-border health information systems. The primary purpose of undertaking a PIA is to protect the rights of service users. The process involves the evaluation of broad privacy implications of projects and relevant legislative compliance, through describing how data are collected, processed, disseminated and published. Where potential privacy risks are identified, a search is undertaken, in consultation with stakeholders, for ways to avoid or mitigate these risks and to facilitate solutions which help safeguard privacy. As PIA considers the future privacy consequences of a proposed project that involves the collection and use of personal health information, it is most beneficial when conducted in the early stages of a project, and ideally at the planning stage (3).

Related with the goals of the PARENT project, a very useful PIA initiative has been identified with the EUBIROD project. EUBIROD explored privacy issues at the level of systems' users, assessing the variability of data processing approaches in MS and their deviation from EU privacy standards and

legislation, and by using the adapted version of the Canadian PIA Guidelines. Key elements of data protection (factors) were selected to ascertain the compliance/non-compliance with privacy principles/norms of data processing operations occurring in EUBIROD registries.

Registry privacy and data protection which should be investigated when conducting PIA are:

“accountability of personal information”; “collection of personal information”; “consent”; “use of personal information”; “disclosure and disposition of personal information”; “accuracy of personal information”; “safeguarding personal information”; “openness”; “individual access to personal information”; “challenging compliance”; “anonymisation process for secondary uses of health data” (10).

References

1. Kohn LT, Corrigan JM, Donaldson MS, eds. To Err Is Human: Building a Safer Health System. Institute of Medicine, Washington: Committee on Quality of Health Care in America, 2000.
2. (AHRQ) Registries for Evaluating Patient Outcomes: A User's Guide, 3Ed, Volume 2. In: Guide, editor, 2012.
3. HIQA) Guiding Principles for National Health and Social Care Data Collections. Dublin: HIQA, 2013. Available at: <http://www.hiqa.ie/publications/guiding-principles-national-health-and-social-care-data-collections>
4. Arts DGT, Bosman RJ, de Jonge E, Joore JCA, de Keizer NF. Training in data definitions improves quality of intensive care data. *Critical Care* 2003;7:179-184.
5. EURORDIS-NORD-CORD Joint Declaration of 10 Key Principles for Rare Disease Patient Registries. Available at: http://download.eurordis.org/documents/pdf/EURORDIS_NORD_CORD_JointDec_Registries_FIN_AL.pdf
6. 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.
7. Wang, RY, Strong DM. Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*. 1996;12(4):5-33.
8. Bray F, Parkin DM. Evaluation of data quality in the cancer registry: principles and methods Part I: comparability, validity and timeliness. *Eur J Cancer*. 2009;45(5):747-755.
9. (NHS) Privacy Impact Assessment: care.data. Chief Data Officer, NHS England, 2014. Available at: <http://www.england.nhs.uk/wp-content/uploads/2014/04/cd-pia.pdf>
10. Privacy Impact Assessment Report. The EU.B.I.R.O.D. PIA Team, 2010. Available at: http://www.eubirod.eu/documents/downloads/D5_2_Privacy_Impact_Assessment.pdf
11. Canadian Institute for Health Information, The CIHI Data Quality Framework, 2009 (Ottawa, Ont.: CIHI, 2009). Available at: http://www.cihi.ca/CIHI-ext-portal/pdf/internet/data_quality_framework_2009_en
12. Statistics Canada Quality Guidelines, Fourth Edition, October 2003. Available at: <http://www.statcan.gc.ca/pub/12-539-x/12-539-x2003001-eng.pdf>
13. Ehling, Manfred, Körner, Thomas. (eds.) Handbook on Data Quality Assessment Methods and Tools Eurostat, European Commission, Wiesbaden, 2007. Available at: <http://unstats.un.org/unsd/dnss/docs-nqaf/Eurostat-HANDBOOK%20ON%20DATA%20QUALITY%20ASSESSMENT%20METHODS%20AND%20TOOLS%20%20I.pdf>
14. Australian Bureau of Statistics. ABS Data Quality Framework [Online]. Available at: <http://www.abs.gov.au/AusStats/ABS@.nsf/Latestproducts/5AFFD020BC4D1130CA25734700151AA5?opendocument>
15. Parkin DM, Bray F. Evaluation of data quality in the cancer registry: Principles and methods Part II. Completeness. *Eur J Cancer* 2009; 45:756-64.

5 GENERAL REQUIREMENTS FOR CROSS-BORDER USE OF PATIENT REGISTRIES

*Persephone Doupi, Arto Vuori, Katariina Peltonen,
Antti Tuomi-Nikula, Haralampos Karanikas, Yannis Skalkidis*

Key messages:

- In the PARENT JA vision, the cross-border use of registries is predicated on a continuous IT-assisted chain of health data capture, storage, processing, transmission and utilization.
- Fulfilling the prerequisites for cross-border operations essentially means achieving interoperability in the broadest understanding of the term, i.e. on legal, organizational, semantic and technical levels as well as the establishment of effective, sustainable solutions for cross-border registry collaboration.
- Policy context: Creation, maintenance and development of registries, as well as preparedness for cross-border operations is largely dependent on the positioning of health data resources in national strategic prioritization for scientific data resources and research infrastructures. Equally important is the question of whether registries are perceived as part of regional and/or national eHealth infrastructure.
- Organisational aspects: Transparent procedures for granting access to or sharing data in a cross-border context must be in place, preferably including predefined response time targets.
- Legal and ethical aspects: a patient registry can be established using either of two legal instruments; by explicit consent of the data subject, or based on law. Adoption of a consent model presumes thorough planning of the purposes of the registry and consultation with local data protection authorities or ethical committees.
- Semantic aspects: comparability and transferability of health data across languages and contexts of use is heavily dependent on the adoption and use of accepted coding standards.
- Technical aspects: crucial in ensuring that health data are shareable; hence adopted solutions must support or be compatible with regional/national infrastructures and semantic requirements for patient data collected in the process of healthcare services provision.

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** (e.g. 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 for studying differences between countries in morbidity, effectiveness of health system-level interventions and utility of procedures; the advantages of large international datasets vs. national ones in the timely detection of rare, or previously unknown effects; gathering and promoting information on best practices worldwide.

Independent of the motive driving the utilization of cross-border registries, the success of the endeavour will always rely on the degree of achievement of certain prerequisites, the implementation of which starts at a local level - regional and/or national. PARENT Joint Action aims at the idea of establishing a continuous IT-assisted chain of health data capture, storage, processing, transmission and utilization. Therefore 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 (see chapter 3) as well as the establishment of effective, sustainable solutions for cross-border registry collaboration. 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 10. An exception is the topic of metadata, which we briefly discuss here. A more detailed analysis of organizational interoperability aspects with regard to stakeholders and their roles constitutes part of the business models analysis of PARENT (D6.2. – forthcoming), while policy aspects and necessary actions are discussed in the respective deliverable (D6.1. – forthcoming). It should be kept in mind that the allocation of requirements to respective interoperability aspects is at least to some extent artificial. Several requirements span many if not all levels of interoperability, even if they are discussed under a predominant heading.

5.1 Political context

The creation, maintenance and development of registries, as well as their preparedness for cross-border operations is largely dependent on the positioning (or lack thereof) of health data resources in national strategic prioritization regarding *scientific data resources and research infrastructures*. PARENT has analysed in a parallel activity national strategies and initiatives concerning Health Data Strategies and the ways in which they impact patient registry work. A brief sample of the findings is provided in [Table 5.1](#).

Equally important is the question of whether registries are perceived as part of a regional and/or national *eHealth infrastructure*. At EU level, Member State collaboration in the field of eHealth has until now focused primarily on the creation and exchange of health data at the point of and for the purposes of patient care, as reflected in the work of the eHealth Network on ePrescription and Patient Summary (1, 2). The needs and requirements of secondary use of data, where the formation and utilization of registries also belong, have until recently remained unexplored. However, in order to achieve the vision of electronic collection, processing and re-use of health data throughout its lifecycle while ensuring the fulfilment of interoperability requirements, e-enabled registries need to be included as a target of national eHealth agendas, thereby establishing the link with ongoing EHR initiatives in Member States.

5.2 Organisational aspects – Registries' operations and procedures

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 or access rights. Each one of the steps in this process can take a variable length of time and incur widely differing costs, depending on the registry holding the data in question. Both elements though may turn into considerable barriers, particularly from the perspective of socially and politically urgent research or regulatory work. New solutions for more straightforward application processes and remote access to data are being developed.

Procedures for granting access to or sharing data in a cross-border context must be in place, preferably including predefined response time targets. An organizational culture oriented towards data sharing, as well as appreciation of data utilization beyond own organizational remits, combined with appropriate resourcing are essential elements in achieving a high level of preparedness. Collaboration with other registries' holders is advisable, in order to exchange experiences, advice and ideas.

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.

5.3 Legal and ethical aspects

Utilizing patient registry data is an asset for science and the patient and health care sector. However, the right to privacy and data protection are fundamental rights (European Charter of Fundamental Rights). Therefore, securing the privacy of the patients or research subjects is an essential task when establishing and maintaining a patient registry or when conducting registry-based research. Moreover, when processing personal data, the data controller has to take into account not only legal, but also ethical perspectives.

The most important European law affecting patient registries' operations is the Data Protection Directive (95/46/EC) (4) 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. Currently the implementations and interpretations of the Data Protection Directive vary between Member States. Similarly the roles of Data Protection Authorities and Ethical Committees differ greatly. The legislative process toward the new harmonizing Data Protection Framework is still ongoing. At the same time, the European Union Directives and Regulations considering Medical Devices, 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 adoption of a consent model assumes a thorough planning of the purposes of the registry. The required content of informed consent varies between Member States. For this reason it is important to consult local Data Protection Authorities or ethical committees in the process of formulating the consent model. The Opinion of the European Data Protection Working Party (WP 29) regarding the definition of consent (WP 29, 2011) (5) provides a thorough analysis of the concept as currently used in the Data Protection Directive.

According to the existing Data protection Directive Article 2.h “the data subject’s consent” means any freely given, specific and informed indication of his/her wishes by which the data subject signifies his/her agreement to personal data relating to him/her being processed. The definition of the Directive implies an opt-in strategy of the consent. For the legal protection of the registry holder and patient, it is advisable that the consent is given in written form. If personal data are transferred abroad, this should be communicated in the context of acquiring informed consent.

As noted earlier, by and large a patient registry can be established either by the explicit consent of the data subject, or based on law. The explicit consent, where the data subject approves the processing of his/her personal data, is the primary instrument (if indeed it is possible to ask the consent of the data subject), since the protection of one’s privacy is the fundamental right of the data subject. However, there are no right answers to when it is possible to ask for the subject’s consent and when not. When it comes to small, disease-specific registries, in principle the consent of the data subject is the legal instrument for establishing and maintaining the registry. For example, this might be the case in rare diseases registries. If the data subject has a doctor-patient relationship with the representative of the data controller, what circumstances would justify not asking the consent of the data subject?

The ethical considerations of establishing a registry are far more complicated, if it is uncertain whether or not the data subjects are capable of giving informed consent. However, these situations are case-specific and require in-depth ethical evaluation. Registry holders should contact the respective local Ethical Committee concerning the ethical issues of registry establishment.

When the establishment of the patient registry is based on law, one should presume that the ethical grounds for doing so are reasonable. The registry holder has to maintain the privacy of the data subjects and ensure that operations abide by the data protection regulation (of course this has to be the case even when establishing the registry using the instrument of an explicit consent). Establishing and maintaining a patient registry based on law may be relevant for administrative, statistical, research and some other purposes. For example, in Nordic countries large population-based patient registries were established by law without the consent of the data subjects mainly for statistical and scientific purposes. However, the processing of this registry data is strictly regulated and the data can be utilized only for specific purposes. When the registry is established by law one has to assume that the benefits of establishing the patient registry are greater than following the principal rule of asking patients’ consents for registration. Using patient registry data for research purposes may prevent the unnecessary collection and processing of personal data.

Even though the European Commission has proposed a Regulation as a substitute for the existing Directive (6), it is also likely that some national legal variation regarding patient registries will continue to exist. These disparities reflect differences in Member States’ national health care systems, information infrastructures and legislations. It should not be assumed that legal and ethical interpretations and practices are identical across Europe. Thus, it is always advisable to consult regional or national Data Protection Authorities or ethical committees when establishing a registry. Recognizing the need for more streamlined processes, cross-border models of operations have also been considered especially regarding ethical committees (Nordic Trial Alliance; Nordforsk) (7). One of

the considered models is to develop procedures for mutual recognition of ethical approvals, in which approval in one country would be valid in the other countries, too in cases of cross-border research co-operation. However, these are still aims for the future and do not reflect the current situation. Securing privacy of the research subjects is a fundamental task when establishing and maintaining a patient registry. It has become a generalized interpretation 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.

According to the Data Protection Directive, *personal data* means any information relating to an identified or identifiable natural person ("data subject"). An identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity.

The current Data Protection Directive does not define the often used concepts of pseudonymous data or encrypted data. One way to define *pseudonymous data* (According to the European Parliament's proposal given in March 2014), is that it means personal data that cannot be attributed to a specific data subject without the use of additional information, as long as such additional information is kept separately and subject to technical and organizational measures to ensure non-attribution. *Encrypted data* then means personal data, which through technological protection measures are rendered unintelligible to any person who is not authorized to access it (European Parliament 2014). (8)

It is notable that according to these definitions both pseudonymous data and encrypted data are considered to be personal data. Therefore the Data Protection Law applies to them.

Anonymised data means data in which all identifiers have been removed so that there is no reasonable possibility to link data back to individual persons to whom data relates and no code key exists to link the data to persons. Anonymised data are not personal data as the data has been altered so that the data subjects can no longer be identified. The possibility to re-identify data subjects must be considered on a case-by-case basis. For example, the deletion of names and personal identity numbers is often not sufficient to make data anonymous. Complete anonymity requires that the possibility for both direct and indirect identification is removed and that the code key is destroyed.

Aggregated data means statistical data on individuals that has been combined to show values without the possibility to identify individuals within the aggregated data set. One practice has been to share and hand over *aggregated* or *anonymised data* in order to eschew the data protection norms. Often however, that is not possible as the analysis requires sharing of individual level data whether it is in an encrypted or a pseudonymous form.

The Data Controller of the Patient Registry should always be defined unequivocally. *Data Controller*, according to the Data Protection Directive, means the natural or legal person, public authority, agency or any other body which alone or jointly with others determines the purposes and means of the processing of personal data. Where the purposes and means of processing are determined by national or Community laws or regulations, the controller or the specific nomination criteria may be designated by national or Community law.

Data Processor, according to the Data Protection Directive, means a natural or legal person, public authority, agency or any other body which processes personal data on behalf of the controller. *Third party* means any natural or legal person, public authority, agency or any other body other than the data subject, the controller, the processor and the persons who, under the direct authority of the controller or the processor, are authorized to process the data. *The recipient* means a natural or legal person, public authority, agency or any other body to whom data are disclosed, whether a third party or not.

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.

5.4 Semantic aspects

Operating in an international environment or the readiness to do so requires that solutions regarding linguistic barriers have been thought of and implemented – both at the level of data and at the level of generic information necessary for data sharing (e.g. information on procedures for access to data, application forms etc.)

The comparability and transferability of health data across languages and contexts of use is heavily dependent on the adoption and use of accepted coding standards (see chapters [3.2.5](#) and [10.11.3](#)).

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 are 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 registry’s information model in a specified structure. That is, to make standardization as complete as possible, it must also cover data architecture and programming details. There are numerous standards in use, stemming from different traditions in e.g. statistics, informatics and commerce. The International Organization for Standardization (ISO) has published standards called ISO/IEC 11179 and also ISO15000-3 and ISO15000-4 for representing metadata for an organization in a metadata registry (9, 10). The Dublin Core metadata terms are a set of vocabulary terms which can be used to describe a wide variety of resources (11). Each of its 15 elements can be further defined with other vocabularies (e.g. SKOS, FOAF, ADMS, DCAT, QB). The Data Documentation Initiative (DDI) aims to connect the two aforementioned approaches (12). A specific set of metadata in greater detail can be described in a data model (e.g. RDF, Topic Maps) and further in its presentation format (e.g. HTML, XML). As conditions and requirements vary in patient registry environments, instead of recommending the use of one particular standard, PARENT recommends the use of widely known standards in semantics, data model and presentation formats, as well as ensuring the interoperability in close communication with other registries.

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, pharmaceutical 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. preferably on the basis of agreed standards. This kind of increased visibility and traceability of health data collections would benefit patient registries and lead to new ideas and innovations. Joining yellow-page type services like the PARENT Joint Action Registry of Registries (PARENT RoR), the AHRQ Registry of Patient Registries (RoPR) or other specialized “umbrella” registry is a concrete implementation step and opportunity for identifying further areas for targeting development efforts.

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).

5.5 Technical aspects – Guaranteeing shareable data

There are different levels of implementation for the technical solutions required, starting from the choices made at the level of an individual registry and up to the level of platforms for cross-border sharing of data. It is not the purpose of these guidelines to take a stand in advising for or against the use of specific technological solutions, since these are both context-specific and a constantly moving target as new technologies emerge. However, the technical layer is crucial in ensuring the ‘shareability’ of health data and hence adopted solutions must be such that take into account and support regional/national infrastructures and semantic requirements for patient data collected in the process of healthcare services provision.

On the level of technical operationalization of legal requirements, particularly in terms of data protection and safety, adopted solutions must be robust and reliably proven to perform the expected tasks.

5.6 Effective and sustainable solutions for cross-border registry collaboration

The creation of effective and sustainable solutions for cross-border use of registry data is a process where all the aforementioned requirements must be concertedly brought into play in order to serve clearly defined unique targets, such as those explored in the PARENT Joint Action Scenarios. The added value generated by achieving these targets will act as the key driver for the engagement of stakeholders who in turn can guarantee the sustainability of the required cross-border registry infrastructure and operation environment, a subject discussed in detail in the respective PARENT Joint Action report.

5.7 Health Data Resources in Europe: Mapping national strategies

The development of registries, as well as their preparedness for cross-border operations is largely dependent on the positioning of health data resources in national strategic prioritization regarding scientific data resources and research infrastructures.

One of the priorities of PARENT was to investigate, map and analyse the current policy and strategy landscape regarding the utilization of health data resources in EU Member States. The primary focus

was on strategies concerning existing national databases and registries (possibly established mainly for administrative purposes) but also patient data generated in the process of health care service provision and delivery (e.g. the database of healthcare provider organizations), as well as newly-generated research data.

The reason for embarking on this work was that the creation, maintenance and development of registries generally and particularly in the long run will depend considerably on the position the subject of health data resources has or has not acquired on national level strategic prioritization.

The challenges included the following:

- not all countries have an established strategy on the subject (e.g. Member States with a long registry tradition, such as the Nordic countries clearly do, but that cannot be assumed for everyone)
- the organizations/institutions responsible for the subject and hence authors of a possible strategy were quite diverse, hence a broad scope is needed in conducting the preliminary search
- a significant body of work has been accomplished under broader subjects such as eScience and generally research oriented initiatives
- the work needed to cover both the Member States, as well as the EU-level developments

The following table summarizes indicative countries with the main findings of the national strategies mapping.

Table 5.1: National strategies mapping⁴⁶

STRATEGY	LEAD ORGANISATION	MAIN ACTORS	KEY POINTS
Croatia			
<ul style="list-style-type: none"> • National Health Care Strategy 2012 - 2020 	<ul style="list-style-type: none"> • MoH 	<ul style="list-style-type: none"> • Croatian National Institute of Public Health • Agency for Medicinal Products and Medical Devices of Croatia (HALMED) • Public health IT system 	<ul style="list-style-type: none"> • Central Health Care Information System in Croatia (CEZIH) • Direct retrieval of the patient's data from the Electronic Health Record • Public health IT system manages more than 33 registries which are, however, neither linked nor standardized; • Drafting of the Strategy of integral management of health care archive and registry material in the Republic of Croatia for the period from 2012 to 2020 is in progress • Pharmacovigilance system development based on electronic adverse event reporting
Cyprus			
<ul style="list-style-type: none"> • "Digital Strategy for Cyprus" (2012) 	<ul style="list-style-type: none"> • MoH • Ministry of Communications and Works 	<ul style="list-style-type: none"> • Health Monitoring Unit 	<ul style="list-style-type: none"> • high quality and high speed electronic services, while ensuring the security of infrastructures and information • paperless government & eGovernment services • Integrated Health Care Information System in all hospitals; • Regional health networks, AAL, Telemedicine use
Denmark			
<ul style="list-style-type: none"> • Danish Roadmap for Research Infrastructures (2011) 	<ul style="list-style-type: none"> • Danish Agency for Science, 	<ul style="list-style-type: none"> • National Service Platform (NSP) 	<ul style="list-style-type: none"> • Political focus on creating framework conditions conducive to research & business development in healthcare & welfare

⁴⁶ Data by May 2014, non-validated by Member States representatives

<ul style="list-style-type: none"> National Action Plan for Strengthening Public-Private collaboration on clinical research (2014) 	<ul style="list-style-type: none"> Technology and Innovation National Experimental Therapy Partnership (NEXT) National eHealth authority 	<ul style="list-style-type: none"> The Clinical Trials office Denmark Statens Serum Institut Danish National Biobank Coordinating Centre 	
<ul style="list-style-type: none"> MAKING eHEALTH WORK - National Strategy for Digitalisation of the Danish Healthcare Sector (2013-2017) 			
Estonia			
<ul style="list-style-type: none"> National Health Plan (2009 - 2020) 	<ul style="list-style-type: none"> Ministry of Social Affairs 	<ul style="list-style-type: none"> Other ministries (incl. Ministry of Economic Affairs & Communications, Ministry of Education and Research local governments and civil organisations 	<ul style="list-style-type: none"> HIAS – Health Information and Analysis System: monitoring system to collect data on the health status of different population groups, any possible changes and past interventions plus data management, preservation and initial analysis Ensuring quality of the collected data and comparability of different data sources; Establishing links between different datasets Child health, population sexual and reproductive health and infant health indicators and impact factors through surveys, development of medical registries and health information systems and specification of the content of collected data.
Ireland			
<ul style="list-style-type: none"> Ireland eHealth Strategy (2013) 	<ul style="list-style-type: none"> Department of Health 	<ul style="list-style-type: none"> eHealth Ireland Council of Clinical Information Officers eHealth Ecosystem 	<ul style="list-style-type: none"> eHealth as National Infrastructure Investment Expected outcomes to be delivered in phase one – years 1- 4: High quality data sets are available to indicate public health trends and inform regional and national policy Improved ability to support surveillance and management of public health interventions Improved ability to analyse and report on population health outcomes Benefits to Ireland: Public Health is greatly enhanced by the availability of high quality and accurate data sets which can be readily accessed and mined. Trends in population data can be observed and if necessary interventions can be enacted. Valuable longitudinal studies can be performed from anonymised data sets which can be used to inform national policy. The Health Services Executive (HSE) has established the National ICT Integrated Services Framework (ISF). The ISF is an Interoperability Framework offering a shared standards based tool and language for defining and aligning the business and interoperability context for Ireland's eHealth systems.
Norway			
<ul style="list-style-type: none"> National strategy for health registries for the period 2010-2020 National Health Registry project (2011) Health Registries for Research (2014) Biobank Norway: A national infrastructure for Biobanks and Biobank related activity (2012) 	<ul style="list-style-type: none"> Norwegian government Ministry of Health and Care Services Norwegian Research Council 	<ul style="list-style-type: none"> National Institute of Public Health Universities BBMIR.no 	<ul style="list-style-type: none"> co-ordinate and modernise the national clinical registries and the mandatory national health registries Improve the use of data for research, health surveillance, prevention and quality assurance of healthcare.

United Kingdom			
<ul style="list-style-type: none"> Health and Social Care Information Centre (2013) 	<ul style="list-style-type: none"> Department of Health NHS England 	<ul style="list-style-type: none"> Health and Social Care Information Centre 	<ul style="list-style-type: none"> NHS has established Health and Social Care Information Centre with the aim of driving the use of information to improve decision making and deliver better care. Secondary Uses Service (SUS) is the single, comprehensive repository for healthcare data in England which enables a range of reporting and analyses to support the NHS in the delivery of healthcare services. Data Linkage and Extract Service provide extracts from a range of individual and linked data sets and can add significant value to individual sets of data by combining and matching them at individual record level in a secure environment.
<ul style="list-style-type: none"> Public Health Surveillance - Towards a Strategy for Public Health England (2012) 	<ul style="list-style-type: none"> Public Health England 	<ul style="list-style-type: none"> Public Health England Public Health England Transition Team 	<ul style="list-style-type: none"> Most commonly identified priority by stakeholders for strengthening the public health surveillance function was that greater use be made of existing data through linkage of those data and analyses of those linked data sets. Hold data securely in line with published information governance standards set out in the Department of Health's Information Governance Toolkit. Data and knowledge gateway (Beta site http://datagateway.phe.org.uk/?lk_sr=govphe) A single point of access to data and analysis tools from across Public Health England.
<ul style="list-style-type: none"> eHealth Strategy 2011 – 2017 	<ul style="list-style-type: none"> The Scottish Government 	<ul style="list-style-type: none"> NHS Scotland eHealth Directorate eHealth Strategy Board 	<ul style="list-style-type: none"> Publication of an Information Assurance Strategy and core guidance including Records Management.
<ul style="list-style-type: none"> The UK Administrative Data Research Network: Improving Access for Research and Policy (2012) 	<ul style="list-style-type: none"> Economic and Social Research Council / The Administrative Data Taskforce 	<ul style="list-style-type: none"> Economic and Social Research Council Medical Research Council and the Wellcome Trust 	<ul style="list-style-type: none"> An example of access and linkage problem experienced by researchers concerning health data. A researcher was requested by the Chief Medical Officer (CMO) for Wales to carry out research into the factors underlying excess winter mortality using the Secure Anonymised Information Linkage (SAIL) data bank. In appendix 2 various national and international models of access to national administrative datasets
<ul style="list-style-type: none"> Seizing the data opportunity. A strategy for UK data capability (2013) 	<ul style="list-style-type: none"> HM Government 	<ul style="list-style-type: none"> Information Economy Council E-infrastructure Leadership Council Open Data Institute 	<ul style="list-style-type: none"> A special reference to the “information economy”. The Information Economy Council is developing a long-term strategy to drive the growth of the Information Economy sector. The E-infrastructure Leadership Council to advise the government on the computing infrastructure and skills Over 10,000 public datasets (including health data) published on data.gov.uk, and the ground-breaking Open Data Institute. The Open Data Institute is catalysing the evolution of open data culture to create economic, environmental, and social value.
Sweden			
<ul style="list-style-type: none"> National eHealth – the strategy for accessible and secure information in health and social care (2010) 	<ul style="list-style-type: none"> Ministry of Health and Social Affairs 	<ul style="list-style-type: none"> National High-Level Group for eHealth Ministry of Health and Social Affairs 	<ul style="list-style-type: none"> National High-Level Group for eHealth to successfully coordinate the implementation of the strategy, a National High-Level Group for eHealth has been in place since 2005.

		<ul style="list-style-type: none"> • Swedish Association of Local Authorities and Regions • Association of Private Care Providers • National Board of Health and Welfare • Famna, Swedish Association for Non Profit Health and Social Services • National advisory group 	<ul style="list-style-type: none"> • A new consultative body in the form of a national advisory group was created during 2009 with the aim of preparing and formalising the dialogue with most of the key actors in the health and social care sector. • National Information Structure (NI) describes what kind of information is needed in health and social care documentation on the general level. It also describes how the information should be structured so that it can be used in different contexts, for different purposes, in the health and social care process and for monitoring and managing activities. The National Board of Health and Welfare has developed an initial NI version. • Healthcare service information in civil registries (VIF). The aim of this project is to improve the quality and service of the information exchange between the healthcare services and civil registries at the Swedish Tax Agency and to the National Board of Health and Welfare.
<ul style="list-style-type: none"> • Public Performance Reports on Health Care and Social Services (2010) 	<ul style="list-style-type: none"> • Ministry of Health and Social Affairs 		<ul style="list-style-type: none"> • A national strategy for quality improvement through public performance reports on health care and social services. Public performance reports are a set of indicators that are regularly reported and that describe services with regard to patient satisfaction, waits and delays, quality, prevention, safety, efficiency, etc.
Spain			
<ul style="list-style-type: none"> • Summary of the Strategy in Mental Health of the National Health System (2008) 	<ul style="list-style-type: none"> • Ministry of Health and Consumer Affairs 	<ul style="list-style-type: none"> • Health Institute Carlos III (research) • Autonomic Communities 	<ul style="list-style-type: none"> • Need of using tools which could allow the evaluation of the reform objectives, trying to overcome the information systems of purely administrative character. There is a reference to the current situation on Mental Health and to recommendations. • Minimum Basic Set of Data (CMBD) at hospital discharge.

References

1. Guidelines on minimum/nonexhaustive Patient Summary Dataset for electronic exchange in accordance with the Cross-Border Directive 2011/24/EU, Adopted by the eHealth Network, 19 November 2013. Available at:
http://ec.europa.eu/health/ehealth/docs/guidelines_patient_summary_en.pdf
2. Guidelines on ePrescriptions dataset for electronic exchange under Cross-Border Directive 2011/24/EU. Adopted by the eHealth Network, 18 November 2014. Available at:
http://ec.europa.eu/health/ehealth/docs/eprescription_guidelines_en.pdf
3. European Charter of Fundamental Rights. Homepage at:
http://ec.europa.eu/justice/fundamental-rights/charter/index_en.htm
4. Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data . Official Journal L 281 , 23/11/1995 P. 0031 – 0050
<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML>
5. WP 29, Opinion 15/2011 on the definition of consent. Adopted on 13 July 2011. Available at:
http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2011/wp187_en.pdf
6. DG Justice. Reform of the data protection legal framework in the EU.
http://ec.europa.eu/justice/data-protection/reform/index_en.htm
7. The Nordic Trial Alliance Working Group 6 on Transparency and Registration. Report on transparency and registration in clinical research in the Nordic countries. March 2015. Available at:
<http://nta.nordforsk.org/news/final-report-on-transparency-and-registration-in-clinical-research-in-the-nordic-countries-published>
8. European Parliament legislative resolution of 12 March 2014 on the proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation) (COM(2012)0011 – C7-0025/2012 – 2012/0011(COD)). Available at:
<http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P7-TA-2014-0212+0+DOC+XML+V0//EN>
9. ISO Metadata Standards: <http://metadata-stds.org/>
10. ISO/IEC 11179: http://www.iso.org/iso/catalogue_detail.htm?csnumber=50340
11. Dublin Core Metadata Initiative: <http://dublincore.org/>
12. Data Documentation Initiative (DDI): <http://www.ddialliance.org/>

6 CREATING A REGISTRY

6.1 Planning a registry

Dmitri Wall, prof. Alan D. Irvine, Eoin O'Brien

Considerable resources and support are required to develop a successful, sustainable patient registry, the extent of which may not be obvious at the outset of planning. This section outlines an approach to registry planning that is believed to be logical and capable of enabling registry planners to propose feasible development strategies that will avoid duplication of effort and incorporate best evidence and expertise, while involving the extensive stakeholder support required to create an abiding, exemplary registry that can become a valuable, interoperable cog in the broader health ecosystem.

Key principles:

- Maintaining a focused, open and transparent approach to registry development is vital.
- The purpose, objectives and outputs of the registry should be clearly and succinctly defined, in a manner supported by best evidence and guidance, and that is aware of whether there is overlap with other projects nationally and internationally.
- Key determinants of high-quality data should be considered and essential means of improving data quality embraced.
- The broad health information landscape in which the registry will operate should be examined to ensure that the registry is interoperable and adheres to relevant standards, datasets and terminology.
- Clear guidance with respect to legal obligations should be obtained. This should include consideration of privacy impact assessments, data protection policies, data ownership, data access and intellectual property.
- Obtaining guidance from key experts and stakeholders is crucial and should be facilitated by engaging an advisory board.
- Broad stakeholder engagement can generate involved registry champions and committee members who can facilitate the smooth running and enduring success of a registry.
- The scope of the registry should be maintained as the project develops.
- A governance plan should be established. This should be supported by registry teams that include, at a minimum, a project management team, a scientific committee and a quality assurance committee. Amongst other things, these groups can identify datasets required to ensure the registry meets its predefined outputs, in addition to creating clear data access policies and ensuring that quality assurance is maintained.
- Inclusion of opinions of patients and real-world users of the proposed registry is essential.
- The feasibility of the registry should be iteratively considered as the project progresses and the funding strategy should aim to ensure lasting success of the registry. This should include proactive risk identification.
- Should the proposed project prove to be feasible, an implementation plan that includes a proof-of-concept pilot project to test the registry in a real-world environment is recommended.

This guide 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. Though 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 and efficiency, and has well-defined, valuable outputs that can be measured, so that the quality and success of the registry can be verified.

Though this section evolves in what we believe is 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. We do, however, feel that addressing each section will add value to the registry and increase the likelihood of developing a successful registry.

During each phase of planning, we advise 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 we envisage the role of registries becoming progressively more valuable. This will only happen if they are developed in a manner that is cognisant of the importance of interoperability.

We also suggest that 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. We endorse in particular the creation of a resource such as PARENT's [Registry of Registries](#) or the AHRQ's [Registry of Patient Registries \(RoPR\)](#) which are helping to connect registries, while raising the standard of registries considerably (1, 2). We strongly advocate ensuring that any registry created forthwith joins such initiatives.

It should be noted that two resources in particular were of considerable help in informing the authors and in structuring this chapter. We highly recommended their utilisation as reference documents of extraordinarily high standard. These are "Registries for Evaluating Patient Outcomes: A User's Guide" and the "ISPOR Taxonomy of Patient Registries: Classification, Characteristics and Terms" (3, 4).

6.1.1 Defining the Purpose, Objectives and Outputs of the Registry

6.1.1.1 Purpose(s)

The first step is to clearly define the overarching aim(s) or purpose(s) for which the registry is being established. This may emerge from a clinical need, a post-marketing requirement, or an interest of patients or clinicians, but the purpose(s) should be capable of being realised through the prospective, non-interventional, scientific approach that a registry should adhere to.

So far as possible, the purpose(s) should be limited in scope and number to ensure focus. As will become apparent, the expansion of ideas is likely to occur rapidly once stakeholders become involved and it is important to limit this at an early stage so as to prevent it becoming too unwieldy to manage. As with any scientific endeavour, this process will be greatly facilitated by conducting a literature review to analyse what information already exists within the scope of the proposed registry. This might demonstrate that while the purpose and objectives are reasonable, a clinical trial or other study design might be a more appropriate means of delivering the required outputs. Furthermore, a literature

review will highlight relevant experts and stakeholders in the field of interest who might be contacted as part of a stakeholder evaluation or for expert advice.

6.1.1.2 Objectives

To facilitate the generation of a valid scientific question, the registry's purpose(s) should be divided into specific objectives, which together will achieve that overarching purpose(s) of the registry. It is worth considering how each objective might translate into a dataset and to imagine whether a scientific methodology could be applied to help validate whether the objectives are achievable.

6.1.1.3 Outputs

Ultimately, a registry's findings are only valuable if the data they generate can be translated into information capable of improving health outcomes. This is more likely to occur if outputs are considered at an early stage, so as to achieve the following objectives:

- Ensure that a registry is purposeful as there will be measurable end-points against which its success can be judged
- Identify potential experts required to advise the development of the registry
- Identify potential stakeholders
- Facilitate buy-in through the identification of outcomes of interest
- Identify the target audience for whom the information gained from a registry might be valuable. This will facilitate the most effective dissemination of results and will also help identification of unforeseen requirements. For example, the primary purpose of a drug registry might be to identify its effectiveness in a real world setting; however, mandatory reporting of adverse effects will also need to be considered.

6.1.1.4 Process outcome

Defining the purpose, objectives and outputs of a registry will typically clarify a registry as belonging to one of three groups (or a combination of these). Conversely, by considering a registry as belonging to one of these groups, may facilitate defining a registry's purpose, objectives and outputs.

- **Condition based registry**
 - **Purpose:** Though there are many listing registers, which identify patients suffering from a particular condition, 'condition-based registries' in this document refer to registries that aim to describe outcomes related to a particular condition.
 - **Purpose example:** description of the natural history of chronic obstructive pulmonary disease (COPD).
 - **Objective example:** identification of depression in patients with severe COPD.
 - **Output example:** Defining the prevalence of depression in COPD and examining how this might be more effectively be detected in the COPD care pathway.
 - **Registry example:** The Malta National Cancer Registry.
- **Product based registry**
 - These registries typically focus on medical devices or pharmaceutical products.
 - **Example:** Arthroplasty Registries

- There are multiple registries in the vast majority of EU Member States in different stages of development, which monitor approximately 4 million patients worldwide at present. Shortcuts to websites of relevant arthroplasty registries for further information are available at www.eur.euroform.org/registers.aspx
 - **Purposes (Example):**
 - The assessment of:
 - Real-world effectiveness
 - Safety and cost effectiveness of a new device
 - Outcome monitoring of performance and potential safety issues over the entire life cycle
 - Early signal detection of inferior outcome of device and surgical techniques
 - The impact of patient profile/comorbidities/risk classes on patient side on the outcome
 - Market monitoring concerning implants and health care providers
 - Feedback to health care providers
 - Identification of fields for improvement and monitoring of effects.
 - **Objective (Example):**
 - Defining the number of post-operative complications related to the device insertion to facilitate feedback to stakeholders in order to support decision-making.
 - **Outputs (Example):**
 - Demonstrating that a device or surgical technique is associated with increased postoperative complications
 - Fulfilment of post-marketing obligations
 - Validation of realisation of expected value by innovations and/or premium products
 - Transparent ranking of quality achieved by implants and health care provider
- **Services based registry**
 - These registries aim to evaluate the quality of service provision or correlate an intervention with outcomes.
 - **Registry example:** The Slovenian Hospital Discharge Registry
 - **Purposes (Example):**
 - The assessment of:
 - All hospital discharges (one day or longer) due to illnesses, injuries, poisonings, childbirths, stillbirths, sterilizations and new-borns in all Slovenian hospitals
 - Information for monitoring, planning, management and development of health care system
 - Health status of the population
 - Cost effectiveness
 - Patient safety quality
 - Other quality indicators
 - **Objectives (Example):**

- To assess the health status of the population and specific subgroups
- Setting the priorities for developing national policies for improvements of health care system
- To assess potential inequalities in health
- **Outputs (Example):**
 - Prevalence in the population for a certain disease or condition in a specific time period (e.g. year)
 - Determination of quality level for certain quality indicators:
 - patient safety indicators, e.g. postoperative complications, obstetric traumas,
 - quality indicators related to acute care, e.g. 30 day in-hospital and/or out of hospital mortality
 - Calculation of burden of specific diseases
- **Combination**
 - As is obvious from the examples above, some registries may have aspects that belong to more than one registry type.

6.1.2 Data Considerations

The success of a registry will ultimately be judged on its ability to meet the goals it was created for. This requires the collection and analysis of sufficiently high quality, targeted data specified by research hypotheses and the dissemination of the results of these analyses. High-quality data are also a key component in enabling interoperability (discussed later in this chapter). Though this section might be expected to occur later in the sequence of registry planning, success-by-design warrants the consideration of the determinants of high-quality data at an early stage of planning a registry. This should result in a focus on instilling key building blocks of quality, making the process of verifying the quality of the registry much easier when audit and other quality assurance processes are conducted post-implementation.

Knowledge of key determinants of data quality and how to achieve it will raise awareness of possible obstacles that might threaten the creation of a registry, such as the absence of an electronic health record to provide useful data. This knowledge also serves to underline the importance of a considered stakeholder evaluation that avails of input from groups that the registry planners may not have considered.

6.1.2.1 Data Quality

Firstly it is worth recalling that data quality is influenced by a number of factors, categorized into four groups. These factors, including data quality, are considered in much greater depth in chapter 4 'Quality dimensions of registries', which the reader is strongly advised to review before proceeding. In this chapter, these components are integrated within a suggested sequence of steps in planning a registry rather than factor by factor.

6.1.2.1.1 Data and Information types

Data or information may be considered primary or secondary. Primary data, or information, refers to data collected to "provide health or social care to the data subject" (5). "Secondary use of information

relates to information collected in the course of providing care, being used for purposes other than direct service-user care” (5). The use of data for secondary purposes, such as research, is governed by legislation, which varies across Member States. As such it is advised that legal expertise is sought. It might be prudent to adopt a position that secondary use of data requires explicit consent from patients or full anonymisation, which should be performed in keeping with local data protection regulation.

6.1.2.1.2 Data Quality Dimensions

“The delivery of safe and effective healthcare depends on access to, and use of information that is accurate, valid, reliable, timely, relevant, legible and complete” (5). Data quality dimensions are presented in [Table 4.1](#) of chapter 4, so that registry planners might consider what dimensions are significant in the context of the purpose and objectives of the registry that they are planning. ‘Seven essential means of improving data quality’ are also described there ([Table 4.2](#)) so that registry planners might consider at a high-level, how these will be addressed by their registry (5, 6).

6.1.2.1.3 Method of data capture

The quality of data will be significantly affected by the manner in which data are collected. Data collection can be considered with respect to two major domains; data source and data provider:

- Data sources
 - Paper-based
 - Questionnaire
 - Paper health record review
 - Documentation review
 - Laboratory reports
 - Other
 - Electronic
 - Questionnaire
 - Electronic Health Record
 - Laboratory reports
 - Databases
 - Mobile applications
 - Health devices
 - Social media
 - Other
- Data provider
 - Clinical units
 - Laboratories/central services
 - Discharge registries
 - Patients and families
 - Patients user groups (associations/federations)
 - Disability registries
 - Centres of expertise
 - Birth registries
 - Cause of death registries
 - Insurance funds (public and private)
 - Other registries
 - Other

Paper-based methods

Paper-based records have the advantage of being relatively inexpensive to create and distribute. However, in an era where health is becoming progressively more connected, paper is potentially very restrictive and does not take advantage of many error avoidance techniques that electronic methods offer. It is also worth considering that at some point, the data will need to be collated electronically to facilitate analysis. Paper can still play a core role in registry design however. Questionnaires and process flows can be created using paper which can be far more accessible for primary stakeholders. Once a prototype has been created using this method, it can facilitate the development of an electronic solution.

Electronic-based methods

Though the design of bespoke electronic solutions can be expensive, their advantage is that of connectivity, error minimization and reduction of duplication. While Electronic Health Records are still in evolution they are certainly not ubiquitous and they still have significant difficulties associated with their use. When they are available, and adhere to appropriate interoperability and terminology standards, they can offer an exceptional source of data for a registry. The list of potential electronic sources is large and it is for this reason that it is highly recommended that registry designers make considerable efforts to liaise with national, and possibly international, health, information and registry bodies to maximize resource utilisation.

Future developments

The recent explosion in mobile Health (mHealth) warrants consideration. As noted previously, there is an ever-increasing facility to utilize technology to connect data that has previously been unimaginable. Similarly, social media has established an almost ubiquitous presence and the extensive data networks that have resulted are of enormous potential to registries. It may be advisable to contact universities and connected health centres to consider what initiatives and ecosystems a registry could form part of to maximize the potential of mHealth and social media.

6.1.3 Overview of the Current State and the Importance of Interoperability

6.1.3.1 Overview of Current State

Having an appreciation for the organizational structure of registries and other healthcare information networks or ecosystems nationally and internationally is of vital importance in ensuring that a registry is best placed to make use of existing resources. This knowledge will also help orientate how a registry's role can best be positioned to fit into "the bigger picture" and contribute to the direction of health policy. There may be relevant data sources that could be integrated within your registry or vice versa. It may be that your proposal has previously been assessed, but was determined not to be feasible. Furthermore, significant resources might be spared through the identification of existing solutions the proposed registry might otherwise have replicated.

6.1.3.2 Interoperability

Interoperability is the means of ensuring that a registry will be able to integrate within "the bigger picture". Interoperability is defined by the Institute of Electrical and Electronics Engineers (IEEE) as the "ability of a system or a product to work with other systems or products without special effort on the

part of the customer” (7). Interoperability is a core component of good communication and as a result, effectiveness and safety.

Meta-analysis has demonstrated the importance of good communication within healthcare scenarios, suggesting that “interventions to improve the quality of information exchange increases effectiveness” (8). In addition, the value of improving information transfer has been noted by major organizations, such as the Agency for Healthcare Research and Quality (AHRQ) in the United States, as an “important patient safety practice” (9). Another US organization, the Institute of Medicine, having identified the extent of the risk posed by medical error in the landmark paper “To Err is Human”, have suggested the development of improved communication systems as core components of modern healthcare systems (10, 11).

6.1.3.3 Planning for integration

As “interoperability is made possible by the implementation of standards”, liaising with national regulation/quality improvement authorities, which can be a valuable source of advice regarding access to and appropriate use of relevant standards, is essential (7).

Of particular relevance from a registry development perspective, is the selection of standard datasets and terminology to facilitate local and cross-border interoperability. For general areas, such as demographics, PARENT is an excellent source of guidance with respect to standard datasets and terminology or to facilitate contact with a registry in another state with a structure and composition that can be adapted or adopted for a new registry’s needs. At a national level, regulatory bodies will typically be able to advise best use of classification systems such as the World Health Organisation’s (WHO) International Classification of Diseases (ICD) or terminologies such as the International Health Terminology Standards Development Organisation’s (IHTSDO) Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT®). For more specific areas, national or international professional clinical groups can be a rich source of information. It should only be a last resort that non-standardised terminology/datasets are developed by a registry team and this should only be considered after discussion with appropriate experts/standards bodies to advise on how the dataset should be best developed.

An ultimate end point of achieving interoperability is to prevent potentially valuable data becoming trapped in information “silos” and facilitate more accurate representation of concepts and comparison of data across international borders.

The same connections that will facilitate interoperability are likely to be able to provide information regarding the current state of the art in registry design. In addition, we suggest contacting health authorities that are likely to have useful guidance regarding supportive ecosystems, including other registry groups. They may also be able to provide a clear picture of current and planned national health and information strategies and existing projects that could provide data sources for the registry, such as an electronic health record, or a national data architecture. PARENT will be able to offer further registry establishment advice and tools for registry development.

“Connected health” is a developing concept which “encompasses terms such as wireless, digital, electronic, mobile, and tele-health and refers to a conceptual model for health management where devices, services or interventions are designed around the patient’s needs, and health related data are shared in such a way that the patient can receive care in the most proactive and efficient manner

possible. All stakeholders in the process are ‘connected’ by means of timely sharing and presentation of accurate and pertinent information regarding patient status through smarter use of data, devices, communication platforms and people”(12). As connected health democratizes health information, there is significant potential for a registry to increase:

- Awareness and interest
- Dissemination and impact of outputs
- Collaboration opportunities
- The volume and variety of data sources available
- Resource sharing

It is therefore worth liaising with centres promoting connected health, such as universities, or non-profit-making groups such as the [ECHAlliance](#) to establish how a registry might integrate in the process (13). Conversely, the considerable organization required to develop a registry may facilitate the development of an ecosystem that can drive connected health.

6.1.4 Considering Legal Aspects and Confidentiality

While there are many important aspects to planning a registry, ensuring compliance with data protection regulations is not only vital, but a legal requirement, the breach of which may result in termination of the registry project. Furthermore, adopting a gold standard, transparent data protection practice is likely to increase the confidence that registry participants will place in a registry and add to its value. As such it is essential to prioritise contacting the relevant national Data Protection Commissioner’s Office early in the design of the registry. Contact details for EU member Data Protection Commissioners are outlined in [Table 6.1](#). More information about the legal aspect is available in chapter [5](#).

Table 6.1: Data protection authorities and contact details for EU Member States

EU Member State	Data Protection Authority	email
Austria	Österreichische Datenschutzbehörde	dsb@dsb.gv.at
Belgium	Commission de la protection de la vie privée	commission@privacycommission.be
Bulgaria	Commission for Personal Data Protection	kzld@cpdp.bg
Croatia	Croatian Personal Data Protection Agency	azop@azop.hr ; info@azop.hr
Cyprus	Commissioner for Personal Data Protection	commissioner@dataprotection.gov.cy
Czech Republic	The Office for Personal Data Protection	posta@uoou.cz
Denmark	Datatilsynet	dt@datatilsynet.dk
Estonia	Estonian Data Protection Inspectorate	viljar.peep@aki.ee
Finland	Office of the Data Protection	tietosuoja@om.fi
France	Commission Nationale de l'Informatique et des Libertés	
Germany	Der Bundesbeauftragte für den Datenschutz und die Informationsfreiheit	poststelle@bfdi.bund.de
Greece	Hellenic Data Protection Authority	contact@dpa.gr
Hungary	Data Protection Commissioner of Hungary	peterfalvi.attila@naih.hu
Ireland	Data Protection Commissioner	info@dataprotection.ie
Italy	Garante per la protezione dei dati personali	garante@garanteprivacy.it
Latvia	Data State Inspectorate	info@dvi.gov.lv
Lithuania	State Data Protection	ada@ada.lt
Luxembourg	Commission nationale pour la protection des données	info@cnpd.lu
Malta	Office of the Data Protection Commissioner	commissioner.dataprotection@gov.mt
Netherlands	Dutch Data Protection Authority	info@cbpweb.nl
Poland	The Bureau of the Inspector General for the Protection of Personal Data	sekretariat@giodo.gov.pl
Portugal	Comissão Nacional de Protecção de Dados	geral@cnpd.pt
Romania	The National Supervisory Authority for Personal Data Processing	anspdcp@dataprotection.ro

Slovakia	Office for Personal Data Protection of the Slovak Republic	statny.dozor@pdp.gov.sk
Slovenia	Information Commissioner	gp.ip@ip-rs.si
Spain	Agencia de Protección de Datos	internacional@agpd.es
Sweden	Datainspektionen	datainspektionen@datainspektionen.se
United Kingdom	Information Commissioner's Office	casework@ico.org.uk
England	Information Commissioner's Office	casework@ico.org.uk
Northern Ireland	Information Commissioner's Office	ni@ico.org.uk
Scotland	Information Commissioner's Office	Scotland@ico.org.uk
Wales	Information Commissioner's Office	wales@ico.org.uk
EU	Data Protection Officer of the EU	DATA-PROTECTION-OFFICER@ec.europa.eu

6.1.4.1 Privacy and Privacy Impact Assessments

“Privacy is the right of individuals to keep information about themselves from being disclosed” (14, 15). A privacy impact assessment (PIA) is a process that “facilitates the protection and enhancement of the privacy of individuals” and is best conducted at a planning stage to protect the registry and its participants from potentially irreconcilable personal and organisational breaches that may be damaging at a later stage (15). This will facilitate the identification of risks to privacy breaches and examination of how these risks can be allayed. Detailing the process involved in a PIA is beyond the scope of this chapter, however the Health Information and Quality Authority of Ireland provide an excellent range of [resources](#) in this area, including a review of international PIAs, a tool to establish whether a PIA is required, details regarding how it should be conducted and a sample report (15-17).

6.1.4.2 Data Protection Policy

Even following a PIA, it is advisable to develop a data protection policy for the registry project and ensure that all involved with design and implementation of the registry are appropriately trained in this regard and regularly made aware of their responsibilities. A local Data Protection Commissioner’s Office or health authority may provide links with groups who have a policy that can be adapted for the purpose of the registry.

6.1.4.3 Data Ownership, Access and Intellectual Property

While considering data security it is prudent to consider data ownership, access and intellectual property. This is likely to require dedicated expert guidance and, to ensure transparency; it is advised that the outcome of this process is formalized in a policy document. This document should also consider the scenario in which the registry project is terminated so that it is clear how the data might best be protected.

6.1.5 Eliciting Expert Opinion & Generating an Advisory Board

Expert elicitation refers to the “solicited exchange of knowledge, information, or opinion from an expert” (18). If the initial planning processes suggest that there is a valid opportunity to establish a registry, further planning can be greatly facilitated by expert guidance. We suggest the establishment of an Advisory Board consisting of a knowledgeable panel with expertise relevant to the registry domain and those who are committed to the establishment of the registry. This will not only facilitate the implementation of best practice, it will also help identify stakeholders who might not be immediately apparent to the group establishing the registry. Finally, the selection of appropriate

representatives for an advisory board is likely to increase the engagement of potential stakeholders with the project by virtue of their involvement, which can be vital to the success of the project.

While, from a practical perspective, it is most likely that experts, with a view to establishing an Advisory Board, are likely to be selected from the country in which the registry is to be implemented, it is advised that input, where possible, should be sought from international experts who have established registries in relevant domains. These experts can contribute vital, hard-won experience, resources and support that can improve the efficiency, sustainability and cost-effectiveness of developing a registry, as well as identifying potential unseen obstacles, thereby significantly increasing the chance of successful implementation of a registry and the relevance of its outputs.

6.1.6 Defining the Scope of the Registry & Building a Registry Development Team

It is advisable, at this point, to consider with the advisory board and funders what the scope of the registry will be. Though this may seem obvious once the purpose(s), objectives and outputs have been defined, these may be challenged by the open nature of stakeholder involvement and there is a significant risk of losing focus if clear limitations are not imposed. In addition, though an open stakeholder engagement process is likely to engage stakeholders' imaginations and promote innovative ideas and engagement with the project, false promises can lead to significant disappointment at a later stage in the project.

The scope should aim to highlight the value of achieving the purpose(s), objectives and outputs of the registry with the minimal complexity possible, and in a manner that is most likely to be successfully accepted by users. Financial resources should be defined and a rough timeframe be agreed to give invited stakeholders an opportunity to plan when they can engage.

As there will be considerable time and preparation involved in developing the registry from this point, it is advised that a project development team is established that is proportional in size to the level of resources available to the registry. This might be a registry champion or person with an interest in the area in question, but thought might also be given to involving a research fellow with an interest in registries, healthcare informatics or the area targeted by the registry. Ideally, this person would be a primary stakeholder with a long-term interest in the area the registry is focused on. This will facilitate development of skills that can improve the long-term success of the registry, while also ensuring that the registry is designed in a fashion cognizant of end-users requirements.

6.1.7 Performing Stakeholder Engagement and Analysis

The Health Information and Quality Authority of Ireland (HIQA) have produced a document entitled "Guidelines for Stakeholder Engagement in Health Technology Assessment" which provides a comprehensive overview of stakeholder assessment that is extremely relevant to registry planning (18).

HIQA note that "Stakeholder engagement is an iterative process of actively soliciting the knowledge, experience, judgment and values of individuals selected to represent a broad range of direct interests in a particular issue" (18, 19). Though stakeholder analysis might also involve the experts identified in [4.1.3](#), the aim in stakeholder analysis is to avoid solicited advice and instead facilitate wider engagement on the topic (18).

The process of stakeholder engagement should also be seen as an inclusive “hearts and minds” campaign. An effort to be inclusive and respectful of all stakeholders’ contributions can significantly improve the registry’s later adoption and success.

6.1.7.1 Identification of Stakeholders

Though the definition of what constitutes a stakeholder varies, for the purpose of a registry, two subtypes can be considered (3):

- Primary stakeholders are intrinsically involved in the design and funding of the registry, but may also include parties with a regulatory capacity.
- Secondary stakeholders may be affected by and involved in using and operating the registry, but do not have direct involvement in its design.

6.1.7.2 Engagement

As the stakeholders of a registry may be extremely diverse, it is recommended that a flexible approach is adopted towards engagement. None-the-less, to facilitate transparency, consistency and relevance it is advised that a standard information document is prepared and distributed in advance, where feasible. Ideally, this document would support the conduct of a semi-structured interview.

6.1.7.3 Recording Stakeholder (& expert participation)

Even with a focused registry, the number of potential stakeholders and registry contacts can increase significantly beyond the expected scope. As such we would advocate using a tool to monitor involvement at a high level. The table in [Appendix B](#) is a re-usable means of collecting information about possible registry stakeholders and recording high-level outcomes from meeting with them, relevant for the purpose of designing the registry.

High-level categories of contacts include:

- Clinical groups
- Public health and regulatory groups
- Product and device manufacturers
- Health care service providers
- Health funding and insurance groups
- Patient and advocacy groups
- Academia
- Relevant experts
- Professional groups and societies
- Registry groups
- Registry sponsor groups
- Development groups (informatics and management)
- Other international groups

6.1.7.4 Content of the Stakeholder Evaluation

Though the content of the evaluation will vary greatly depending on the nature of the registry and its stakeholders, we suggest a process that has been adapted from Registries for Evaluation Patient Outcomes: A User's Guide" (3). We recommend providing the stakeholder with a document, striking a balance between delivering information and being concise, that consists of the following components:

1. Introduction to the group designing the registry, the current state and the motivation for developing a registry
2. A brief introduction to Registries
3. The purpose of the document
4. Engagement requesting input from stakeholders regarding
 - a. Purpose & Objectives
 - b. Available relevant information sources
 - c. Key stakeholders
 - d. Feasibility – barriers and motivators to establishing a registry
 - e. Registry team membership
5. Description of the further steps required to establish the registry

Though direct feedback is likely to be limited given the time constraints of busy stakeholders, this process is likely to not only be an exercise in clarifying the registry design process, but can create a template for a semi-structured interview with stakeholders at a later stage and also helps develop awareness of the project and confidence in the design process.

6.1.7.5 Stakeholder Evaluation Output

Given the diverse nature of stakeholders it is difficult to ensure consistency and as such a scientific document is unlikely to be produced. None-the-less, by following the process described above, the opportunity has been presented for frank and honest engagement and useful information and requirements will be made apparent.

The analysis of the stakeholder evaluation also provides a useful opportunity to personalize further interactions with stakeholders and to provide relevant information at conferences or with other stakeholders to increase awareness and further engagement.

6.1.8 Re-defining the Scope of the Registry

Following stakeholder assessment it is advisable to reconsider the scope of the project. While factors likely to improve stakeholder engagement and ultimately increase the chance of the registry's success are important, these should be weighed against the considerable expense the extra scope is likely to add. It is also worth noting that increasing the volume of data collection is typically associated with a decrease in completeness of data entry.

A final scoping document will facilitate the creation of a business case and will better inform selection of data elements of the registry and the registry data model.

From this point, changes to the scope may result in significant resource utilisation and, as such, a change management strategy should be created which outlines how further adaptations to the scope should occur in the future.

6.1.9 Governance, Oversight and Registry Teams

Before considering data elements for the registry and beginning to focus on the practical implementation of the registry, it is advisable to establish a governance plan and to develop teams that will facilitate design of the registry and maintenance following implementation.

This serves a number of purposes:

1. Creating teams can involve end-users, increasing buy-in.
2. Facilitating a better understanding of how the registry will operate and how intellectual property will be handled.
3. Creating the governance framework for data sharing and dissemination of data or information created by the registry.
4. Ensuring oversight and that the registry development is progressing as planned.

Particularly when the scope of the registry is small, there may need to be overlap, however, at a minimum, we suggest prioritisation of a project management team, scientific committee and a quality assurance committee. It is suggested that, though specific teams/committees will benefit from members with specific skill sets, that members be selected to ensure that all stakeholder groups are adequately represented. In particular, it is advised that patient groups should be asked to contribute to ensure that the patient's voice is represented appropriately as their data are the subject of the registry process.

6.1.9.1 Project management team

The involvement of a person skilled and experienced in project management is advised. If this is not possible, it would be worthwhile considering training for a project manager and consideration given to the use of project management software. [Table 6.2](#) outlines a tool to facilitate registry team organisation and selection.

6.1.9.2 Scientific Committee

The aim of the scientific committee should be to ensure that the registry is outcomes- driven and that the data collected are disseminated effectively. It is suggested that the committee aim to meet four main objectives:

- Question identification
- Data element identification and selection
- Dissemination of results
- External data access/study proposal adjudication

As such, this group should consist of subject matter experts, ideally with a track record in publication of scientific results. It would also be ideal to include members of the group with statistical/epidemiological and health outcomes analyses experience, so that these factors remain in

focus throughout the design, implementation and life of the registry. It is suggested that a transparent approach is undertaken with respect to member and topic selection.

6.1.9.2.1 Question identification

Based on the scope identified by the advisory board and the input of the stakeholder evaluation, the committee should identify specific questions that the registry will address. These questions will inform the selection of data fields that the registry will record.

6.1.9.2.2 Data element identification & selection

It is suggested that this process be considered an iterative one that considers the dimensions of data quality discussed previously.

Rough selection:

In the first instance, it is advised that the scientific committee consider a rough map of possible data fields. This should then be submitted for statistical analysis based on the scientific questions that have been proposed.

Statistical and Epidemiological analysis:

This process is vital to ensure that the registry is developed to an appropriate scale that ensures that the purpose and objectives it was created for are met.

Extra data fields add considerable complexity and cost because of data validation requirements. A statistical analysis can help highlight the essential fields for registry success and help maintain as much simplicity as possible; reducing the resources required ensuring completeness of data entry when the registry is implemented. It will also reduce the effort required to validate and analyse data.

It is advisable that this process is conducted by statisticians and epidemiologists trained in registry science. If the registry development group has no formal attachment with experts with skills in this area, it is worth checking with universities or other registry groups, who might identify relevant experts.

Health Outcomes/Pharmacoeconomics analysis:

At the same time as a statistical analysis review of potential registry outcomes from a health outcomes and pharmacoeconomics perspective should be considered. Increasingly, the relevance of real-world effectiveness is being prioritised and the relevance and attractiveness of a registry can be greatly increased by engraining it within national and international strategies. In Sweden, for example, the establishment of a hip and arthroplasty registry resulted in the avoidance of 7,500 revisions between 2000 and 2009, with a saving of \$140 million in costs.⁴⁷ This is also a mechanism of scientifically establishing the potential economic worth of the registry and as a means of creating a benchmark against which the registry might later be

⁴⁷ National disease registries for advancing health care. Lancet. 2011; 378(9809):2050.

evaluated as a marker of success. This can be of particular consequence when funding organisations are approached with a view to ensuring the long-term feasibility of the registry project.

Final Selection of elements:

The final selection of data elements is only likely to occur at the time of implementation of the registry, or ideally, after a pilot project has been conducted and after a financial analysis has identified the scope that can realistically be supported. The aim of the data selection process at the planning stage should therefore be to outline the data fields that will be required to a level adequate to conduct a feasibility study.

6.1.9.2.3 Dissemination of Results

Dissemination of registry data increases the potential impact of a registry and facilitates peer review. This process enables registry methods and data to be independently scrutinized, which in turn can validate the quality of the registry. Planning how registry data will be disseminated can help develop a timeline for implementation as well as ensuring that adequate funding is considered for this purpose.

6.1.9.2.4 External data access/study proposal adjudication

If a registry collects high-quality data, it is both likely and desirable that external requests will be received requesting access to data or proposing studies that can utilize registry data. To ensure transparency and facilitate best use of data, it is suggested that the scientific committee establish a formal plan to adjudicate on such requests. This might involve defining the grounds for collaborative agreements where external parties, in addition to gaining access to data, can benefit from the experience and expertise of committee members aware of the context in which the data were collected.

6.1.9.3 Quality assurance Committee

Ensuring that the registry's quality is validated will increase the value of the registry. Though the project management team and scientific committee will together increase a registry's quality, it is advisable to have an independent committee established to assess whether this is the case through the creation of a formal audit and quality assurance plan. In addition, this group might be well placed to handle complaints or to ensure that ethical and legal obligations are being met in the absence of a specific group to manage this.

This group would ideally comprise experts familiar with registry analysis and who have experience of audit and quality assurance. There should also be consultation with regulatory groups to ensure that all regulatory requirements are met; this is of particular relevance when the registry is focused on safety assessment, such as devices. Conflicts of interest should be considered and declared during this process.

Table 6.2: Re-usable table to facilitate selection and recording of possible registry team members

Registry teams	Technical expertise required	Name	Group	Comments
Project Management				
To oversee the management of the overall project				
Clinical/Subject Matter Board				
To determine the scope of the data captured				
Scientific Committee				
To guide scientific utilisation of registry data assess external applications for utilisation of data	Health Outcomes			
	Epidemiology			
	Statistics			
	Data mining			
	Data standards			
	Social Media			
	eCommerce			
Data Collection & Database Management Board				
To guide with respect to attainment of best data standards	Data standards			
	Data linkage			
	Data quality			
	Databases			
	Data security			
	Data mining			
	Clinical Standards Manager			
Legal/patient privacy				
To ensure compliance with legal requirements	Health information Act		Health Providers	
	Health law		Regulatory Bodies	
	Project management lead		Data Commissioner	
Quality Assurance & Liaison				
To ensure the quality of the registry is maintained	Data quality expert		Regulatory Bodies	
	Epidemiology			
	Patient representation			
	Health Outcomes			
Other Comments				

6.1.10 Resource requirements

Resource requirements will vary significantly depending on the scope of the registry project. The steps followed to this point should identify the extent of resourcing that will be required to meet the objectives outlined. Resources to consider include:

6.1.10.1 Human Resources

Though registry committees can perform a large quantity of work, there is likely to be either dedicated or intermittent need for staff to ensure proper set-up and maintenance of a high-quality registry. Depending on the scope of the registry project, this might include staff to meet needs in:

- Administration
- Project management
- Data management
- Data collection
- Study design, epidemiology & statistical support
- Data dissemination
- Programming
- Design (question and graphic)
- Training
- Financial
- Legal/data security & protection
- Clinical

6.1.10.2 Information Technology Resources

Depending on the environment in which the registry is to be established, requirements can range from analysis software to an extensive hardware and software budget. It should be stressed that information technology support with experience of registry design is extremely valuable. Gaining advice from other registries, registry groups such as PARENT and local regulatory bodies is invaluable and should be sought to ensure any system delivered is designed appropriately and with interoperability in mind.

6.1.10.3 Financial Resources

Though the outlay for the initial design and implementation of a registry is the most obvious requirement, consideration should be given to the long-term sustainability of the registry project. Financial resources will vary significantly depending on the scope of the registry; however, by following a planning process with an inclusive stakeholder assessment, it is more likely to identify appropriate funding avenues and collaborations that may maximise financial investments in addition to the financial value of registry outputs. Examples of funding avenues include public-private partnerships, governmental funding, patient groups, and sponsorship from charities or pharmaceutical companies.

Finally, it is necessary to take account of the financial implications of closing down a registry and what arrangements would need to be made to ensure that data security is maintained in this scenario.

6.1.10.4 Other Resources

The list of potential other resources is extensive, however, particular note should be drawn to office space. It is ideal if the registry emerges from a group that can provide accommodation, which is important for a number of reasons, including from a data security perspective.

6.1.11 Funding Strategy

It is likely that for each group which proposes a registry, there is a funding source that has helped bring the idea to this stage. By including a directed stakeholder evaluation, it is likely that further opportunities might present themselves. Of particular significance, however, is the need to consider how funding might influence how the outputs of the registry are interpreted. At all times, funding should be arranged in a manner that is transparent and without conditions that might undermine the validity of the scientific study.

6.1.12 Risks and feasibility

Risks accompany each component of the registry establishment and maintenance process, from excessive dataset selection and lack of adherence to recognized standards through to a failure to consider a registry termination strategy. Of all these, however, we suggest that a failure to be aware of the extensive preparations required to develop and maintain a registry are the most significant. The process described in this section may seem over-burdensome, but it can present a myriad of advantages, such as identifying collaborative opportunities and identifying funding opportunities. Apart from this, as registry science evolves, regulation is likely to follow and create obstacles that might threaten the creation and survival of registries that have already consumed significant resources. As such, we recommend that an extensive planning process be undertaken under the guidance of experts familiar with the process of registry design and with stakeholders. Once this has been completed, an informed feasibility assessment can be undertaken. This should review whether the objectives and purpose of the registry are likely to be met within the timeline considered, the budget available, the scientific model proposed and within the environment in which the proposed registry is due to be implemented.

6.1.13 Developing an Implementation Plan

If the feasibility study reaches a positive conclusion, it is likely that most of the components will now be adequately developed to create an implementation plan, which should also include a plan that details how and when the registry will be evaluated. It is suggested that a further review of the steps involved in planning the registry is undertaken to develop an action plan and timeframe for each step in conjunction with the appropriate expert or stakeholder identified by the planning process. Within this, rate-limiting steps should be identified to help determine the “critical path” which will dictate how long the project is likely to take. It is suggested that at this point, particularly in the case of a large registry project, an experienced project manager is involved to help deliver the project on schedule.

As part of the implementation plan, it may be useful to consider a pilot project as a proof-of-concept model before proceeding with a full implementation. This can generate significant support for a registry, create useful outcomes and identify significant obstacles that may not have been initially obvious. It can also create a wealth of knowledge and experience at a manageable level that can increase the chances of ultimate success.

A project proposal should be formalized with firm time and budgetary constraints outlined to facilitate regular oversight by the project management committee (or similar). Though numerous measures of quality have been mentioned, ultimately, the registry will need to be regularly evaluated against the objectives and purpose it was designed to meet. This can facilitate review and adjustment of the registry that can further improve outcomes, efficiency and ensure that relevance is maintained.

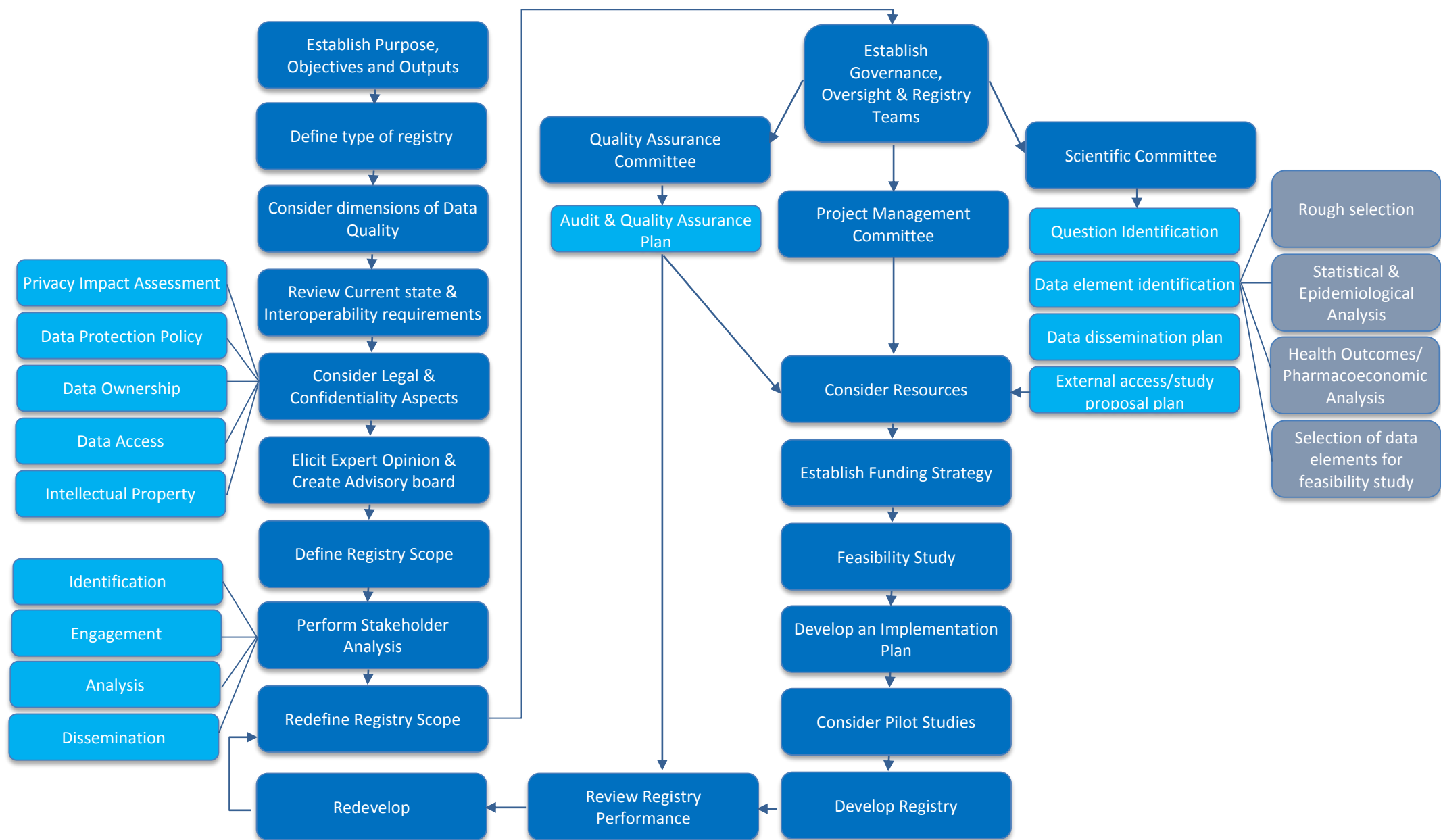


Figure 6.1: Planning a Registry Process

References

1. PARENT (PATient REGistries iNiTiative). PARENT Pilot Registry of Registries 2014 [24th May 2014]. Available from: http://www.parent-ror.eu/-/state/list_all.
2. Agency for Healthcare Research and Quality. RoPR | Registry of Patient Registries: U.S. Department of Health & Human Services; 2014 [24th May 2014]. Available from: <https://patientregistry.ahrq.gov/>.
3. Gliklich R, Dreyer N, Leavy M, eds. Registries for Evaluating Patient Outcomes: A User's Guide. Third edition. Two volumes. (Prepared by the Outcome DEClDE Center [Outcome Sciences, Inc., a Quintiles company] under Contract No. 290 2005 00351 TO7.) AHRQ Publication No. 13(14)-EHC111. Rockville, MD: Agency for Healthcare Research and Quality. April 2014. <http://www.effectivehealthcare.ahrq.gov/registries-guide-3.cfm>.
4. ISPOR Taxonomy of Patient Registries: Classification, Characteristics and Terms. Lawrenceville, NJ2013.
5. Health Information and Quality Authority. Guiding Principles for National Health and Social Care Data Collections. Dublin, Ireland: HIQA; 2013.
6. Health Information and Quality Authority. What you should know about Data Quality. Dublin, Ireland: HIQA, 2012.
7. IEEE Standards Glossary: IEEE; [22 May 2014]. Available from: http://www.ieee.org/education_careers/education/standards/standards_glossary.html.
8. Foy R, Hempel S, Rubenstein L, Suttrop M, Seelig M, Shanman R, et al. Meta-analysis: effect of interactive communication between collaborating primary care physicians and specialists. *Annals of Internal Medicine*. 2010;152(4):247-58.
9. Shojania KG, Duncan BW, McDonald KM, Wachter RM, Markowitz AJ. Making health care safer: a critical analysis of patient safety practices: Agency for Healthcare Research and Quality Rockville, MD; 2001.
10. Institute of Medicine. Crossing the quality chasm: A new health system for the 21st century. Washington: National Academies Press, 2001 0309072808.
11. Levit L, Balogh E, Nass S, Ganz PA. Delivering high-quality cancer care: charting a new course for a system in crisis. Institute of Medicine Washington, DC: Institute of Medicine. 2013.
12. Caulfield BM, Donnelly SC. What is Connected Health and why will it change your practice? *QJM*. 2013;106(8):703-7.
13. ECHAlliance | Delivering leadership for the development of Connected and MHealth market and practice across Europe 2014 [24th May 2014]. Available from: <http://www.echalliance.com/>.
14. Erickson JI. Caring for patients while respecting their privacy: renewing our commitment. *Online journal of issues in nursing*. 2005;10(2).
15. Health Information and Quality Authority. Guidance on Privacy Impact Assessment in Health and Social Care. Dublin, Ireland: HIQA; 2010.
16. Health Information and Quality Authority. International Review of Privacy Impact Assessments. Dublin, Ireland2010.
17. Health Information and Quality Authority. Sample Privacy Impact Assessment Report Project. Dublin, Ireland2010.
18. Health Information and Quality Authority. Guidelines for Stakeholder Engagement in Health Technology Assessment in Ireland. Dublin, Ireland: HIQA; 2014.
19. Deverka PA, Lavalley DC, Desai PJ, Esmail LC, Ramsey SD, Veenstra DL, et al. Stakeholder participation in comparative effectiveness research: defining a framework for effective engagement. *Journal of comparative effectiveness research*. 2012;1(2):181-94.

20. Irish Platform for Patients' Organisations, Science & Industry [07/12/2013]. Available from:
<http://www.ipposi.ie/>.
21. Irish Platform for Patient Organisations, Science and Industry,. Towards a National Strategy for Patient Registries in Ireland. Dublin, Ireland: 2008.
22. Donohue F. Presentation IPPOSI Meeting Dublin 10 May. Dublin 2011.

6.2 Registry research design

Marcel Kralj

Early in the registry development phase it is necessary to determine some kind of an overall research plan/design that defines the registry characteristics and (future) operation from a more research-methodological point of view. Various elements have to be considered at this stage of registry development.

Key principles:

- Research questions or hypotheses should be properly formulated.
- It is essential to clearly define the target population of the registry. Registry should be also defined in terms of geographical and organizational coverage.
- Definition of cases that are going to be included in a registry should exist. Inclusion and exclusion criteria have to be determined and clearly stated.
- It is necessary to understand which study model can be applied in a registry (e.g. cohort, case-control, nested case-control etc.).
- Anticipated registry size and duration should be estimated.
- Registry data collection procedure has to be determined. It must support the highest possible data quality, lowest possible burden for the reporting units and lowest possible costs for the registry.
- In case of follow-up a clear strategy should exist.
- Thorough documentation for the entire data collection protocol, including guides for data providers and other users should be prepared.
- Representativeness and generalizability of a registry should be considered and appropriately described for data interpretation purposes.
- During registry design phase time, costs and registry resources need to be constantly taken into account.

When the purpose and main objectives of the registry are defined, the next step is to define the data to be collected, and determine the methodology/protocol with which the registry will try to achieve the defined goals. At this point, the registry holder needs to consider many issues, including the defining of the registry target population, anticipated registry size and duration, study design, data sources for the registry, registry dataset and data collection methods/procedure. At the same time, the registry holder needs to look at the registry resources, costs and consider the quality aspect. This chapter describes those registry's elements and covers the important aspects that are necessary to take into account during that development stage.

6.2.1 The population covered by a registry

Enrolment of the patients for a registry starts with a clear understanding of the target population, which is a population to which the registry would like to generalize its results and findings (e.g. patients with multiple sclerosis in Slovenia). When building a registry it is important to accurately define the target population since it is a key factor in forming the registry population. The registry holder should understand and determine whether the registry is a hospital-based registry⁴⁸, population-based registry or even a population registry⁴⁹. It is necessary to define the registry in terms of geographical and organisational coverage.

In addition to target population, it is recommended that a registry provides a case definition which is a detailed specification of the patients/cases that are going to be included in a registry. The registry team should specify so-called eligibility or inclusion criteria that are a set of conditions that a patient must meet to be eligible for inclusion in a registry, and generally include geographic (e.g. hospitals in a particular region of the country), demographic (e.g. age, gender, ethnicity), disease-specific (e.g. a certain diagnosis, stage of disease), time-specific (e.g. specification of the included dates of hospital admission), laboratory-specific, and other criteria (e.g. size of the hospital in terms of number of patients) (2, 35). Exclusion criteria, on the opposite side, are those criteria that disqualify subjects from inclusion in the registry. Inclusion and exclusion criteria often reflect considerations such as cost and practical constraints (sometimes subjects are not included, not because they are out of interest, but due to the additional cost or burden of including them), ethical concerns, people's ability to participate (e.g. their health condition may prevent participation), and design consideration (it is sometimes advantageous to have a more homogeneous population as a means for reducing confounding, but in terms of generalizability, stringent inclusion criteria might reduce the generalizability of the registry findings to the target population) (24). Inclusion and exclusion criteria should therefore be defined carefully and many aspects need to be taken into account while defining those criteria, as the selection of inclusion/exclusion criteria can optimize the internal validity or generalizability of the registry, improve its feasibility (also in terms of follow-up and attrition), and lower its costs (25). Besides very clear definitions of the inclusion and exclusion criteria it is crucial that criteria are well documented, including the rationale for these criteria.

6.2.2 Anticipated size and duration

Estimation of anticipated registry size is an important part of the planning process. Some registries try to include all cases from the defined population, but often registries include only a sample of a population. In that case it is recommended to prematurely estimate how many cases the registry is

⁴⁸ Registry that aims to record information on all patients seen in a given hospital or group of hospitals irrespective of geographical areas (37).

⁴⁹ Description of the population registry and population-based registry is provided in chapter 2.2: 'Types of patient registries'.

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. Hence, it is important to adequately plan the registry's size. Various components impact on estimating registry size and need to be considered, including (2, 30):

- the study outcome and its frequency/variability
- size of clinically important effects, the desired precision of estimates (e.g. the width of a confidence interval);
- timeframe (e.g. for analyses, dissemination of results);
- available resources and money, feasibility;
- support for regulatory decision-making (e.g. if registry is intended to support regulatory decision-making, the precision of the estimate is important);
- anticipated drop-out rate

Many methods for sample size calculation exist and are described in general statistics textbooks (31-33). There are also different tools that can assist in sample size calculation. Besides software programs (e.g. G*Power, nQuery Advisor, PASS, STATA) there are also online tools that allow free sample size calculations, such as:

- [Russ Lenth's Power and Sample Size](#)
- [David Schoenfeld's Statistical Considerations for Clinical Trials and Scientific Experiments](#)
- [UCLA Calculator Service](#)
- [The Survey System's sample size calculator](#)
- [Raosoft's sample size calculator](#)

These tools should be used with caution, since they are not always reliable or suitable for any situation.

Although a patient registry is generally considered as a long-term and sustainable action, the anticipated duration of a registry (taking into account the enrolment and follow-up phase) should also be specified when developing a registry. The duration of a registry depends on what type of registry it is, what the specific procedures in the registry are, and what objectives need to be met. Some registries collect data at only one time point while 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 registry's objectives is expected to have accrued (3). If we neglect the funding as the biggest factor for registry duration and sustainability, the factors that the registry holder, together with the key stakeholders, should consider when estimating registry duration, include the induction period for desired outcomes, sufficient follow-up time for the exposure, data collection method; sample size, complexity of data being collected, anticipated accrual of enrolled subjects, and deadlines for dissemination of results (2, 35).

It is worthwhile to note that registry size can also refer to the number of sites included in a registry, and to the volume and complexity of data being collected (3). Hence, a registry holder can consider these perspectives as well.

6.2.3 Registry dataset

The registry needs to develop the dataset that will serve the purpose and objectives of the registry. Although some key variables/data elements can be identified and determined soon in the developing

process, this can be a very lengthy activity and should not be underestimated since it is the registry dataset which will eventually determine the usefulness and success of the registry. More information on developing a registry dataset is available in chapter [6.3](#).

6.2.4 Data collection procedure

The decision on how the registry will collect the data is affected by several factors, namely the characteristics of the registry's target population, the information that needs to be obtained and other specific goals of data collection, available data sources, registry resources and time limits. The registry data collection procedure must support the highest possible data quality, lowest possible burden for the reporting units and lowest possible costs for the registry. The registry holder needs to identify and evaluate all available data sources and determine which one will be used. The registry must make an agreement with the data providers and develop the technical protocol⁵⁰ for the data acquisition. (More information on data sources is available in chapter [6.4](#) 'Data sources for registries'.)

When developing a registry data collection procedure, the registry should take into account the technological aspect of data collection (e.g. paper-based forms, web-based data entry, use of personal computers, handheld computers, scanners, mobile phones) and be aware of advantages and disadvantages of both, paper-based and electronic approaches. The choice of which system to use depends on where the data are captured, by whom, and what resources are available for the particular reporting unit. It is important that the approach is practical and reliable. In addition, a registry designer needs to look also from the perspective that is especially well-covered in the field of survey methodology, where a great emphasis is placed on the modes of data collection, their characteristics and principles of good practice. This includes, for example, the consideration as to whether the case report forms or questionnaires are understandable and easy to use, questions or instructions are worded correctly, whether they are measuring the right things, whether the presence of the interviewer/data collector (e.g. nurse) would influence a patient's answers; the self-administered mode would yield more honest answers or produce a lower response rate, telephone data collection could be used to obtain data more cost-effectively, this mode would enable response from all patients, etc.

Registry data collection can be transversal, where all defined patients are registered once, or longitudinal, where the data are collected at different time points for the same patient. In case of the longitudinal design, the registry should carefully determine (a) which data needs to be (re)collected, (b) at what time points (e.g. every 6 months), (c) how long (e.g. for 10 years) and (d) with what means (e.g. with the telephone, by visiting a general practitioner, by data linkage to other records). When developing the follow-up strategy it is important to consider the costs which can increase significantly when the follow-up is implemented via personal contact, the extra work that will be put on the data providers and the burden that will be imposed on the patient. The latter can quickly become an issue as the preparedness of the patient to provide data is easily exhausted. This may result in loss to follow-up which can lead to the biased results, especially if these losses are not random. For example, if in a follow-up process only data from satisfied patients with encouraging outcomes are obtained, meanwhile unsatisfied patients with less promising outcomes do not want to participate in a follow-up, then the registry does not reflect the true picture. The registry should therefore, develop a good patient retention plan that is suitable to the target population.

⁵⁰ Protocol that includes a requirement for granting access, username and password creation, etc.

In all this, a registry should prepare thorough documentation for the entire data collection procedure and provide methodological guides/standard instructions and rules for data collectors/providers and other data users. This typically includes information on reporting dynamics, what data needs to be collected and how, means of data transmission, established controls for the acquired data (e.g. readability of data, adequacy of records and their number) and access rights. It is often advisable to describe also the typical data flow of the registry, where the information on how the data travels from the source to the registry, together with the other additional information (e.g. key persons/stuff included in the process, type of technology and data collection method used, access rights, data transmission, timetables) is clearly specified. The description of the data flow can help the registry team and other stakeholders (e.g. company that will provide the technical solution) to better understand the whole data collection protocol. Among the other things, it can serve also when performing evaluations of the data collection protocol (e.g. identification of potential sources of errors etc.)

Table 6.3: Example of the data flow description

- | |
|---|
| <ul style="list-style-type: none"> - In the hospital a nurse collects the data from the patient via the paper-based form; - the data are then entered into the web-based system; the nurse and the doctor are the only ones who have access to the data and can modify the data; - the data are transmitted via web server to the central database; - after 6 months the patient is contacted by telephone and asked three additional questions; data are collected via paper-based form and then entered into the web-based system; - ... |
|---|

6.2.5 Research-based registries - additional points to consider

Nowadays, many registries are being developed that are taking a more research approach. These study-oriented or research-based registries possess different characteristics, therefore some additional points need to be considered when developing this type of registry. However this does not mean that points described below should be entirely ignored by registry holders who aim to develop more 'classical', wide encompassing registries. All in all, also the latter can be seen in some way as research-based registries (i.e. there are always some research questions that that registry tries to answer).

6.2.5.1 Research questions and hypotheses

When the purpose and main objectives of the registry are clearly defined the next step is to take that purpose or idea and shape it into a researchable question. Research questions and hypotheses narrow the purpose of the study and become major 'signposts' for guiding the overall study (1).

Research questions for registries range from purely descriptive questions aimed at understanding the characteristics of people who develop the disease and how the disease generally progresses, to highly focused questions intended to support decision-making (2). Research questions in registry-based studies are generally hypothesis generating (i.e. developing hypotheses after the data are collected and new knowledge is gained) or evidence building, rather than hypothesis testing. However, registries focused on determining clinical effectiveness, cost-effectiveness or risk assessment are commonly hypothesis driven (2-4). Regardless of the nature of research questions (or hypotheses) it is crucial for a registry planner to define them because all further decisions (e.g. registry population, what data will be collected and analysed) and work in a registry development process are guided by research

questions of interest. Proper formulation of a research question or hypothesis is not an easy task and should not be underestimated. An improperly defined, unfocused or underdeveloped research question or hypothesis can generate a risk for not getting the right results and accomplished objectives of a registry. Accordingly, it is highly recommended that a registry developer invests/spends required time to suitably develop a research question or hypothesis.

(Research) ideas as a foundation for developing a research questions or hypotheses are typically gathered by literature review, critical appraisal of the published clinical information, brainstorming with colleagues, seeking experts' opinions, and evaluating the expressed needs of the patients, health care providers (2, 5). The clinical questions of interest can also be defined by payers/sponsors of the registry. Thus, it is not uncommon that multiple questions are set as a result of the interests of different stakeholders. In that case a registry planner should be aware that a higher number of research questions can increase the complexity of a registry study design and subsequent collection of data and statistical analysis. Registry developers should therefore assess whether it is feasible to answer every question of interest.

When defining research questions or more specific research hypotheses it is important that they are accurate, understandable and focused enough for a specific registry. The clinical epidemiology literature offers various instructions on research questions and hypotheses, such as, for example, FINER (6) and PICOT (7) criteria for a good research question. An example of a research question and hypothesis for a registry is presented in [Table 6.4](#).

Table 6.4: Example of a research question and hypothesis for registries

Idea/Interest/Purpose	Research questions/hypotheses
Monitoring clinical effectiveness of hip implants	Hypothesis: In Europe, exchangeable neck hip stem implants have significantly higher revision rate than hip implants with un-exchangeable neck.
Natural history of patients with diabetes disease	Research question: What is the incidence and prevalence rate for diabetes type 1 disease among children and adults in Slovenia?

6.2.5.2 Key exposures and outcomes

In a simplified way we can describe the exposure and outcome as a relationship, where one event (i.e. exposure) affects the other (i.e. outcome). In the field of patient registries, the term 'exposure' refers to treatments and procedures, health care services, diseases, and conditions, while outcomes generally represent measures of health, onset of illness or adverse events, quality of life measures, measures of health care utilization and costs (2).

It is crucial to identify the key exposures and outcomes at the very beginning of a registry development, since the selection of exposures and outcomes will affect further registry development (e.g. registry study design, data collection process). The identification of key exposure and outcome variables is guided by the registry research questions that are defined at the registry's outset. When identifying the key exposures and outcomes it is important to know that sometimes more outcomes 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 (2, 8). For example, if we select smoking cigarettes as an exposure for measuring a particular outcome (e.g. heart disease) probably it would not be enough to have only one binary variable for exposure (i.e. smoking

or non-smoking), but to include also other information, such as dose (e.g. how many cigarettes per day) and duration (i.e. how many years of smoking). During the identification of key exposure variables it is therefore necessary to consider also this aspect, and it is useful to take into account independent risk factors for the outcomes, and confounding variables as well. More information on selecting data elements for a registry is provided in chapter [6.3](#).

6.2.5.3 Study design

Registry studies are observational studies in which the researcher merely observes and systematically collects information, and, unlike in the experimental studies, does not assign specific interventions to the study subjects being observed. In observational studies the researcher chooses what exposures to study, but does not influence them.

Although patient registries are generally considered as prospective observational studies, the registries, from the time perspective, could be both – prospective and retrospective studies. Prospective studies are designed to gather data about events that have not happened yet, while retrospective studies are designed to gather data about events that have already happened. Thus, prospective studies look forward in time and retrospective studies look backward (9).

It is not always simple to define which study design⁵¹ the registry follows, using traditional epidemiological terms. For example, in some situations study design for a registry can be considered as an opened cohort or simply a case series of patients under some specific diagnosis (22). Sometimes even the registry's nature itself does not require clear specification of its study design. However, it is necessary for a registry designer to understand which study model can be applied in a registry. Several study designs that are more commonly applied in registries are cohort study, case-control study, nested case-control study, case-cohort study, and case series. Besides these, also some other designs are sometimes used, such as cross-sectional study and case-crossover design. Readers are encouraged to consult textbooks and articles of epidemiology for more information on study designs (2-3, 10-20).

6.2.5.4 Comparison groups

A registry can also include and collect data on one or more comparison groups. Although registries usually do not use comparison groups, they are essential when it is important to distinguish between alternative decisions, to assess the magnitude of differences, or the strength of associations between groups. Based on the registry's objectives three types of comparison groups can be used:

- internal comparison group (data are collected simultaneously for patients who are similar to the focus of interest, but who do not have the condition or exposure of interest),
- external comparison group (data have been collected outside the registry for patients who are similar to the focus of interest, but who do not have the condition or exposure of interest),
- historical comparison group (refers to patients who are similar to the focus of interest, but who do not have the condition or exposure of interest, and for whom information was collected in the past, for example, before the introduction of an exposure or treatment or development of a condition)

⁵¹ 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 (21).

When deciding about including a comparison group in a registry, the registry developer should consider also that adding a comparison group may add complexity, time, and cost to a registry (2).

6.2.5.5 Sampling frame and sampling method

Registries sometimes try to include all units of the target population, but often they include just a sample of the target population from which inferences about the whole population can be made. The need for including only a sample of the target population typically arises because of limitations of time and resources but also due to other constraints (26). The activity of selecting cases (i.e. patients, institutions, object or events) into a sample from a larger collection of such cases, according to a specific procedure, is called sampling. Ideally the sample is drawn directly from the target population but usually this is not the case, because a sample can be drawn only from cases to which registry/participating sites have access (i.e. accessible population). Hence, the accessible population represents the sampling frame from which a sample is selected. Sometimes the accessible population is the same as the target population, but usually is a subset of the target population. In terms of a precision of registry's estimates/results, a registry planner should be aware of this issue, since non-coverage of certain parts of a target population can lead to biased estimates (27, 28). In other words, if cases of the target population who cannot be sampled (because there is no access to them) are different from those who can be drawn into a sample, the registry findings can be biased. During a sampling phase a registry planner needs to assess what impact on the registry findings a sampling frame and its potential non-coverage issue could have.

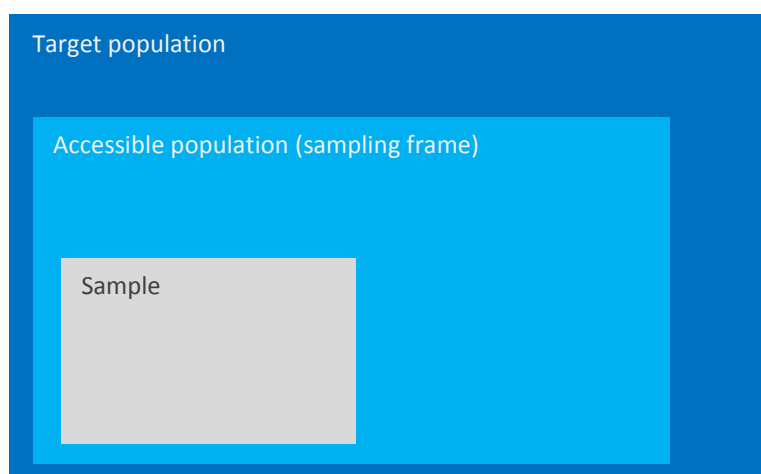


Figure 6.2: Key concepts in sampling (adapted from Sim J and Wright C, 2000)

Many different sampling methods can be used when selecting cases for a registry. Sampling designs are classified as either probability sampling or nonprobability sampling. In general, probability sampling is the preferred method, in which the selection of individual cases (e.g. patients, events) is left to chance, rather than to the choice or judgement of the person. However, in some situations probability sampling is not feasible and nonprobability sampling is more useful. Some sampling methods that are often used for generating samples include simple random sampling; stratified random sampling; systematic sampling; cluster sampling; multistage sampling; case series or consecutive (quota) sampling; haphazard, convenience, volunteer, or judgmental sampling; modal instance; purposive; and expert sampling (2).

6.2.5.6 Representativeness and generalizability

When selecting patients, hospitals or events it is important that consideration about representativeness is made, since the representativeness is essential component of a registry study. If the sample is not properly representative, conclusions/generalization may be incorrect. The registry developer should consider representativeness in terms of patients (e.g. men and women, children, the elderly, racial and ethnic groups), sites (e.g. geographic location, practice size, academic or private practice type) and events (e.g. type of events/services on a particular day) (2). The registry developer should critically assess how the potential lack of representativeness can affect the results of a registry. For example, suppose that the purpose of the registry is to monitor the clinical effectiveness of specific surgeries. If a registry includes only academic centres/hospitals with high technical support, then the results probably would not reflect a true picture. On the other hand, for example, when a registry is not representative in terms of gender (e.g. a higher number of women in a registry), this would have no impact on the representativeness of the registry findings if the outcome that is observed (e.g. clinical effectiveness of a specific drug) does not vary with gender.

Associated with the representativeness, the generalizability concept is often used, which refers to the extent to which the conclusions of the registry study can be generalized/applied to populations other than those sampled and included in the registry. Strong generalizability or external validity is achieved by the inclusion of a typical patient sample which is often more heterogeneous (e.g. different demographic characteristic, comorbidity). Patient registries are generally designed to have strong external validity so that their population will be representative and relevant to decision makers. It is important to note that the way in which patients are included, classified and followed directly affects generalizability (2, 3). In terms of data interpretability it is important to describe and document the representativeness and generalizability of a registry, and whether it covers the relevant patients, events and periods of interest.

References

1. Creswell John. Research Design: Qualitative, Quantitative and Mixed Method Approaches, 3rd Edition
2. Gliklich RE, Dreyer NA, eds. Registries for evaluating patient outcomes: A User's Guide. 3rd ed.
3. ISPOR: Taxonomy of patient registries: classification, characteristics and terms.
4. The Yeshiva Fatherhood Project. Introducing qualitative hypothesis-generating research.
5. Lobiondo-Woow G, Haber J. Chapter 2. Nursing Research - Methods and Critical Appraisal for Evidence-Based Practice. 2013.
6. Hulley S, Cummings S, Browner W, et al. *Designing clinical research*. 3rd ed. Philadelphia (PA): Lippincott Williams and Wilkins; 2007.
7. Brian Haynes R. Forming research questions. *Journal of Clinical Epidemiology* 2006; 59:881-6.
8. International Agency for Research on Cancer. Cancer Epidemiology – Principles and Methods.1999.
9. Stark Nancy J. Registry Studies for Medical Devices – Whitepaper and Workshop Invitation. 2010.
10. Verhamme K. Study designs in Paediatric Pharmacoepidemiology. *European Journal of Clinical Pharmacology* 67, S1 (2010) 67-74
11. Song Jae W, Chung Kevin C. Observational Studies: Cohort and Case-Control Studies. *Plast Reconstr Surg*. 2010 December, 126(6): 2234–2242
12. Rose S, van der Laan M. J. Why Match? Investigating Matched Case-Control Study Designs with Causal Effect Estimation. *The International Journal of Biostatistics*, Volume 5, Issue 1 2009 Article 1.
13. Carlson D. A. M, Morrison S. R. A User's Guide to Research in Palliative Care: Study Design, Precision, and Validity in Observational Studies. *JOURNAL OF PALLIATIVE MEDICINE*, Volume 12, Number 1, 2009
14. Jepsen P., Johnsen S. P., Gillman W. M., Sorensen H. T. Interpretation of observational studies heart 2004; 90:956–960.
15. Ernster, V. L. Nested case-control studies. *Preventive medicine*, 23, 587-590, (1994).
16. Langholtz, B. Case-Control Study, Nested. Volume 1, 646-655. In *Encyclopedia of Biostatistics*, 2nd Edition. Eds. Armitage, P. and Colton T. John Wiley & Sons, Ltd, Chichester, 2005.
17. Wacholder S, Silverman DT, McLaughlin JK, Mandel JS. Selection of Controls in Case-Control Studies. *American Journal of Epidemiology* Vol. 135, Mo. 9. 1992
18. Kooistra B, Dijkman B, Einhorn TA, Bhandari M. How to design a good case series. *The Journal of bone and joint Surgery*. 2009, 91 Suppl 3:21-6.
19. Dekkers OM, Egger M, Altman DG, Vandenbroucke JP. Distinguishing case series from cohort studies. *Annals of Internal Medicine*, 2012 Jan 3;156(1 Pt 1):37-40
20. Andrews N. Epidemiological designs for vaccine safety assessment: methods and pitfalls. *Biologicals*. 2012 Sep;40(5):389-92
21. WHO. Training manual for community-based initiatives: A practical tool for trainers and trainees. 2006.
22. EPIRARE. Deliverable D4: Guidelines for data sources and quality for RD Registries in Europe. 2014.
23. Laforet P et al. The French Pompe registry. Baseline characteristics of a cohort of 126 patients with adult Pompe disease. *Revue Neurologique*, Vol 169, Issues 8–9, pages 595–602. 2013.
24. Polit DF, Beck CT. *Nursing Research: Generating and Assessing Evidence for Nursing Practice*. 8th Edition. Lippincott Williams and Wilkins, a Wolters Kluwer business.

25. Eduardo Velasco. Inclusion Criteria. In Encyclopedia of Research Design, eds. Neil J. Salkind. 2010. SAGE Research methods. Available at: <http://srmo.sagepub.com/view/encyc-of-research-design/n183.xml>
26. Sim J, Wright C. Research in Health Care: Concepts, Designs and Methods. 2000. Stanley Thornes (Publishers) Ltd.
27. Brick, Michael J., Ismael Flores-Cervantes, Kevin Wang in Tom Hankins. 1999. Evaluation of the Use of Data on Interruptions in Telephone Service. Proceedings of the Survey Research Methods Section of the American Statistical Association.
28. Groves, Robert M. 1989. Survey Errors and Survey Costs. New York: Wiley.
29. Cochran WG. Sampling Techniques. Third ed. Wiley; 1977.
30. Noordzij M et al. Sample size calculations: basic principles and common pitfalls. Nephrol Dial Transplant (2010) 25: 1388–1393.
31. Altman DG. Practical Statistics for Medical Research. London, UK: Chapman & Hall; 1991.
32. Bland M. An Introduction to Medical Statistics. 3rd ed. Oxford, UK: Oxford University Press; 2000.
33. Lwanga SK, Lemeshow S. Sample size determination in health studies - A Practical Manual. World Health Organization 1991.
34. Carey TS, Sanders GD, Viswanathan M, et al. Methods Future Research Needs Reports, No. 8. Rockville (MD): Agency for Healthcare Research and Quality (US); 2012 Mar.
35. Gregg, M.B. Field Epidemiology. New York: Oxford University Press, 2002
36. Statistični urad Republike Slovenije. Metodološki priporočniki: Smernice za zagotavljanje kakovosti, št 2. Ljubljana, 2012.
37. European Network of Cancer Registries. 2011. Guidelines on Confidentiality and ethics for population-based cancer registration and linked activities in Europe.

6.3 Registry dataset

Marcel Kralj

Selecting the appropriate datasets to produce relevant and valid indicators that inform decision making is one of the most critical tasks in building a registry. The process of developing registry dataset is a lengthy task, and it typically requires a team work of clinical experts, health informatics, statisticians and epidemiologists who need to consider many different aspects during this process.

Key principles:

- General principles for developing data elements' definitions and value domains should be followed.
- Setting validation rules for data elements is highly recommended as it helps to reduce various errors, ensures internal consistency and in general improves the data quality.
- Many standard data elements, definitions, classifications, clinical terminologies and common datasets already exist. When developing registry dataset these standards should be examined and used whenever possible.
- During registry dataset creation it is necessary to consider the costs and burden of data collection and also what is the expected quality (i.e. can reliable information be collected) and coverage for the individual data element.
- Registry dataset should be tested before it is really used.
- It is crucial to understand registry dataset. Clear methodological guide and data dictionary should be prepared and easily accessible to all relevant stakeholders.

Before deciding on what data to collect in a registry, it is important to be clear about the purpose of the registry. Once the registry's purpose and goals are determined, the data that are required to meet those objectives can be identified. The selection of data elements⁵² 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 (1-3):

- **Patient domain** (Data that describe the person, such as demographic information, contact information, information about medical history, health status and patient identifiable information. The inclusion of a patient identifier is often necessary and can bring many advantages. For example, it enables linkage to other data sources; tracking a patient through time and place; it enables gathering information when the personal contact with patients is required; and data quality checks. However, the main issue in that field is the privacy aspect. More about this is described in chapter 5.)
- **Provider domain** (Data that describe the characteristic of the individuals providing health care interventions to the patients included in the registry)
- **Exposure domain** (Data that describe the patient's experience with the product, disease, device, procedure, or service of interest to the registry.)
- **Outcome domain** (Data that are of main interest to the registry. Often this refers to measures of health, onset of illness or adverse events, quality of life measures, measures of health care utilization and cost.)
- **Covariate/confounder domain** (Data that are not of primary interest to the registry but their inclusion and measuring is still important, since they are related to the exposure or outcome or both. Inclusion of covariates allows controls/adjustments during analyses.)
- **Administrative domain** (Information related to registration process; for example, date of previous or next follow-up, date of reminder.)

After the identification of data domains for a registry, decisions about which data specifically will be collected by a registry need to be made. 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 a data collection process inducing higher costs and burden, it may jeopardize 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 process of building a dataset is undertaken by a team which typically includes clinical experts, health informatics, statisticians and epidemiologists. During the process various tools can be used, such as mind mapping (e.g. XMind, FreeMind) or/and spreadsheet (e.g. Microsoft Excel) tools.

6.3.1 General principles for building a registry dataset

Minimalist approach in building a dataset

Data elements need to be carefully considered in relation to the purpose of the registry. Every data element must support the purpose and goals of the registry. If there is no strong argument for its collection, it should not be included.

⁵² Data element is any named unit of data used to record information in a registry or database. It is characterized by a name, a definition, representation terms and the set, range and/or format values (8). The term is often used interchangeably with 'variable'. Patient's date of birth is an example of a data element.

The burden and costs for data collecting

The success or failure of a registry is often determined by the costs and burden of data collection. When building a dataset for a registry it is necessary to consider the burden of data collection that will be put on a patient, physician/health provider, and a registry team as well. The likelihood of loss to follow-up or limited usability due to the burden of data collection should be also considered.

Availability of data sources for data elements

It is recommended to identify existing data sources and assess their usefulness. Linkage to other data sources can significantly lower the cost and burden of data collecting.

Privacy aspect

During the selecting and developing data elements, registry planners must take into account security policies and privacy issues. They must assess whether the dataset complies with information privacy principles, and how the inclusion of data elements that are private or confidential in nature will affect the patient's response.

Consideration of data quality for data elements

Data elements of uncertain quality or coverage should not be included in the registry dataset. Unless reliable information can be collected on a majority of cases, the item should not be part of a registry dataset.

Use of data standards

The use of data standards is one of the most important aspects in building a registry. Standard data elements and definitions should be used when possible. Standards promote consistency, comparability, and common understanding of data elements. The use of existing data standards, such as classifications, clinical terminologies and common data sets enables comparison of results, data exchange and reuse – the activities that are nowadays invaluable and highly supported by the European Union (see term “semantic interoperability” in chapter 3.2.5).

Explicit definitions

When there are no suitable internationally standardized data elements or they cannot be used in a specific registry, the registry team needs to define and select their own data elements. Definitions of data elements should be explicit and should ensure that there is no variation in concept, collection or format between institutions and individuals collecting and reporting on the data. ISO/IEC (10) specifies requirements and recommendations on the formulation of data definitions that are specified in Metadata Registries. According to the ISO/IEC 11179 – 4 (2004) a data definition should (a) be stated in the singular, (b) state what the concept is, not only what it is not, (c) be stated as a descriptive phrase or sentence(s), (d) contain only commonly understood abbreviations, (e) be expressed without embedding definitions of other data or underlying concepts, (f) state the essential meaning of the concept, (g) be precise and unambiguous, (h) be concise, (i) be able to stand alone, (j) be expressed without embedding rationale, functional usage, or procedural information, (k) avoid circular reasoning, (l) use the same terminology and consistent logical structure for related definitions, and (m) be appropriate for the type of metadata item being defined.⁵³

⁵³ Detailed explanation of the above-mentioned recommendations is provided at: <http://standards.iso.org/ittf/licence.html>.

Selecting value domains, setting validation rules

For each data element a set of permitted values (i.e. value domain) must be determined. A value domain can be enumerated, where the value domain is specified by a list of all its permissible values (e.g. 1=male, 2= female), or non-enumerated, where the value domain is specified by a description rather than a list of all permissible values (e.g. a value domain for person's age might be "18 years and older") (11). It is important that value domains are determined thoroughly and clearly. This can be achieved by the use of various attributes that are associated with a value domain (specifying, e.g., representation class, datatype, format, maximum character quantity, unit of measure). The Australian Institute of Health and Welfare (6) in its data development guide provides some of the key recommendations regarding the value domain, namely (a) ensure that the value domains are consistent and mappable to (inter)national data standards, where these exist, (b) where a classification or code set is used as a value domain, the edition of the classification or code set must be clearly referenced to avoid ambiguities about which edition is in use, (c) permissible values must be exhaustive within the value domain and mutually exclusive, (d) consider the proper degree of granularity, (e) when using 'other', to ensure an exhaustive set of permissible values, using a code value that is contiguous with the last code in the permissible value sequence should be avoided since this allows adding another enumerated category to the list of permissible values without renumbering the codes, and (f) use the supplementary values to capture missing values in order to accommodate statistical analysis.⁵⁴

Setting validation rules is another activity that is highly recommended. Selecting possible ranges of the values (e.g. person's age cannot be above 120 years, body height in centimetres cannot contain more than 3 characters, date of injury cannot be a date from the future) taking into account also internal consistence with regard other variables (e.g. if person is male or his/her age is higher than 55, he/she cannot be pregnant) or any other errors (e.g. empty cell) helps to reduce the number of errors and improve the data quality. This is especially in the case of electronic data collection, where a mechanism can be established to automatically alert a user when information being entered is inconsistent, not within the expected range of values, not given in the correct format etc.

Minimum dataset

The registry team should decide on the minimum/core dataset which is a list of variables that are essential to collect the data for any case/subject. It should be carefully considered and specified whether a data element is mandatory (i.e. always required to collect the data), conditional (i.e. required to collect the data when a certain condition is met) or optional. When a data element is of conditional type, the condition must be clearly documented (e.g. the number of cigarettes smoked daily is required if the patient is a regular smoker).

Modifying data elements

A registry operating over a long period of time will be faced with the possibility that either the data elements or the indicators for policy making that such elements inform will change. When changing the data elements a registry team should try to comply with the existing standards and to retain longitudinal comparability. In any case, it is important that a registry considers the impact that these changes will have on a collection and interpretation of findings. (See also the chapter [9.1](#) 'Changing an existing registry'.)

⁵⁴ More information on value domain concept and its attributes can be found at:
<http://www.aihw.gov.au/WorkArea/DownloadAsset.aspx?id=6442458038>
<http://standards.iso.org/ittf/licence.html>

Testing dataset

When the first version of the registry dataset is developed, it should be tested. Each data element should be checked separately whether its definition, value domain, any rules or other descriptions are properly determined, comprehensive and understandable. Looking on the entire dataset, a registry team should check the overall consistency of the dataset, assess the data collection burden and evaluate the possibility of making errors in the data collection process.

Methodological guide

Normally, every dataset, together with the data collection process, requires a methodological guide that includes detailed information about what is collected and how. It is used to provide the user with advice or interpretation on how to treat particular data elements and successfully perform the data collection. The guide may include (a) the interpretation of data element's definition and value domain, (b) the explanation of what exactly is collected/included in the observation and what is not, covering all unclear cases/situations, (c) the introduction of rules and restrictions for specific data elements, including the information about the data element's format and about whether the data element is of a mandatory type, and (d) the information about the data collection and data reporting, such as who should collect the data on the specific data element/variable, when he/she should collect the data and by which method/instrument, who is obliged to report the collected data and what are the dynamics of the reporting.

Well-documented and accessible data elements

Data elements should be well-documented and readily accessible to everyone who is interested in a registry's dataset. Well-documented and transparent data elements give an understanding of the collected data and ensure consistency in the data collection process. Visibility and usability of the dataset are important characteristics, meaning that the dataset can be easily noticed and reused by others. This promotes standardisation and comparability. Hence, it is important that a registry establishes a data dictionary which is the inventory of all data elements/variables included in a registry (see chapter [6.5.6.6](#) 'Data dictionary').

(2, 4-11)

6.3.2 International coding systems, terminologies and common data sets

As already mentioned, a registry should use existing standards wherever possible since this facilitates consistency, comparability, data exchange and reuse. When developing a registry dataset, the registry developers and steering committee should together identify the existing standards that could potentially be used, and determine the most advisable standard to adopt. [Table 6.5](#) presents several international standard coding systems and terminologies that are widely used in the health domain (see also chapter [3.2.5.1](#) 'Standards, models and tools' and [10.11.2](#) 'eHealth standards')

Table 6.5: International coding systems and terminologies

Area	Standard	Developer	Website
Diseases	ICD-10-CM ICD-9-CM ICD-O	WHO	www.who.int/classifications/icd/en
	ORPHA-codes	ORPHANET	www.orpha.net
Medical Nomenclature	SNOMED	International Health Terminology Standards Development Organization	www.ihtsdo.org/snomed-ct
Devices	Global Medical Device Nomenclature (GMDN)	GMDN Maintenance Agency	www.gmdnagency.com/
	Universal Medical Device Nomenclature System (UMDNS)	WHO Collaborating Centre ECRI	www.ecri.org.uk/umdns.htm
Drugs	ATC/DDD Index	WHO Collaborating Centre for Drug Statistics Methodology	www.whocc.no/atc_ddd_index/
	MedDRA (Medical Dictionary for Regulatory Activities)	International Conference on Harmonization (ICH)	www.meddra.org/
	WHO Drug Dictionary	WHO	www.umc-products.com/DynPage.aspx?id=73588&mn1=1107&mn2=1139
Adverse Reactions	WHO-ART	WHO, maintained by the Uppsala Monitoring Centre	www.umc-products.com/DynPage.aspx?id=73589&mn1=1107&mn2=1664
	EU SPC ADR database	EMA	www.imi-protect.eu/methodsRep.shtml
	MedDRA (Medical Dictionary for Regulatory Activities)	International Conference on Harmonization (ICH)	www.meddra.org/
Disability	ICF	WHO	www.who.int/classifications/icf/en/
External Causes of Injury	ICECI	WHO	www.who.int/classifications/icd/adaptations/iceci/en/
Primary care	ICPC-2	WHO	www.who.int/classifications/icd/adaptations/icpc2/en/
Procedures	ICD-10-PCS ICD-9-CM Vol. 3	WHO	www.who.int/classifications/icd/en
Health Interventions	ICHI	WHO	www.who.int/classifications/ichi/en/
Medical Laboratory Observations	LOINC	Regenstrief Institute	loinc.org/
Genes, genetic disorders and traits	Online Mendelian Inheritance in Man (OMIM)	McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD)	www.omim.org/
Genes	HGNC	Human Genome Organization (HUGO)	www.genenames.org/about/overview

A registry team should also look for the existing data elements, or even further, common datasets. Before deciding on data elements, a registry team should make an overview of the current state of the art on the domain that the registry covers, and try to identify the already developed data elements and datasets that could be reused in their case. Reusing the commonly used and accepted data elements, in addition to the above mentioned advantages, could also mean a saving in the effort that is needed for the development of new data elements. However, it should be noted that when the existing data element is not relevant or is too constraining for the needs of the registry, it should not be used (6).

In recent years, important steps have been made towards harmonisation between registries and other data sources, when various organisations and projects started developing common datasets for their own domains. Here, it is certainly worth mentioning the epSOS project, which has done important work in the field of sharing information about the patient. Its so-called Patient Summary dataset, which aims to support safe, high-quality cross-border care for emergency or unplanned care events, consists of approximately 70 variables and comprises patient administrative data and patient clinical data. Thus, in case of data that describe the patient, it is recommended that a registry reuse these data elements. Similarly, for example, registries from the field of rare diseases should check and reuse data elements that were developed by the EPIRARE project, arthroplasty registries the EFORT EAR's datasets, etc., as long as they are relevant to their purpose, of course.

Table 6.6 shows a non-exhaustive list of common datasets that exist in the EU health domain.

Table 6.6: Existing common datasets

Area	Author	Common dataset	Link to the dataset
Patient summary	epSOS	epSOS Patient Summary	ec.europa.eu/health/ehealth/docs/guidelines_patient_summary_en.pdf
Rare diseases	EPIRARE	EPIRARE common data set	www.epirare.eu/_down/del/D9.3_ProposalforCDE_FINAL.pdf
Arthroplasty	EFORT-EAR	EFORT EAR Minimal datasets	www.ear.efort.org/
Cancer	ENCR	ENCR Standard dataset	www.encl.eu/images/docs/recommendations/recommendations.pdf
Cardiology	CARDS	CARDS Data Standards	www.escardio.org/Policy/Pages/data-standard-cards.aspx
Diabetes	B.I.R.O.	BIRO Common Dataset Summary	www.biro-project.eu/documents/downloads/D3_1_Common_Dataset_v1_7.pdf
Neuromuscular disorder	TREAT-NMD	FSHD Core Dataset	www.treat-nmd.eu/downloads/file/registries_toolkit/FSH_core_dataset.pdf
Multiple Sclerosis	EUREMS	EUREMS Core data set	eurems.eu/attachments/article/93/EUREMS%20Data%20Mask_August2014.pdf
Prescriptions	eHealth Network	ePrescriptions Dataset	ec.europa.eu/health/ehealth/docs/eprescription_guidelines_en.pdf

Table 6.7: Example of the epSOS Patient Summary dataset

PATIENT ADMINISTRATIVE DATA				
Variable (nesting level1)	Variables (nesting level 2)	Variables (nesting level 3)	DEFINITION AND COMMENTS	BASIC (Basic)/ EXTENDED (Ext) DATASET
Identification	National healthcare patient ID	National healthcare patient ID	Country ID, unique to the patient in that country. Example: ID for United Kingdom patient	Basic
Personal information	Full name	Given name	The first name of the patient (example: John). This field can contain more than one element.	Basic
		Family name/surname	This field can contain more than one element. Example: Español Smith Note: some countries require surnames to be the birth name (to avoid potential problems with married women's surnames).	Basic
	Date of birth	Date of birth	This field may contain only the year if the day and month are not available, e.g. 01/01/2009	Basic
	Gender	Gender code	This field must contain a recognized valid value	Basic
Contact information	Address	Street	Example: Oxford Street	Ext
		House number	Example: 221	Ext
		City	Example: London	Ext
		Post code	Example: W1W 8LG	Ext
		State or province	Example: London	Ext
		Country	Example: UK	Ext
	Telephone no.	Telephone no.	Example: +45 20 7025 6161	Ext
	e-mail	e-mail	Example: jens@hotmail.com	Ext
	Preferred HP/HPO to contact	Name of the HP/HPO	Name of the HP/HPO that has been treating the patient. If this is an HP, the structure of the name will be the same as described in 'Full name' (given name, family name/surname).	Basic
		Telephone no.	Example: +45 20 7025 6161	Basic
		e-mail	Email of the HP/legal organization	Basic
	Contact person/legal guardian (if available)	Role of that person	Legal guardian or contact person	Ext
		Given name	The first name of the contact person/guardian (example: Peter). This field can contain more than one element. Example: Español Smith	Ext
		Family name/surname	This field can contain more than one element. Example: Español Smith	Ext
		Telephone no.	Example: +45 20 7025 6161	Ext
		e-mail	e-mail of the contact person/legal guardian	Ext
Insurance information	Insurance number	Insurance number	Example: QQ 12 34 56 A	Ext

References

1. Australian Commission on Safety and Quality in Health Care. Operating Principles and Technical Standards for Australian Clinical Quality Registries. 2008.
2. Gliklich RE, Dreyer NA, eds. Registries for evaluating patient outcomes: A User's Guide. 3rd ed.
3. ISPOR: Taxonomy of patient registries: classification, characteristics and terms.
4. Rare diseases task force. Patient registries in the field of rare diseases. 2011.
5. WHO. Planning and developing population-based cancer registration in low- and middle-income settings. International Agency for Research on Cancer, 2014
6. Australian Institute of Health and Welfare (AIHW) 2007. A guide to data development. AIHW Cat. no. HWI 94. Canberra: AIHW. Available from:
<http://www.aihw.gov.au/WorkArea/DownloadAsset.aspx?id=6442458038>
7. Health Information and Quality Authority (HIQA). Guiding Principles for National Health and Social Care Data Collections. 2013.
8. EPIRARE. Deliverable D4: Guidelines for data sources and quality for RD Registries in Europe. 2014.
9. National Health Information Management Group (NHIMG). Minimum Guidelines for Health Registers for Statistical and Research Purposes. 2001
10. ISO/IEC 11179-4: 2004 (E). Formulation of data definitions. Available from:
<http://standards.iso.org/ittf/licence.html>.
11. ISO/IEC 11179-3:2013(E). Information technology — Metadata registries (MDR) — Part 3: Registry metamodel and basic attributes. Available from:
<http://standards.iso.org/ittf/licence.html>

6.4 Data sources for registries

Metka Zaletel

Different data sources could be used in the process of registry set-up and maintenance. It is necessary to determine which of already existing sources could be used or whether it is really necessary to establish a new data collection mechanism. The impact of legal issues, data quality issues, the importance of unit identifier, are among the top priority topics one should care about.

Key principles:

- Analyze the potential data sources taking into account the existing data sources.
- It is necessary to review the legal background that regulates the possibilities of merging and re-using of existing data sources.
- When using secondary data sources, the whole team shall take into account the different initial purpose of the secondary data source. Collaboration of the whole registry team is essential when analyzing the potential re-use of existing data sources due to possible drawbacks.
- When using secondary data sources, the critical analyses of data quality should be done. At the same time, the primary data sources bring different quality issues and challenges.
- The impact of using and merging/linking of different data sources should be carefully studied and, later on, presented to the potential users of final data.

6.4.1 Definition of Primary and Secondary Data Sources

The definition of primary and secondary sources is not strictly connected to patient registries and could be commonly used also for other research or statistical purposes.

The short definition of the **primary data source** explains it as data collected from the individuals to create (or supplement) the patient registry. Individuals or data providers in these cases could be either patients or clinicians, caregivers, pharmacists or other persons involved in health care. When the registry is completely or partially built on primary data sources, these sources are collected for the direct needs of the registry. When the primary data source is the only data source for a particular registry, the inclusion of the identifier is not necessary, but desirable also for the cases when the registry serves as the secondary source for another registry.

Secondary data sources are sources that were established or collected previously for other purposes. Examples of these sources are EHRs, medical charts, different databases (e.g. hospital administration database, census database). If the registry is built on two or more data sources, the identifier (e.g. personal identification number or some other unique identifier) must be included in all of the used data sources to enable merging the sources.

Among secondary data sources one should emphasize the importance of “non”-health data sources, especially statistical data sources. The latter can serve as an important source of socio-demographic or socio-economic variables and therefore enable us to decrease the burden on patients or health care workers.

It is worth emphasizing that the primary data source for one registry can act later on as a secondary data source for another registry.

There are many pros and cons when using primary or secondary sources.

Primary data sources are in most cases costly and time consuming, but on the other hand can provide data of higher quality with 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.

Secondary data sources are on the other hand less costly; they are easier to gather – providing that there is a sufficient legal background. There are the following considerations concerning usage of the secondary data sources:

- In most of the cases the secondary data sources are collected for other purposes (e.g. collected for insurance fund analyses, but later on used to perform different health care analyses)
- Data from secondary sources are usually used either by transfer into the registry or they are linked to other data sources to create a new, larger dataset for analysis.

Emerging challenges:

- Whenever two or more data sources are merged, the identifier of high quality is requested.
- Analyses of the data sources are quite often limited due to a different initial purpose of primary data collection.

- If two or more data sources are merged, each of them carries its own level of quality and influences the final quality level.
- The sufficient legal background should be taken into account when data are merged or linked from different institutions.
- Confounders in each dataset can compromise reliability of conclusions.
- Secondary data source can increase the analytical potential and the quality of conclusions by additional and in depth information concerning the populations covered; certainly, under condition that sufficient granularity of data is provided and evaluation processes follow transparent procedures.
- Secondary data can be used for validation and censoring processes in efficient standard processes, links have to be provided in regular processes.

6.4.2 Identification of Available Sources

When planning a new registry, all possible and available sources should be analysed. In most of the EU countries, there are legal acts defining registries, their ways of data collections, data providers and possible users. Based on these acts, the future registry holder can observe some of the potential data sources.

As mentioned earlier, when decision on usage of secondary sources is made, the quality and reliability of the source should be explored; on the other hand, the usability of the source for the “new” purpose should not be neglected. It is important to say that level of quality and reliability of one particular registry could be close to perfection for its own purpose, but could be quite unusable for some other purposes.

The most important data sources that the future registry holder should examine and analyse:

1. Primary data sources:

- a. Patient reported data are – as described above – usually resource-consuming in the sense of data collection, coding, keying, validating. There are many pros when using these data, especially gathering information not covered elsewhere, like opinions, life style, herbal supplements, etc.
- b. Clinician reported data could offer much more information since it is collected directly from the source with more clinical context. On the other hand, this could be burdensome for the data user since it is necessary to code the data or perform contextual analyses. Again, it is a resource-consuming task.

2. Secondary data sources:

- a. EHRs are information on routine medical care and practice. Usually they are structured, and the information is coded according to different classification systems. Since these data were prepared for the patients’ treatments, they can be very extensive, sometimes even in image formats, and the historical data could be hard to retrieve. But nevertheless, EHRs are the most valuable secondary data source for the patient registries.
- b. Human resource and financial databases could be used in some of the registries when the main purpose is evaluation of the staff or financial resources usage. The most important challenges are classifications since in many of these databases classifications are adapted to other financial systems and are hardly used for statistical or analytical purposes.
- c. Population databases or registries (like Census database, Central Registry of Population, National Patient Registry, etc.) are in most cases valuable sources of socio-demographic

or socio-economic variables. The user should bear in mind that these sources were prepared either for administrative purposes (like a Central Registry of Population) or for statistical purposes for the field of demographic statistics.

- d. Other health registries are maybe the most important data source among above mentioned sources. Again, these registries were probably set up for different reasons and purposes which should be explored in quite some detail. On the other hand, we should reduce the burden on reporting units (e.g. clinicians, hospitals). If there exists data reporting on particular issue (e.g. hospital admission), these data should be re-used as much as possible whenever the legal framework allows it.

As written in the previous subchapter, if one wants to use one of the above mentioned secondary sources, a patient identifier is necessary. For statistical purposes, matching without identifier is also possible with certain probabilistic methods, but not recommended in the process of building the registry. Therefore, the prerequisites for using these sources are:

- Unique patient identifier which is used in all sources that are going to be used
- Documentation explaining what is really the content of the source (e.g. target population, metadata on variables...)
- Adequate level of quality for selected variables.

At the end, it is worth mentioning that every day new registries are born. The registry holder should take this into account by monitoring the new registries (appearance and quality). Based on the new registries, the registry design could be changed if the new sources are of high quality, available, and their inclusion would bring a burden decrease.

References

1. Wallgren A., Wallgren, B.: Register-based Statistics – Administrative Data for Statistical Purposes, Wiley, 2007
2. Polygenis D, ed. *ISPOR Taxonomy of Patient Registries: Classification, Characteristics and Terms*. Lawrenceville, NJ; 2013.
3. Health Information and Quality Authority (HIQA). Guiding Principles for National Health and Social Care Data Collections. 2013
4. Gliklich RE, Dreyer NA, eds. Registries for evaluating patient outcomes: A User's Guide. 3rd ed. 2014.

6.5 The role of information system methodologies and techniques in the phase of patient registry creation

Vesna Lešnik Štefotič, Živa Rant, Ivan Drvarič

In the phase of development (design) of the patient registry, it is recommended to use models - an abstraction of an existing real world system. With different types of models, we can explore a patient registry from many perspectives (process, data etc.). To prepare proper models information system methodologies and techniques can be of a great value.

Key principles:

- a model helps us understand real world systems and is designed to see only the most important issues by preventing us from getting distracted by all the details, not important at this time
- it is very useful to involve information system (IS) experts - persons with knowledge and experience in IS methodologies and techniques as early as possible in the development of the patient registry
- IS experts are there only to facilitate the process of defining the PR and to provide guidance on how to accomplish these with different IS techniques, health domain experts (usually clinicians) are those who define the content
- communication across all team members and especially between health domain experts and IS experts is the key issue when modelling PR
- there are a lot of IS models and techniques which can be used in PR design – for example:
 - UML (Unified Modelling Language) with use case diagram
 - BPM (Business process modelling) to model PR processes using BPMN (Business process modelling notation) or EPC (Event-driven process chain)
 - data modelling using E-R diagram (Entity – Relationship diagram)
 - knowledge management using OpenEHR

Patient registries can be **computerized to a different degree**. Some registries use IT tools only for processing, analysing and representing the data, some also for gathering information in an electronic way directly from the information source (patient, clinicians etc.) or indirectly from other information systems (IS) such as electronic health records (EHR's).

Regardless of the degree of computerization of the patient registry it is **very useful to take advantage of information system development methodologies, techniques and tools in the phase of development (design) of the patient registry content and functions**.

In this subchapter the following will be 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 and
- the importance of standard terminologies and code lists.

The main purpose of the following text is to briefly introduce some of the most 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 common used modelling notations⁵⁵ and terminology to be able to read a model;
- understand the role of IS expert (or other person with experience in IS modelling techniques) in the patient registry creation;
- understand the importance of using standardized terminologies and code lists if they exist.

For more information on this subject and described techniques the reader is encouraged to explore the provided links to free tutorials and additional reading.

6.5.1 Why modelling?

“A picture is worth a thousand words.”

A **model** is usually a **human construct** to help us **understand real world systems**. When we are **modelling** then we **construct an abstraction** of an existing real world system (or of the system we are envisioning). Modelling help us to see only the most important issues by preventing us from getting distracted by all the details which are not important at this time.

For example a world map is a model of a world. If the search is only for continents, than the country borders on the map are not needed. But when searching for the number of countries per continent, than a world map is needed with a greater detail, including country borders.

Real world issues can be seen from different perspectives using different modelling techniques and different standardized way of presenting it (=notation). For example collecting data for a patient

⁵⁵ Notation = standardized way of presenting models (usually real world issues, like processes, things etc.)

registry (PR) can be explored from the process point of view (how the process is performed, which tasks are executed, who is participating in the tasks, what the inputs and outputs are) with **process modelling techniques** such as **business process modelling** and a **process model** of collecting data can be prepared using **business process management notation**. Then collecting data for a PR can be further explored only from the data perspective (what data elements are collected, how the data elements relate to each other etc.) with **data modelling techniques** such as **entity- relationship modelling** and a **data model** can be prepared using **entity-relationship diagram**.

Models are an excellent tool to communicate with others. The prerequisite is that all the participants understand at least the basic notation standard of the presented model.

6.5.2 The role of IS expert (system analyst, process modeller, health informatics expert etc.)

For useful application of IS methodologies and techniques in PR creation **it is recommended to involve persons with knowledge and experience in IS methodologies and techniques** such as system analysts and/or business process modellers or other persons with the knowledge in this domain **as early as possible in the development of the patient registry** (see [6.1.10.1](#) 'Human Resources').

It is very important to clearly understand **the role of IS experts** in the process of patient registry creation. They are only there **to facilitate the process of defining the right content and to provide guidance on** how to accomplish these important tasks with different IS techniques. **Health domain experts (usually clinicians) are those who define the content**, as they have the knowledge of the patient registry domain. IS experts cannot and should not define on their own the scope, content, outcomes, etc. of the patient registries. They are only facilitators of the PR creation process and responsible for proper modelling.

Communication across all team members and especially between health domain experts and IS experts is the key issue when modelling the PR. As already stated, the IS expert is responsible for proper modelling and to be able to do so it is crucial to gather the right information from the right people. The commonest way of gathering information is to conduct **guided interviews** with health domain experts; another option is to have an **interactive modelling workshop**, where the model is prepared during the session. In both cases it is very important to **properly manage the process of information gathering** from preparation, execution to post execution phase.

In the following two subchapters some tips will be given on how to conduct a guided interview and what an interactive modelling workshop is and how to execute it.

6.5.2.1 Guided interview

A guided interview will provide an IS expert with a wealth of information. It is usually divided into three phases: preparing the interview, conducting the interview and the post interview phase.

A. In the preparation phase – project team (or analyst) should:

- **Define the purpose and objectives** of the interview – why the interview is being conducted , what is its objective
- **select the right people** to be interviewed

- **prepare a set of questions** which will guide the interview; it is recommended also to prepare short checklists not to miss any important information
- **arrange a venue**
 - location
 - date/time
 - equipment
- **send out invitations; explain to the participants in advance** the purpose/objectives of the interview, what kind of input is expected from them (documents, examples of reports, work instructions, etc.)

B. Execution phase – conducting of interview shall begin with an opening statement of purpose for the interview. This purpose statement is to ensure that the interview has a clear overall goal. It should also be used during the interview to ensure the interview stays on topic.

In the beginning it is recommended to use general and open ended questions. To clarify a particular issue closed questions are recommended. **Active listening** is very important. Some rules for being a good listener:

- focus on the speaker,
- be aware of non-verbal signs
- respond (verbally or non-verbally) to the speaker to encourage him or her to continue
 - using encouraging words and body language (head- nodding, smile etc.)
 - repeating the received sentence in your own words
 - reflecting your understanding of their position
 - asking questions to clarify his/her message.
 - summarizing etc.

During an interview it is very important to **provide feedback** to the speaker about his/her message. Provide feedback if you have not understood the message with additional questions to clarify what was meant. Repeat the sentence in your own words (paraphrasing) to show the speaker that you are an active listener and, at the same time, potential mistakes in the given message can be corrected. It is strongly recommended to summarize periodically the speaker's messages.

At the end of the interview, the participants should be thanked and asked to review the results (written document and graphical models) of the interview at the next session if necessary.

C. Post interview phase

After the interview is completed, the project team should review the information gathered and prepare a written document and prepare a model. Both should be reviewed together with the interviewees.

6.5.2.2 Interactive modelling workshop

Interactive modelling workshop is a guided interview combined with real time modelling. The result of such a workshop is a model, confirmed by consensus, with resolved ambiguities.

A modeling workshop is usually led by a system analyst/designer. He/she must involve all the other participants in the discussion. Usually a whiteboard is used where models are drawn and notes taken. If it is possible, the addressed issue is also in real-time, modeled with some dedicated software and presented to the participants with an overhead projector.

The preparation and execution phase of the process design workshop are similar to that of a guided interview and all above mentioned recommendations should be followed.

6.5.3 Short description of frequently used modelling techniques and notations

There are a lot of different methodologies and techniques in IS development which can be useful also in the process of patient registry creation, but it is not possible to list and describe all of them. So in the next subchapters only some **useful IS methodologies and techniques** will be presented which can be applied in a patient registry design. **Unified Modelling Language (UML)** will be explored as a set of different modelling techniques notations used typically in software development and some of them (for example use, case diagramming) are very useful also in the creation of a PR. Then **process management techniques** will be explored which will help to **model processes** to fully understand tasks, roles, inputs and outputs, drivers/events of different processes related to patient registry design, patient registry execution and patient registry improvements. The modelling techniques presented next will deal with data or in broader view knowledge modelling. In these subchapters “classical” **data modelling** will be explored using **entity - relationship modelling techniques** (part of a single model approach) and also dual level approach to **knowledge modelling using archetypes**. A very important part of data modelling (especially in the context of semantic interoperability) is represented also by **terminologies and code lists**. Therefore a separate subchapter will explain why it is recommended to use internationally approved terminologies and code lists when collecting health data.

The main purpose of this subchapter is to emphasize how useful engineering methodologies and techniques of IS design are in the patient registry creation and to show some examples of how some of them can be used in this process. It is expected that this subchapter will evolve/grow in the future and more examples of IS methodologies and techniques will be added.

The techniques are presented in random order. Where they can be applied in PR creation is described in the subchapters.

6.5.4 UML

Unified Modelling Language (UML) is a standardized (visual) modelling language consisting of an integrated set of diagrams. It was developed by Jim Rumbaugh, Grady Booch and Ivar Jacobson in 1994 to help system and software developers accomplish the following tasks: specification, visualization, architecture design, construction, simulation and testing, documentation. Today UML is adopted by the Object Management Group (OMG) a consortium of over 800 companies dedicated to developing vendor-independent specifications for the software industry (1).

UML is methodology independent, this means that the process of gathering requirements, analysing and modelling them is not formally defined; only diagramming notations of different diagrams are prescribed. UML 2.0 defines **thirteen types of diagrams**, divided into three categories (2):

- **Structure Diagrams** include the Class Diagram, Object Diagram, Component Diagram, Composite Structure Diagram, Package Diagram, and Deployment Diagram.
- **Behaviour Diagrams** include the Use Case Diagram (used by some methodologies during requirements gathering); Activity Diagram, and State Machine Diagram.
- **Interaction Diagrams**, all derived from the more general Behaviour Diagram, include the Sequence Diagram, Communication Diagram, Timing Diagram, and Interaction Overview Diagram.

For the purpose of PR creation (and due to the limited space) only **Use case diagram will be presented**. It is a very valuable tool to define the user requirements (goals) of the PR. It must be stressed that other types of diagrams can also be used when creating a PR and especially when developing software for a PR (PR set-up).

Additional information on UML can be found on the official UML web page (www.uml.org).

6.5.4.1 Use case diagram

Use case analysis is a major technique used to find out the **functional requirements of a system**. **Use case**, an important concept in use case analysis, represents **an objective a user wants to achieve** with a system. It can be in text form, or be visualized in a **use case diagram**.

A use case describes a system's actions from an external point of view (user's point of view). Use cases are named with a verb or a verb and noun phrase for example "Make data quality check".

A use case diagram provides a graphical overview of goals (represented by use cases) users (represented by actors) want to achieve by using the system (represented by a system boundary but is often "opt out" in diagram). Use cases in a use case diagram can be organized and arranged according to their relevance, level of abstraction and impacts to users. They can be connected to show their dependency, inclusion and extension relationships.

A UML use case diagram is mainly formed by **actors**, **use cases** and **associations** (connectors); sometimes also by system boundaries.

An actor is any **person** (also organisational unit) or **external system** (machines, IT system, sensors) that interacts with the system in achieving a user goal. It is drawn as a named stick figure.

Questions to find all relevant actors of a use case are:

- Who are the system's primary users?
- Who requires system support for daily tasks?
- Who are the system's secondary users?
- What hardware does the system handle?
- Which other (if any) systems interact with the system in question?
- Do any entities interacting with the system perform multiple roles as actors?
- Which other entities (human or otherwise) might have an interest in the system's output?

A use case is a **function of a system**. It is named by verb and drawn as an ellipse.

Associations are connections between actors and use cases; drawn by a line (sometimes with arrow).

System boundaries are drawn as a rectangle across use cases performed by a system.

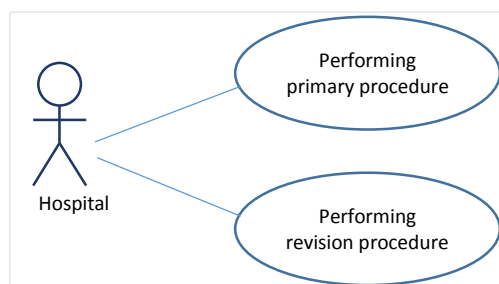


Figure 6.3: Example of use case diagram (»Hospital« as Actor, »Performing primary procedure« and »Performing revision procedures« as Use cases)

First it is usual to draw a **top level use case diagram** or a **context diagram**. It is a special kind of use case diagram, where the individual use cases are hidden and represented by the system of interest interacting with all the actors. It is very useful to define the context (environment of a system) in this case of a PR.

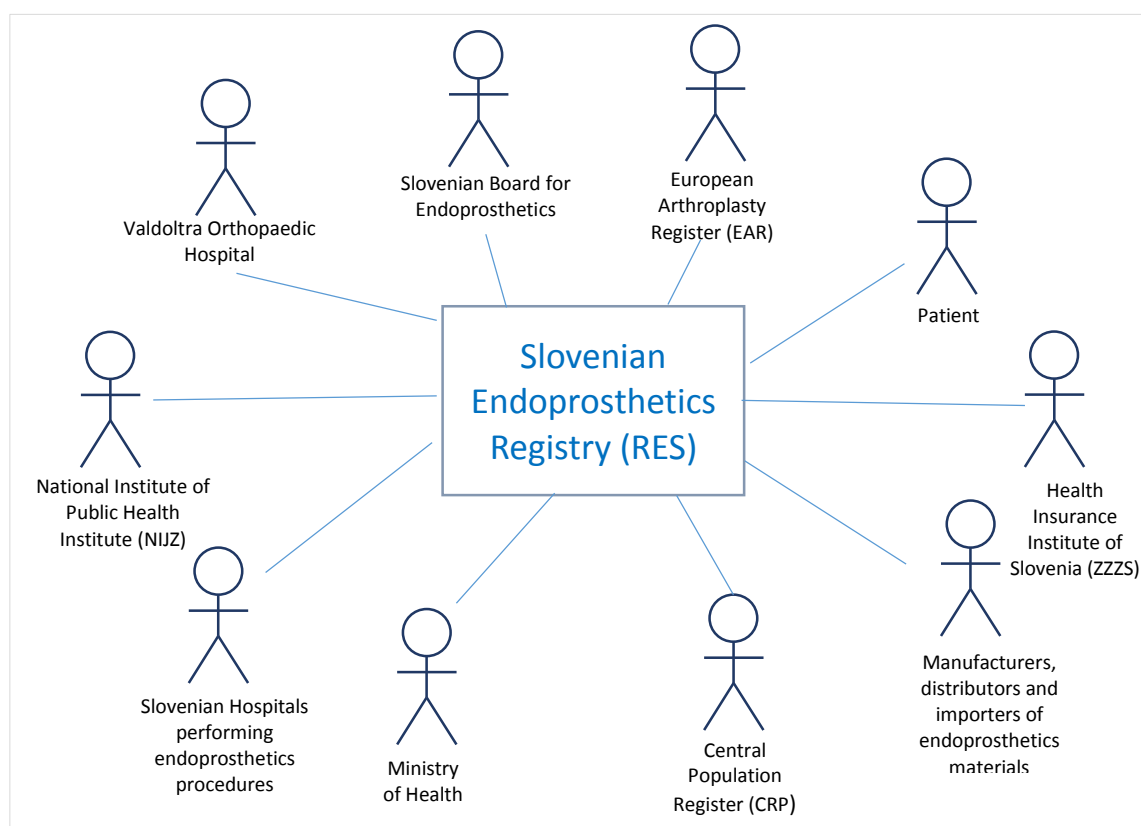


Figure 6.4: Context diagram for The National Arthroplasty registry of Slovenia (RES)

Example of using Use Case diagrams in Patient registries:

- TNPCR–AERRO Central Cancer Registry Business Use Case Diagram
www.cdc.gov/cancer/npcr/informatics/aerro2/hospitals/h_business_use_case_diagram.htm
- NPCR–AERRO Central Cancer Registry Operations Use Case Diagram
www.cdc.gov/cancer/npcr/informatics/aerro2/pdf/c_operations_use_case_diagram.pdf

A useful tutorial on creating use cases can be seen on Pace University, available at:
csis.pace.edu/~marchese/CS389/L9/Use%20Case%20Diagrams.pdf

6.5.5 Process Modelling

Process model design is an engineering technique and part of **Business process management (BPM)**, a disciplined approach to business processes. To fully understand the purpose and the ability to use process management in patient registries creation it is necessary to quickly introduce Business process management as a whole and then **Business process modelling** as an important step to model existing or future processes using **Business process modelling notation (BPMN)**.

6.5.5.1 Introduction to Business process management

In the 1990's the focus on **processes** instead of functions was introduced in organisations (3). Michael Hammer was the originator of reengineering and the process enterprise changed forever how businesses do business (5).

Typically, organizations (also healthcare organizations) are structured into divisions and departments based on the functionality of each division and department (for example, IT department). Each division or department performs its own specific tasks and determines its own competency. Therefore, this organization structure tends to create 'Silo Thinking'; each department stands alone with less or no interaction with other departments within the same organization. Compared to Silo thinking, **processes**, on the other hand, **cut across** these **functional silos**. Where different activities in a process require different skills, the process is likely to involve a number of people and departments.

Business process thinking in organizations introduced new methodologies focusing on business processes such as **Business process management (BPM)**.

BPM is defined by Association of Business Process Management Professionals as "a disciplined approach to identify, design, execute, document, monitor, control, and measure both automated and non-automated **business processes** to achieve consistent, targeted results consistent with an organization's strategic goals. BPM involves the deliberate, collaborative and increasingly technology-aided definition, improvement, innovation, and management of end-to-end business processes that drive business results, create value, and enable an organization to meet its business objectives with more agility." (4)

Business process is a sequence of tasks / activities which transform inputs to outputs, that is of value to the customer, performed by human or machine (for example purchase process).

Business Process Management defines **process lifecycle** in eight steps:

1. Identify the process
2. Model the process
3. Discuss, audit, review the process
4. Automate the process⁵⁶
5. Implement the process
6. Track (measure) the process
7. Optimize the process
8. Dismiss the process

⁵⁶ This step is not mandatory.

From the above mentioned steps it can be seen that Business process modelling is an important part of Business process management.

Further reading:

- Hammer, Champy: Reengineering the Corporation: A Manifesto for Business Revolution (5),
- Keen, Knapp: Every Manager's Guide to Business Processes (6)
- Keen: The Process Edge, Creating Value Where It Counts (7)

6.5.5.2 Introduction to Business process modelling

Process modelling is a widely-used engineering approach to determine and **describe existing processes and future process scenarios**. We can say that process modelling comprises all the activities that need to be undertaken to document a process.

Process modelling considers the following elements:

- Process scope
- Process start
- Process activities and their interconnections
- Process numbers (measurable figures of its activities: e.g. duration, number of staff, maximum load, etc.)
- Process KPI (key performance indicators)
- Process end
- Process connections with other processes

All these elements need to be documented before starting implementation (and automation).

For business process modelling there exist a lot of modelling techniques. Graphical/visual representation of processes with process diagrams is a common way to describe a process. There are many diagramming techniques and notations to model processes, but the ones mostly used are **event based process chain (EPC)** – process modelling language invented by prof. Scheer and colleagues at the University of Saarland in 1992 (8) and **Business process management notation (BPMN)** from OMG.

Business process management notation (BPMN) is a de-facto standard notation widely used in the BPM community (9). In the following subchapters the point where business process modelling could be useful in patient registries development will be presented and EPC and BPMN will be introduced.

6.5.5.3 Business process modelling and patient registries

Patient registries can benefit from BPM:

A. in phase of planning PR

- **to better understand current state** (see chapter [6.1.3.1](#) 'Overview of Current State'; chapter [6.1.6](#) 'Defining the Scope of the Registry & Building a Registry Development Team'; chapter [6.1.7](#) 'Performing Stakeholder Engagement and Analysis') we can model the current processes
- **to identify all data sources** for Patient registry (chapter [6.4](#) 'Data sources for registries')
- **to model new process for PR operation** (collecting data, processing and analysing data, reporting data)
- **to model PR supporting processes** such as for example *Perform annual audit*

B. in phase of PR set-up

- **to model processes to be automated** (chapter 7 'Patient registry information system development and implementation') – to gather user requirements

C. in phase of running a PR

- **to improve processes** (chapter 8.2 'Overarching Processes')

6.5.5.4 Event-driven process chain (EPC)

"Event-driven process chain (EPC)" is the main ARIS model for representing processes. It is a dynamic model bringing together the static resources of the business (systems, organization, data, etc.) and organizing them to deliver a sequence of tasks or activities ('the process') that adds business value (10).

An event "activates" an activity and activity will always "create" one or more new events (Figure 6.5: Example of the EPC diagram).

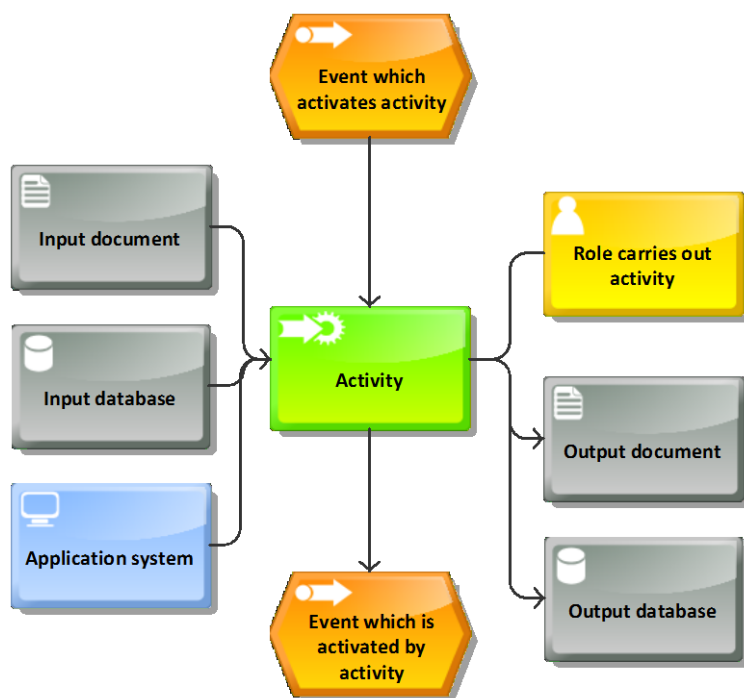


Figure 6.5: Example of the EPC diagram

6.5.5.4.1 The EPC Objects

Essentially, there are four types of objects used in the EPC:

- Events,
- Functions,
- Rules,
- Resources (data, organisation, system, etc).



Event represents the changing state of the world as a process proceeds:

- External changes that trigger the start of the process
- Internal changes of state as the process proceeds
- The final outcome of the process that has an external effect

To describe events, we typically use the convention 'noun-verb'.

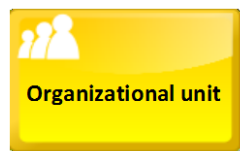


Functions/activities represent the activities or task carried out as part of a business process; ideally with each one adding some value to the business. Function may be carried by people or by IT systems. They have inputs (information or material), create outputs (different information or a product) and may consume resources.

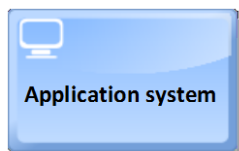
To describe function/activity we use the convention 'verb-noun' or more specifically.



Rules: Real processes do not just consist of sequential steps. The need to cope with parallel paths, decisions, multiple triggers and complex flows is the reason that modelling tools are used to represent processes. To model a process flow 'Rules' are added to the functions and events previously described. There are three basic types of rule: OR, XOR and AND.



Organizational objects represent the people who perform the process tasks represented by functions/activities. Specific people, departments, roles or teams can be represented, depending on the context and detail of the model.



Application Systems resources represent the computer and software applications used to support the business.



Activities have inputs (information or material) and create outputs (different information or a product). Usually they are data (in a database) or documents.

Further reading:

- ARIS online Academy
http://cdn.ariscommunity.com/aris_online_academy/what_is_bpm3/50bfqndn/player.html
- ARIS Architect & ARIS Designer
www.softwareag.com/corporate/images/SAG_ARIS_BusArchDesign_FS_4PG_Jan14-Web_tcm16-78556.pdf

6.5.5.5 Business process management notation (BPMN)

Introduction

Business process management notation (BPMN) principles originate from flowcharts, which were invented in 1946 by Goldstine and von Neumann. In the following years, many similar modelling techniques appeared. BPMN 1.0 was introduced in May 2004 by Business Process Management Initiative and was in 2005 acquired by Object Management Group (OMG). BPMN 2.0 was released in 2011 (11).

In the last ten years, BPMN has spread and has become the most widely used process modelling technique in the world, supported by the main IT companies. It is an open specification, therefore no royalty fees need to be paid. BPMN is a standard for graphical modelling and also for transformation to execution model (XML code).

Where BPMN can be used?

BPMN has many fields of application. It can be used:

- To capture the existing state of the process (AS-IS process);
- To gather requirements for the new information system (description of the system behaviour and interaction with the user);
- To optimize the process (describe the AS-IS and then TO BE process);
- To simulate behaviour using special software tools;
- To be automatically translated into execution mode (the program code), but it has to be defined in great detail.

6.5.5.5.1 BPMN Basics Elements

BPMN uses a set of graphical elements to define a **Business Process Diagram (BPD)**. BPD is a network of graphical objects which are activities and the flow controls that define their order of performance. Using BPMN the following can be described:

- **Processes and activities** (complex and atomic) - units of work (e.g. hip replacement, hospital admission, floor cleaning, inputting user data, sending quote)
- **Events- impulses, conditions or business rules** that start or interrupt the process or activity (e.g. receive document, Tuesday at 6:00, the heart rate is below 50, out of stock)
- **States**
 - **Activity states:** e.g. the 'preparing order' activity can be in the following states: idle, starting, running, finished, interrupted
 - **Object states:** e.g. the 'order' information object can be in the following states: prepared, draft, sent, deleted, archived
- **Decisions and conditions** that directly affect the flow of the process. Decisions can be used to split the process in two parallel or alternative branches or paths (e.g. is number of patients < 5, is blood pressure > 120/90?)
- **Artefacts** - all objects (inputs, outputs) that are used within the process (e.g. material, documents, information, user interfaces, reports, instructions, standards)
- **Roles, actors** - people, information systems or other organizations who perform the process activities (e.g. employees, nurse, surgeon, information systems, databases, external systems)

BPMN has a small set of notation categories so the reader of a BPD can easily recognize the basic type of elements. The four basic groups of elements are (8):

- Flow objects (activity, event, decision);
- Swimlanes (role, sub-role);
- Artefacts (documents, information) and
- Connecting Objects;

A. Flow objects are three core elements:

- **Activity** - represents performed work; we have two types of activity: **task** (atomic activity) and **sub-processes**; drawn as rounded corner rectangle with additions (+, II, loop)
- **Event** - represents what happens during the course of a business process, it affects the flow of the process; drawn as circle; we have 3 types of events:
 - **start event** (single circle), which is used at process start
 - **intermediate event** (double circle), which is used between the process start and end
 - **end event** (thick circle)
- **Gateway** – represents process decisions as well as the forking, merging and joining of paths; drawn as diamond shape, the centre of the shape represents the type of split; 2 types of gateway are distinguished:
 - (X) in the centre represents decision (choose only one process branch)
 - (+) in the centre represents parallel execution of all output branches

B. Swimlane in a diagramming technique is a mechanism to organize activities into separate visual categories. In BPMN swimlanes are representing **roles** (participants) in a process. A role is represented as a horizontal (or vertical) rectangle. Sub-roles or departments within the organization are represented as rectangles within another rectangle.











C. Documents and information are represented as a paper icon with a folded corner.

D. Process elements can be **connected** using:

- **Sequence flow** - this represents the logical sequence of activities, events and decisions (drawn as solid line with a solid arrowhead); connects activities, events and gateways
- **Information flow** - this element represents inputs/outputs of activities and triggers for message driven events. Usually, the document element is attached to the information flow. (drawn as dashed line with an open arrowhead)

Usually, sequence flows are drawn first; information flows are added later, in the process modelling phase.

Table 6.8: Basic elements of BPMN

Activity (Subprocess, Task)	 
Event (Start, Intermediate, End)	  
Gateway	
Documents	
Swimlane (Role)	
Sequence flow	
Information flow	

6.5.5.2 Additional BPMN elements

With the basic BPMN elements it is possible to model almost all real life scenarios. But there are additional types of elements:

- Specialized types of processes and tasks;
- Specialized types of events;
- Specialized types of gateways;
- Various information flow usage.

The various types of specialized processes, events, gateways and flow usage can be quickly viewed on Object Management Group - Business Process Model and Notation website (www.bpmn.org) under 'Quick Guide'. Informative presentation of BPMN can be seen also on Camunda.org tutorial (<http://camunda.org/bpmn/tutorial.html#tutorial>).

It is recommended to download A1 Poster BPMN2_0_Poster.pdf (freely available on www.bpmn.org under Documents/BPMN 2.0 Poster). It is a very useful quick reference guide for using BPMN 2.0. The normative document on BPMN 2.0, can be viewed on www.omg.org/spec/BPMN/2.0/PDF.

6.5.5.3 Steps for graphical modelling of identified business process using BPMN

For creating graphical representation of the business process using BPMN take the following steps (Figure 6.6):

1. Define roles
2. Define activities
3. Arrange activities in pools/lanes
4. Connect the activities using sequence flow
5. Add events
 - Start event (what triggers the process)
 - Intermediate events (pauses, exceptions)
6. Add documents, information and flows
 - Output activities, receiving events

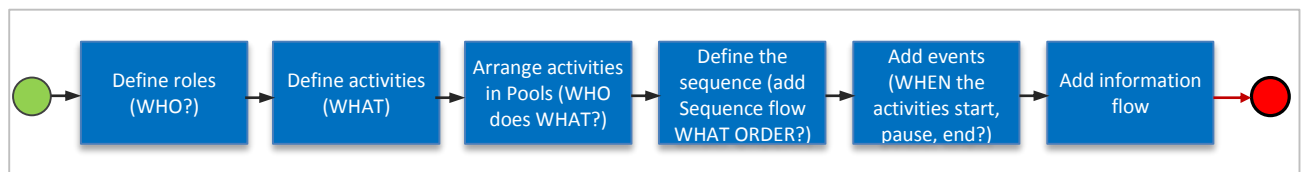


Figure 6.6: Steps for modelling graphical presentation of BPMN

6.5.5.4 Process model decomposition

Processes are usually not simple and their graphical models quickly become big, complex and unreadable. Therefore, they should be decomposed and not presented on a big A1 whiteboard. By modelling processes the rule of 7-10 should be applied, which means: maximum 10 activities per process model should be designed for maximum readability. If the process model contains events and gateways, this number should be lowered.

Two types of decomposition are recommended:

- **top down decomposition** and
- **role decomposition** or **pool focusing** (hiding other roles or pools).

Top-down or levelled process decomposition helps to reduce cluttering on the process models. Different levels are also targeted for different audiences. For example, **level 1** should be read, defined and managed by top level management (CEO, general manager, quality manager, etc.) The

second level should be targeted for process owners and performers, because it defines detailed activities, responsibilities, events, information flows etc. Usually 2 levels of detail are enough if the processes will be performed manually (it also depends on the complexity of processes). The **third level** usually contains technical details and it is targeted for developers, software architects, and performers.

Conceptual models (designed for business users) should not be mixed with technical process models, which include implementation details (designed for technical experts, software designers and programmers).

Pool focusing is the horizontal way of detailing processes. The main idea is to design details (activities, events, gateways, etc.) in ONE pool and to represent other pools as black-boxes (without any details).

6.5.5.6 Documenting business processes

Usually, the process diagram alone is not enough to fully present and completely describe a business process. There are some process properties like process scope, process goals, metrics of the process, etc., which cannot be represented graphically. Therefore additional explanations are needed in text documents.

Documenting the processes includes:

- graphical presentation, where the flow of a process is presented using some graphical notation, and
- textual presentation providing more detailed descriptions of the process and document templates.

The **Process Description Document** is a document which may be provided for each process defined. Its creation is not mandatory, but it is strongly suggested to do so in order to provide the reference information about the process. The content of the Process Description Document should define at least:

- aims of the process, objectives
- graphical representation in selected modelling notation (like BPMN), showing the sequence of activities, roles involved, documents used, etc.
- descriptions of activities providing more information about each activity defined in the graphical representation
- key performance indicators - defining how to evaluate if the work has been performed correctly and efficiently,
- references to other processes; information which processes provide inputs to the process and which use the outputs of the process.

6.5.5.7 BPM tools

There is a lot of BPM tools and platforms. BPM tools can be grouped by licence type into 3 groups:

1. Open source, freely accessible tools: Intalio, Bonita
2. Free, but not open source: ARIS Express
3. Commercial: Signavio BPM, Oryx BPM, Appian, RunMyProcess (Google Apps platform), ActiveVOS – Socrates, ARIS, MS Visio, Lombardi Teamworks, Pegasystems BPM Suite, etc.

6.5.6 Data modelling (using E-R diagram)

As was seen in the previous chapters there are many slightly different definitions of patient registries (for example, see definitions listed in AHRQ – Registries for Evaluating Patient Outcomes: A User's Guide (12), p.35), but almost all of them say that patient registries are “**an organized system for the collection, processing and storage of uniform health data on individual persons in a systematic way for specific and defined purpose.**”

According to Hernandez (13) any data collected in a systematic way and for a specific purpose regardless of the collection method (electronically, paper-based) can be called a database. From this definition it can be concluded that the main ingredients of the patient registries are uniform health data about individual persons organised in a **database** (DB).

What is a database?

“A database is a collection of related data.” (14)

"The database is a tool for efficient storage and manipulation of data." (15)

"A database is a collection of data that is used to model the organization or organizational process. It does not matter whether it is used for a computer program or it is on paper. As long as the data are collected and organized for a specific purpose, we have a database." (13)

What are data?

Data are a representation of facts, concepts and instructions presented in a formalized manner suitable for communication, interpretation, or processing by humans or by automatic means. (ANSI, ISO)

"Data are facts presented by the values (numbers, signs, symbols) that have meaning in a particular context." (15)

"The data are static values stored in the database." (13)

What is information?

"Information is quantified data in a specific situation." (14)

"Information is data that is processed in such a way that meets the needs of the individual." (13)

6.5.6.1 Types of databases

Databases can be roughly divided into 2 categories: **operational databases** and **analytical databases**. The first type – the **operational databases** – are used primarily in transactional systems (**OLTP - On-Line Transaction Processing**), which are mainly intended for daily collection, modification and maintenance of data. The data stored in these databases are dynamic, which means that they change frequently. Operational databases always show the current status. An example of such systems is, for example, an automated teller machine (ATM).

The second type – the **analytical databases** are primary used in analytical systems (OLAP – On-Line Analytical Processing). In an OLAP database there is aggregated, historical data, stored in multi-

dimensional schemas (usually star schema). This type of data are used for example in decision support systems, to analyse trends, etc.

OLTP and OLAP are complementing technologies. OLTP runs a business day by day and analytical databases usually use data from operational databases as a main source. Both types of databases meet the specific tasks of data processing and therefore their development requires a different data modelling approach.

In the following section only modelling the first type of databases will be explored.

6.5.6.2 Data Model (E-R diagram)

Data modelling is originally part of the software engineering discipline. The output of the data modelling is the **data model** – a representation of a real world situation about which data are to be collected and stored in a database. A data model depicts logical relationships among different data elements.

There are a lot of different techniques of data modelling but focus will be placed on the most widespread technique; an **Entity Relationship Diagram** (E-R diagram) will be used to demonstrate data structure.

The ER model was introduced by Peter Pin Shan Chen in 1976 as a conceptual modelling approach that views real world data as systems of **entities** and **relationships**.

With an ER diagram any system can be described but E-R diagrams are most often associated with modelling databases that are used in software engineering. In particular, E-R diagrams are frequently used during the design stage of a development process in order to identify different system elements and their relationships with each other. **In patient registry creation E-R modelling can be used to identify required data elements** (see chapter 6.3 'Registry dataset') **and structure it properly** (to place data elements in a prominent and logical position).

E-R diagrams are a very useful tool for data modelling and **visual presentation of data model**. They are **easy to understand** and do not require a person to undergo extensive training to be able to work with them efficiently and accurately. This means that they can be easily used in communication among team members, developers and end users, regardless of their IT proficiency. E-R diagrams are also readily translatable into relational tables which can be used to quickly build databases.

6.5.6.2.1 Elements of E-R diagram

An E-R diagram is a visual presentation of data with the following elements: entities, attributes and relationships.

An **entity** is a thing (material or nonmaterial) that is relevant to a given system and on which the system must store data. It has to be recognized as being capable of an independent existence and which can be uniquely identified. It may be a **physical object or subject** such as patient or medical device, **an event** such as medical appointment, **a concept** such as an order. For example, a patient registry may include entities: patients, diagnoses, interventions, outcomes, etc. Entities are represented in ER diagrams by a rectangle and named using singular nouns (e.g., Patient).

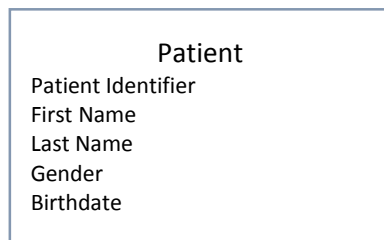


Figure 6.7: Representation of an entity Patient

An **attribute** is a property, trait, or characteristic of an entity or relationship. The attributes describe entity or relationship. Attributes are named using singular names (e.g. Patient Name) and are represented in original notation by oval shapes, but in many other notations as a list inside the entity rectangle.

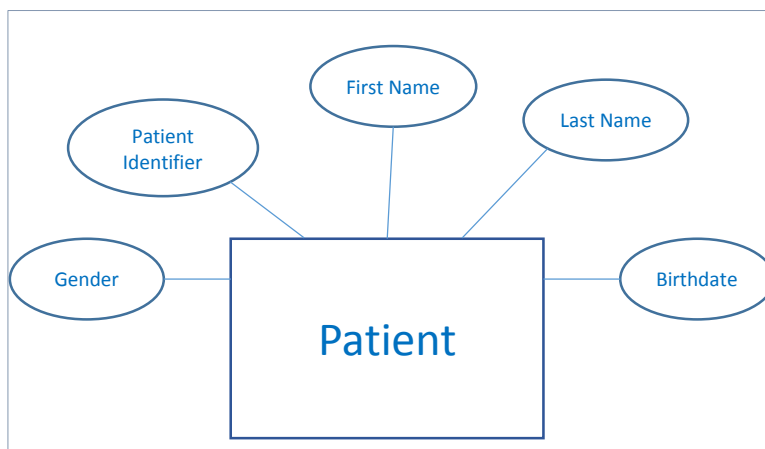


Figure 6.8: Representation of entity Patient with its attributes in original (Chen) notation

A **relationship** represents the interaction between the two entities. The phrase “The patient has a diagnosis” tells us that there is a relationship between the entities Patient and Diagnosis. Usually, the relationships are binary (between two entities), but they can also be ternary (three entities), etc. Specific types of relationships are recursive relationships (entity is in relationship with itself—also called self-referential relationship). Typical of such a relationship is: “Each employee can be a leader and has a leader.”

Relationships are in original notation represented by diamond shapes and are labelled using verbs. Usually they can be read in both directions, for example:

- “The Patient has a diagnosis.”
- “The Diagnosis is assigned to the Patient.”

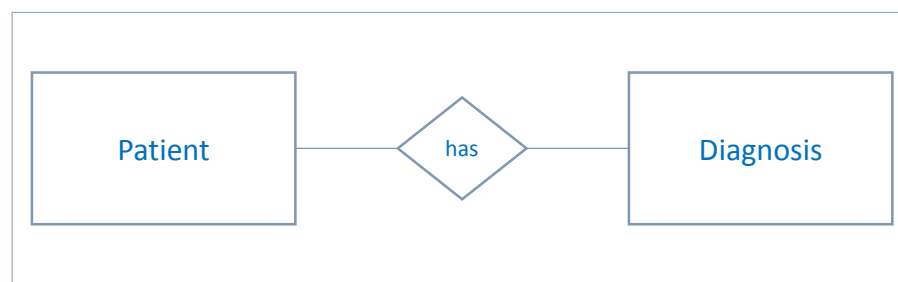


Figure 6.9: Representation of a relationship between entity Patient and Diagnosis in original notation

Cardinality further defines relationships between entities by placing the relationship in the context of numbers. It depends on the rules in a modelled system. To define cardinality of the relationship it is necessary to ask how often the entity occurs in conjunction with another entity. A ceiling is being sought (the maximum number). For example one could ask: “How many diagnoses can a patient have?” and then check also the other direction: “How many patients can have a diagnosis?”

There are relationships with different cardinalities:

- One-to-One (1:1),
- One-to-Many (1:m), and
- Many-to-Many (m:n).

More on E-R diagramming technique can be read on ER Diagram tutorial website <http://creately.com/blog/diagrams/er-diagrams-tutorial/>.

6.5.6.3 Building a Data Model

Data modelling is divided in 3 main activities:

- Developing a **conceptual data model**;
- developing a **logical data model** and
- developing a **physical data model**.

A **conceptual data model** is an abstract representation of a problem domain. A **logical data model** describes the data in as much detail as possible. It is bound to the selected type of data model (for example relational, hierarchical, object – relational) but without regard to physical implementation. A **physical data model** is bound to a selected implementation platform and will not be presented in this document. Below, the basics of conceptual modelling will be explored and a very important part of data models: the data dictionary will be presented.

6.5.6.4 Conceptual data modelling and conceptual data model

A conceptual data model describes the problem domain. The result is a general and abstract description of reality which helps team members to understand the data requirements. Development of a conceptual model is usually divided into three phases (13):

1. Requirements gathering;
2. Designing E-R diagram;
3. Normalization.

6.5.6.4.1 Requirements gathering

To produce an efficient data model it is necessary to document and understand the requirements of the problem domain. At this stage, a definition is needed of the **aims and objectives** of the Patient Registry (see [6.1.1](#) ‘Defining the Purpose, Objectives and Outputs of the Registry’) and also it is necessary to impose **limits on our system** (see [6.1.6](#) ‘Defining the Scope of the Registry & Building a Registry Development Team’), **review and evaluate the existing system of collection, storage and use of data and analyse the current operating environment** (see [6.1.3.1](#) ‘Overview of the Current State’) and **predict future requirements**.

The outputs of this phase are:

1. Defined Mission statement and Goals of PR (see chapter [6.1.1](#))
2. Defined scope of the PR (see chapter [6.1.6](#))
3. Preliminary list of entities (concepts/items/subjects about which the data will be stored- e.g. Person) and Preliminary list of attributes (properties about which data will be stored— e.g. Persons name, Persons address) (see chapter [6.3](#))

6.5.6.4.2 Designing E-R diagram

Designing an E-R diagram can be divided into 3 parts:

1. identification of entities;
2. identification of attributes for each entity; and
3. identification of the relationships between the entities.

The output of this phase is **graphical representation of data model** and **short descriptions of all identified entities**.

1. Identification of entities

At this stage, a list of entities that will be used in the new PR will be prepared.

1. First, on the basis of a preliminary list of attributes defined in the previous phase, identify entities.
2. Next, compare the obtained list of entities with the preliminary list of entities developed in the previous phase.
3. Compare the list of entities with the mission objectives for the database.
4. Add a description of the type of entity: a precise definition of the entity and why it is important for the patient registry project.

Rules to define the name of an entity:

1. The name should be a unique, descriptive name that is understandable by all the PR stakeholders.
2. The name should accurately, clearly and unequivocally identify the entity.
3. Use the minimum number of words that are necessary to describe the subject.
4. Do not use words that describe the physical characteristics (e.g., table, file format).
5. Do not use abbreviations or acronyms.
6. Do not use names that implicitly or explicitly identify more than one entity.
7. Use singular nouns.

2. Identification of attributes

At this stage, **the attributes will be added to the entities** from the preliminary attribute list and **the type of attributes will be defined** (key, non-key).

Keys are special type of attributes and are extremely important because they:

- ensure that each record in the entity is accurately (unique) identified⁵⁷;
- ensure different levels of integrity;

⁵⁷ For example in the entity Patient, PIX (Patient Index) can serve as a primary key of this entity.

- allow the creation of relationships between the entities.

3. Identification of the relationships between the entities.

Identified entities have relationships between them. Relationships are an important part of the E-R diagram.

There are 3 types of relationships with different cardinalities:

- relationship one-to-one (1:1);
- relationship one-to-many (1:m) and
- relationship many-to-many (m:n).

In the conceptual model the many-to-many relationships are not resolved.

6.5.6.4.3 Normalization

Database normalization is the process of designing a database with the desired properties. This optimizes the management of the database by eliminating redundant (duplicate) data and ensuring that only relevant data are stored. Normalization is based on so called normal forms and rules for their creation. The outputs of the normalization process are a **refined E-R model** and **refined descriptions of the entities**.

6.5.6.5 Logical data modelling and logical data model

A logical data model describes the data in as much detail as possible. It is bound to the selected type of data model (for example relational, hierarchical, object-relational) but without regard to physical implementation.

6.5.6.6 Data dictionary

An important part (especially for interoperability purposes) of the data model is also a **Data Dictionary** where all data elements (entities and attributes) are well defined. A typical description of data elements (or metadata⁵⁸ of data elements) includes (but it is not limited to this):

- Identification of the data element (name, short name, alias, ID)
- Definition of data element type (entity or attribute)
- Definition⁵⁹ of data element, where also the clear purpose of the data element is described
- Logical representation of data element (value set⁶⁰, permitted values, default values, data type etc.).

Data dictionaries are very often also called metadata repositories. The well-known, and also in health, often used standard for metadata repositories is **ISO/IEC 11179 Metadata registries**. More on metadata registries can be found on the ISO/IEC 11179 website: <http://metadata-standards.org/11179>. A good example of a metadata registry defined according to ISO/IEC 11179 is METeOR: <http://meteor.aihw.gov.au/content/index.phtml/itemId/181162>, the Australian Health Metadata Registry.

⁵⁸ Data about data.

⁵⁹ Short and helpful guidance on data elements definition could be found in chapter 6.3 'Registry dataset'.

⁶⁰ See also chapter 6.5.10 'The importance of terminologies and code lists'.

6.5.7 Entity-Attribute-Value Data Model in Medical databases

A widely used data modelling/design technique in clinical databases and clinical data repositories is Entity-Attribute-Value (EAV) design. Background component of EAV design is representing arbitrary information on some object as Attribute-Value list. An example of such representation in medical device implant database description would be:

Table 6.9: Units of implants as attribute-value lists

Hip Implant Unit	Ref no	Lot no
Acetabular Cup	9998-00-756	2582612
Inlay	8834-01-453	356321
Femoral Head	5632-01-234	234764
Femoral Stem	2345-03-234	234567

In [Table 6.9](#) we have 3 lists with 4 attribute-value pairs. Conventional columnar data model is represented on [Figure 6.10](#).

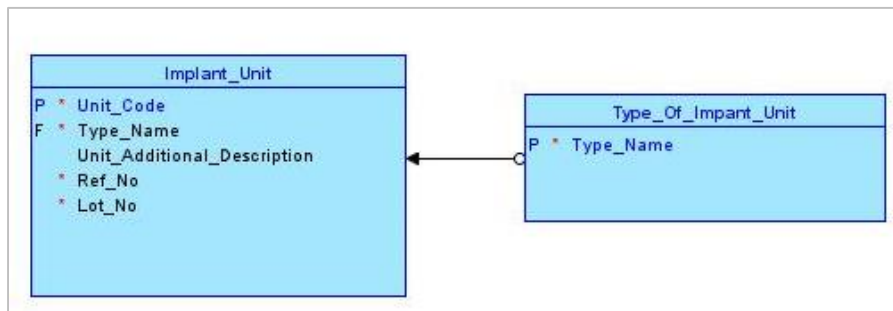


Figure 6.10: Example of conventional relational model with columnar form of attributes

Syntaxes of extensible markup languages like XML (16) are related to attribute-value pairs. XML elements that are delimited in open/close tags represent entities or attributes of entities.

However there are limitations in traditional relational oriented database modelling.

The conventional way to represent attributes for a class/entity in relational database modelling is in the form of columns in a table. That means one column per attribute:

- Hip Implant Component,
- Ref No,
- Lot No.

This approach is suitable when:

- there is a fixed number of attributes describing class or entity
- most or all attributes have values for a given instance of class.

Such columnar representation is not optimal for classes:

- with a potentially large number of attributes for a given instance and
- there may be several instances with attributes having unknown or inapplicable value

This situation is an analogy for the computer science sparse-matrix problem using the term “sparse data”. The term sparse data denotes a situation where there is a discrepancy between numbers of

potential attributes versus actual attributes. It should be also considered that mainstream relational database engines are usually limited to up to 1024 columns per table. It is not rational strategy for partitioning such enormous set of attributes across several tables. Also volatility of the data model should be considered, since the number of attributes (parameters) continually increases as medical knowledge and standards advances. This requires continual modifications to the schema and user interface as well.

The following two figures represent a case of transforming a conventional columnar model for subset of attributes into an EAV row level model.

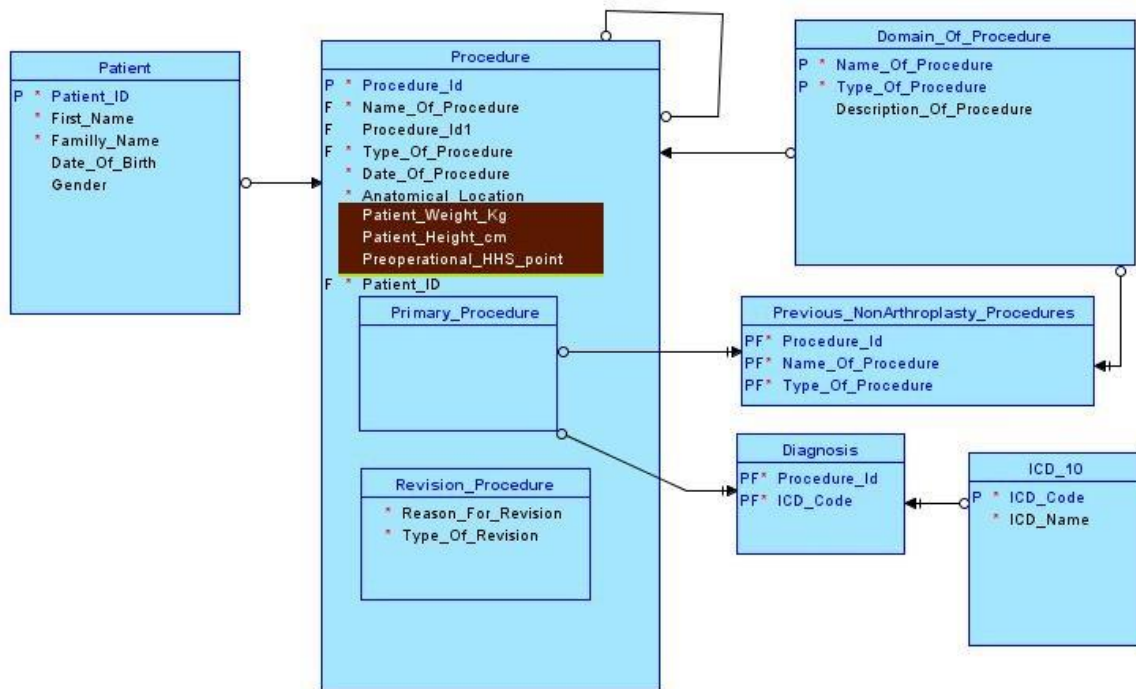


Figure 6.11: Example of conventional relational model before transformation to EAV entities

In [Figure 6.11](#) is represented part of relational data model with entities that have conventional columnar presentation of attributes. However in entity "Procedure" there are 3 attributes (brown rectangle) with allowed "unknown" values:

- Patient_Weight_kg,
- Patient_Height_cm
- Preoperational_HHS_point – orthopaedic measure as assessment of patient's hip disabilities

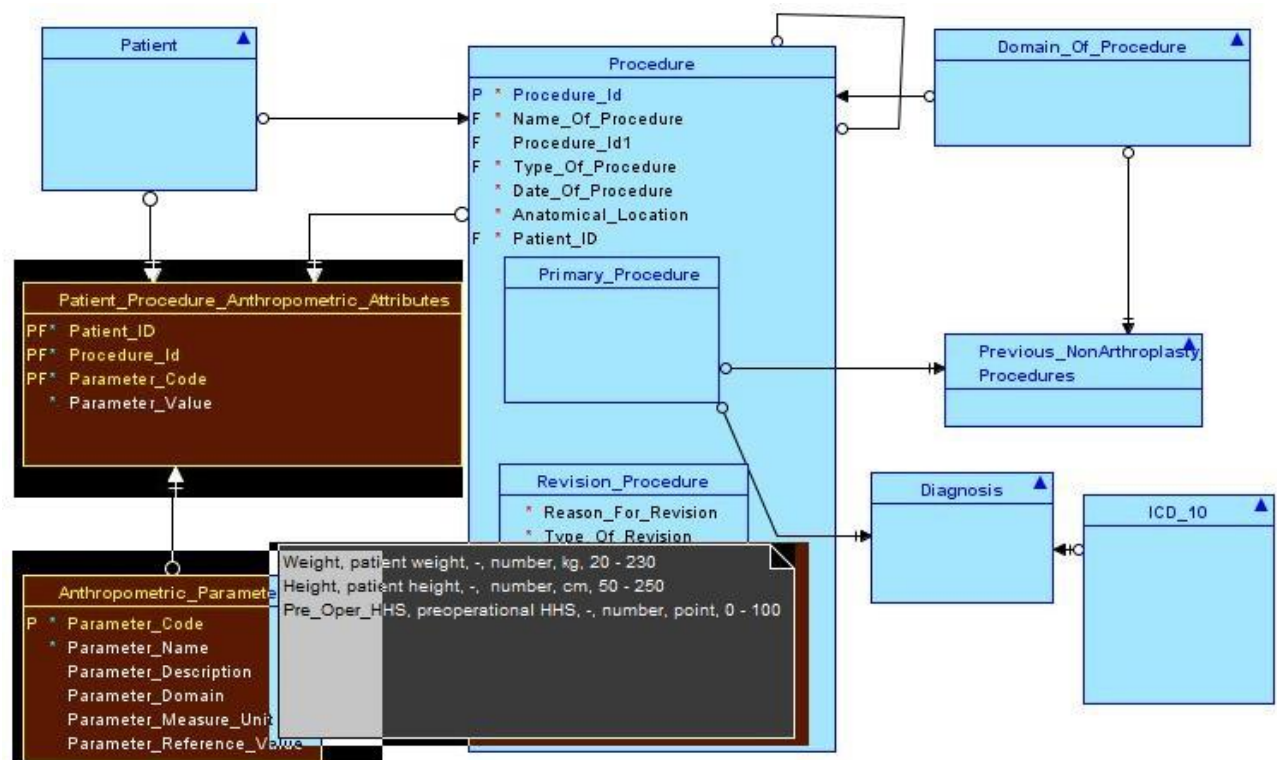


Figure 6.12: Data model with EAV entities

All attributes represent anthropometric measures of the patient. As such they are candidates for separated set of attributes that might be expandable with many other measures. When transforming columnar representations of these attributes into a model that considers EAV modelling/design technique the result model is presented on [Figure 6.12](#).

The model seems more complex at first glance, but provides more flexibility in management of changes in the schema. The basis of the transformation to EAV is incorporation of the metadata describing the nature of the attribute needed: business rules, data type, and unit of measure.

6.5.8 Temporal modelling in medical databases

According to (24) three various type of databases are specific for medical database systems:

- **Administrative databases** that serves as operational support in organizational and economic aspect of information technology support: ERP accounting, ERP assets, ERP Human Resources, CRM. Databases in this realm are not prepared for clinical questions although there supposed to be connections (Patients, Organisation structure with Personnel, Accounting details) to other realms where more clinical orientation of data is structured. Database structure is complex and demands detailed knowledge of transactions to use the data for reference or analyzing.
- **Clinical databases** for tracking procedures and services, electronic patient records, more granular details of medical devices, cases, diagnoses. As such they represent a daily operational tool in:
 - planning therapies, procedures,
 - occupancy or availability of medical assets or human resources like therapists, physicians).

A clinical database can also serve as clinical research infrastructure. Data must be tracked in time with additional business rules on events (individually or on sequence of events in workflow) that control consistency of data as time series.

- **Registries** like disease registry, drug registry, incidents. Registries connect various sources resulting in spatial-time oriented databases spanned above data clusters like patient, disease, medical device, physician, geo-location, date-time. Data from registries are serving risk-management processes, trend and survival statistical analysis, incident modelling. The data coming into registries are based on operational data in the former two types of databases with additional data retrieved from quality assurance, data consolidation processes.

All three types of databases are included in consolidated data subsets (as data mart or data warehouse pivoted form) used in researches, cohort studies, data control in operational data tables with time as a mandatory essential dimension.

The success of the process for data consolidation is significantly dependent on timing data consistency and clear simple presentation of time-dependency.

The technique in conventional modeling that considers **time** as entity specific is called **temporal data modelling**. A database that involves time constraints and controls is called a **temporal oriented database**.

In (25) are defined two main concepts of modelling TIME as attribute in relational or object model:

1. Time instants concept which considers time dependent entities (temporal entities) as series of events (changes) forming log/track or time series of changes (transactions). Records in such time series might be additive meaning we can perform aggregating functions like sum, count on different grouping criteria. Maintenance of time series might be more flexible. To provide consistency the rule of sequencing must be obeyed between various events in the sequence. Time attribute is represented as Dat_Time_Of_Change.
2. Time intervals concepts relating to the situations/states/statuses of the entity/object for specified time interval defined with upper (Date_Time_To) and lower limit (Date_Time_From).

Both concepts are derivable from each other with pivoting techniques. It must be pointed out that in second concepts other time dependent attributes cannot be additive. So aggregating techniques usually requires transformations of data.

Business rules on time attributes must consider controlling the risk of the time interval overlapping the same categories of data (temporal entities).

In the following example both concepts of modelling are presented on the example of the ER diagram in a registry database of patients, implants and procedures which are performed for implant insertions or removal by physician.

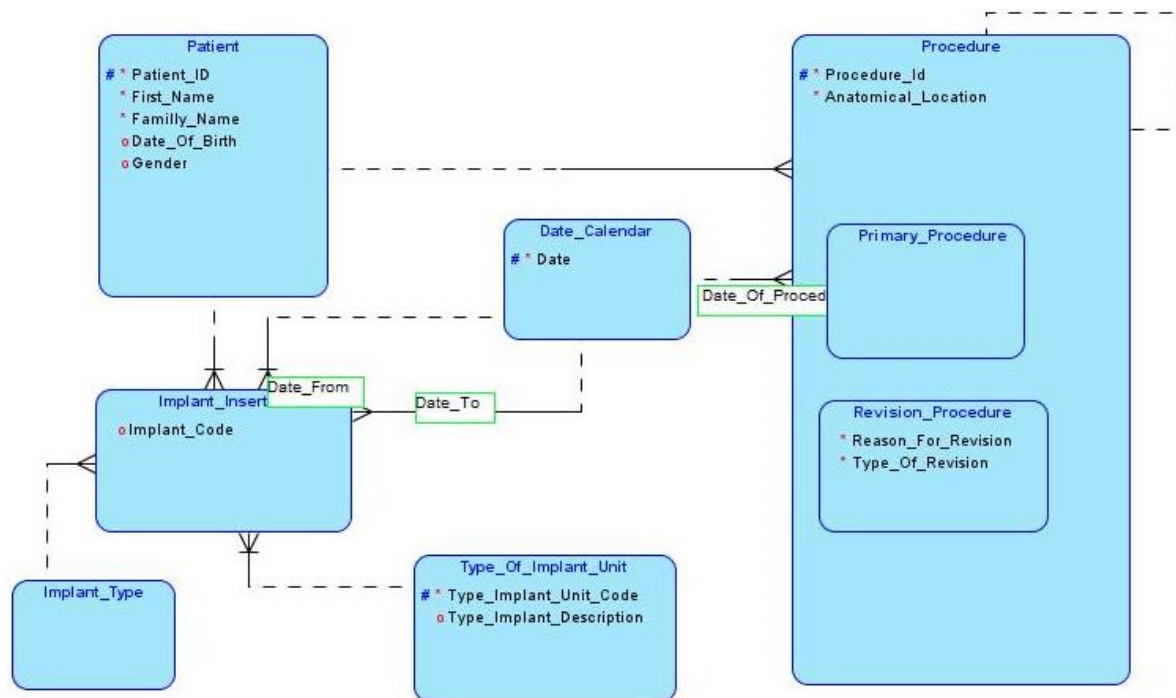


Figure 6.13: Data model with time interval temporal entities

In this example the complete status of the implant used for the patient is represented with the time interval temporal entity Implant Inserted. Such a concept provides a complete history of all implant units and how they form an implant in “patient”. But the time series of procedures and time implants are independent. So consolidation of these two time series must be performed through time interval operators for comparison and, compliance of date of procedure with proper time interval of the implant unit.

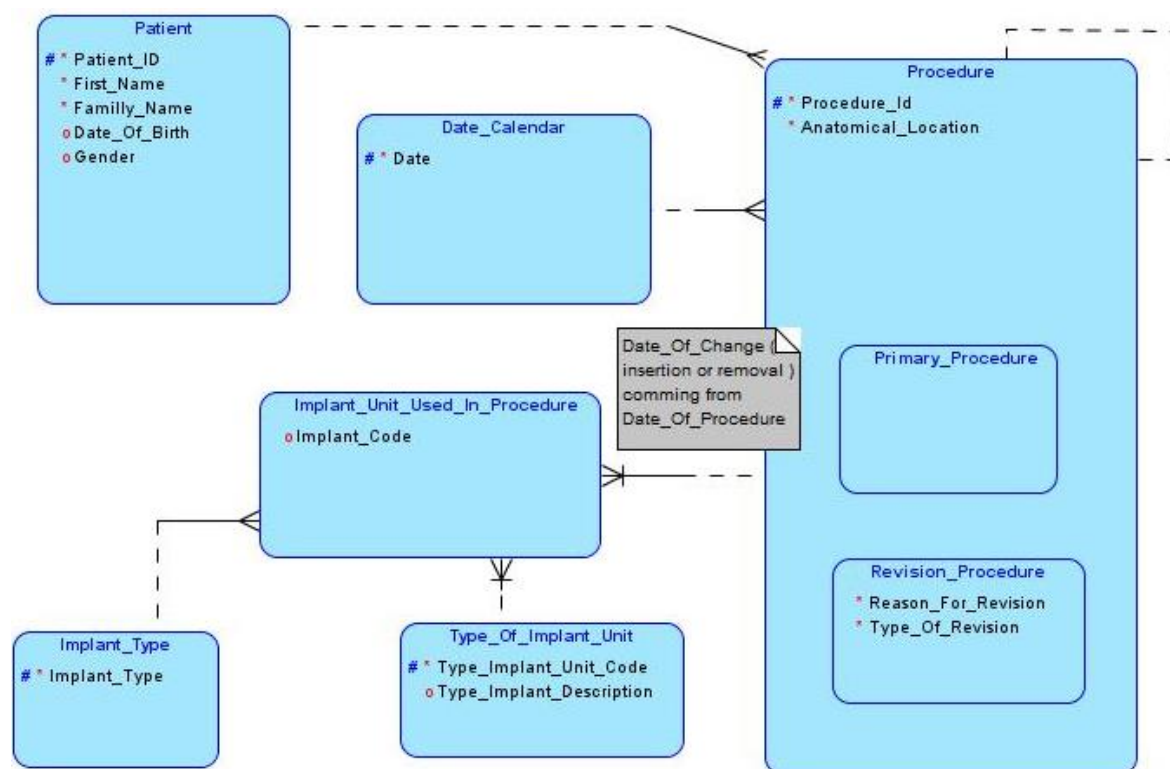


Figure 6.14: Time instant concept of time modelling

In the second example the time instant concept for temporal entity `Implant_Unit_Used_In_Procedure` is used. This way changes or transactions (insertions, removals) are tracked for each individual unit of implant. Time or Date of change is defined with `Date_Of_Procedure`. So `Implant_Unit_Used` is indirectly temporal dependent. The model is much more flexible and clear for performing business rules, but requires a different query technique for retrieving consolidated implant (all units) in every moment of time.

6.5.9 Knowledge management using archetypes

Beal (28) states that the **health domain** is open-ended and there is a huge number of constantly changing concepts. Therefore he proposed instead of a “classical” **single level approach** in which both information and knowledge concepts are hard-coded directly in software and database models, a **dual level approach** where information systems are built from an information model only, and driven at runtime by knowledge-level concept definitions, known also as **archetypes**.

Nowadays health information systems are built using both approaches. But due to an increased need for sharing of patient information across multiple settings and within diverse electronic health record repositories, **dual-level approach is in literature recognized as a promising solution to ensure electronic health records interoperability** (28-31). Therefore also PARENT is using this approach in defining the new Arthroplasty registry as a role model.

In the next subchapters dual level approach methodology will be presented with an emphasis on archetypes modelling using Open EHR.

6.5.9.1 OpenEHR

OpenEHR is a virtual community working on interoperability and computability in e-health. Its main focus is electronic patient records (EHRs) and systems (32). The success of openEHR is in no small part due to the formal acceptance of **CEN 13606** as a European and ISO standard. This standard is based on many aspects of the openEHR design approach, and part 2 of the standard is a snapshot of the openEHR Archetype specifications (32).

OpenEHR is based on multi-level modelling approach where clinical models (**archetypes**) built by domain experts are separated from data representation and sharing (**reference model**). The semantic architecture of OpenEHR is depicted on the picture ([Figure 6.15](#)) and in the following lines the core elements of Open EHR will be presented.

Reference model is mostly technical infrastructure – generic technical artefacts for representing health information (data structures and types, health record organisation, security etc.) and is hidden from content modelling for clinicians.

Archetypes are standardized computable models of discrete clinical concepts, for example: blood pressure, symptom, medication order, family history, Breast cancer histopathology result. They should capture as many clinical perspectives as possible to be universally applicable, but can be designed also for specific local use cases.

Templates are used to create datasets for example for discharge summary, Arthroplasty registry etc. With templates data entry definitions and message definitions are defined for a particular clinical

context or purpose. Templates are an aggregation of archetypes according to specific use-case. In templates archetypes are constrained to make them practical— e. g. unwanted items are removed, set default values, bind the terminology etc.

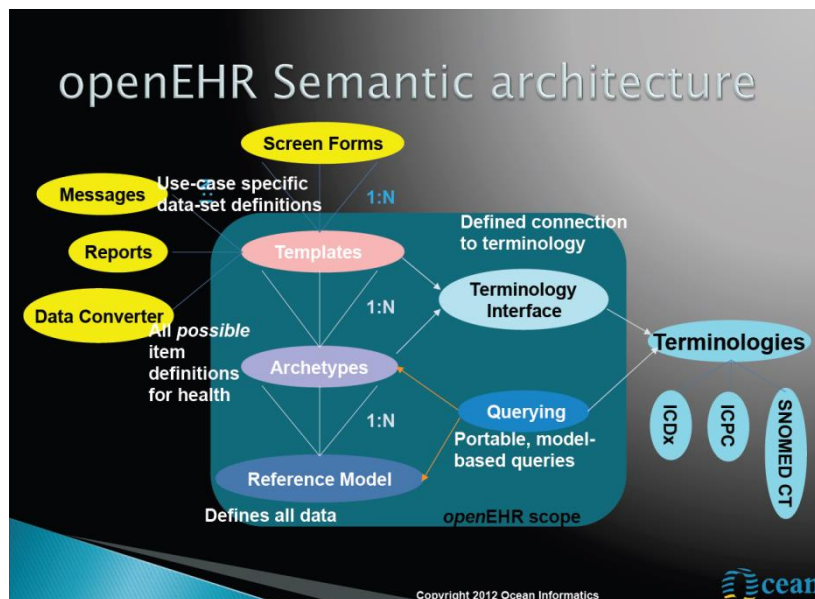


Figure 6.15: OpenEHR Semantic architecture (Beale, Thomas: Architecture Overview, Ocean Informatics, 2012)

6.5.9.2 Modelling Archetypes

To capture clinical knowledge (clinical) archetypes have to be modelled. It is recommended that the archetype design is led by a person with experience in archetype modelling. Process of modelling archetypes (33):

1. Identify all clinical concepts
2. Explore if archetypes for the identified concepts already exists
 - If yes: **3.a Use existing archetypes**
 - If no: **3.b Design new archetypes**

6.5.9.2.1 Identify all clinical concepts

The subject, activity, or task which need to be modelled must be researched. For simple concepts such as *body weight* one archetype has to be designed, but for more complex concepts (for example *pregnancy*) multiple archetypes must be modelled.

To visualize the researched subject, activity or task it is recommended to use some kind of Mindmap tool (for example XMind; www.xmind.net/).

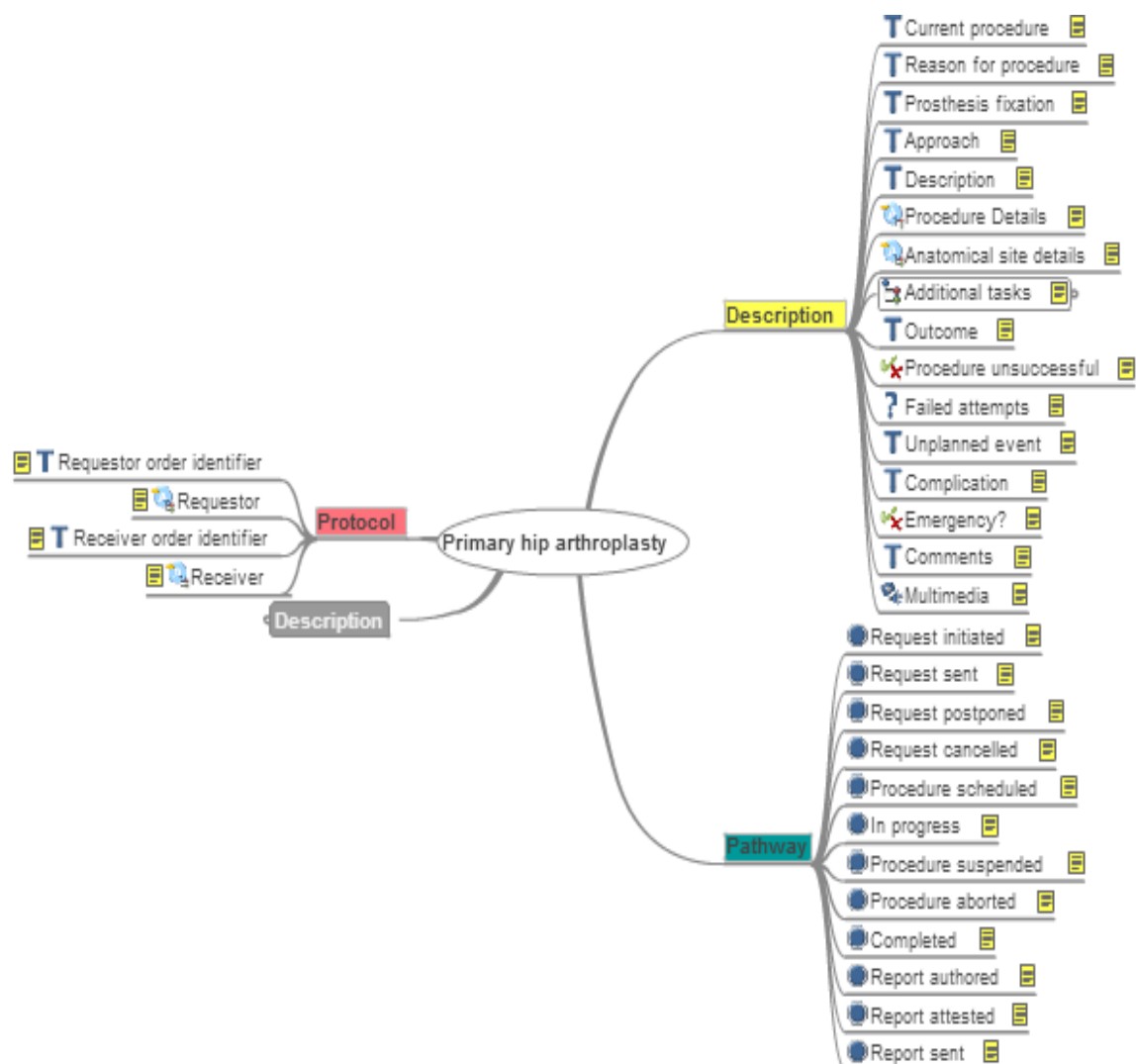


Figure 6.16: Primary Hip Arthroplasty Report presented as a mindmap

6.5.9.2.2 Explore if archetypes for the identified concepts already exists

Search for already defined archetypes in:

- openEHR CKM (www.openehr.org/ckm)
- NEHTA CKM (dcm.nehta.org.au/ckm/)
- NHS archetypes (www.connectingforhealth.nhs.uk/systemsandservices/clinrecords/nccr)
- etc.

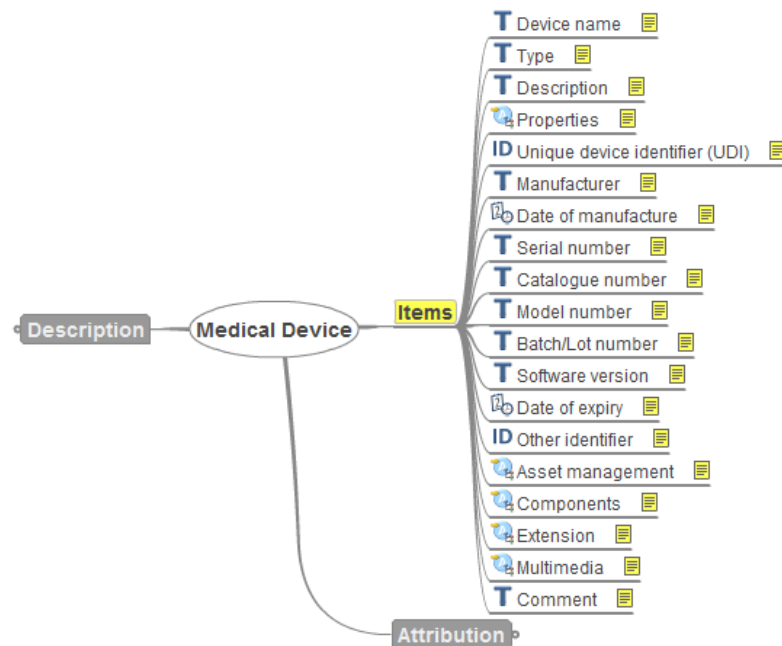


Figure 6.17: Example of updated Medical Device archetype as a mindmap

6.5.9.2.3 Design new archetype

Designing a new archetype can be done in the following steps (33):

1. Gather content
2. Organise the content
3. Choose the archetype class
4. Build the archetype
 - a. Name the archetype
 - b. Select the structure
 - c. Add data types
 - d. Add constraints
 - e. Add metadata
 - f. Add terminology
5. Collaborate & Publish
6. Add to a Template

1. Gather content

Consider the clinical concept from all possible angles and point of views. Think about how different clinicians may record the identified data.

It is recommended that the guided interview technique is used (see chapters [6.5.2.1](#) and [6.5.2.2](#)) or an interactive modelling workshop (using mind map tool) to gather information on clinical concepts. There are a lot of useful sources of information to model clinical content such as: different forms (also in paper), existing computer applications, clinical audit datasets, clinical trials datasets, patient registry datasets, reporting obligations etc. Look for similar projects locally and internationally, search for publications on the identified topic etc. To be as broad as possible research the clinical concept from different perspective: different medical specializations, nursing, researchers, public health, clinical decision support etc.

2. Organise the content

Organize the content using mindmap. Focus on identifying:

- Purpose – container or navigation
- Context
- Data elements
- Protocol
- State – context for interpretation
- Events
- Pathway steps
- Concepts needing coding/terminology

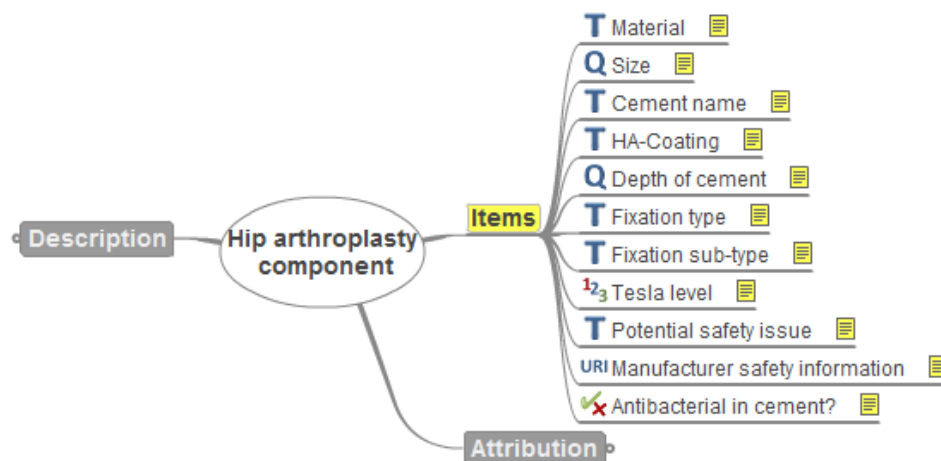
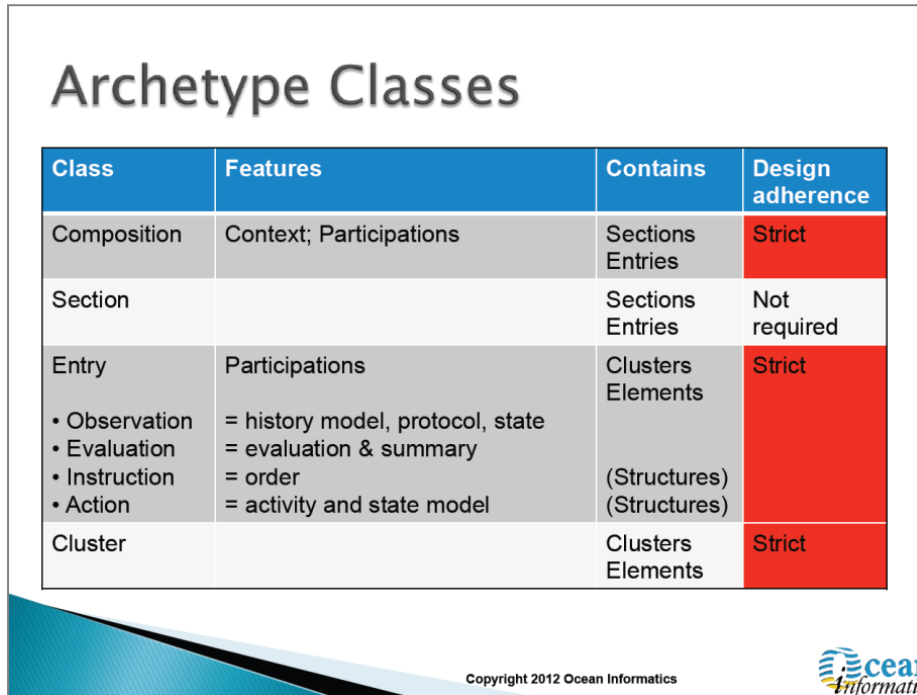


Figure 6.18: Example of newly defined Hip arthroplasty component archetype as a mindmap

3. Choose the archetype class

Although domain experts do not need to care about openEHR reference model (RM), the author of the archetype should know the possible “archetype types” called **archetype classes** presented in the RM.

As can be seen (see [Figure 6.19](#)) the existing archetypes classes are: Composition, Section, Entry and Cluster. More on this topic can be read in OpenEHR wiki (www.openehr.org/wiki).



Class	Features	Contains	Design adherence
Composition	Context; Participations	Sections Entries	Strict
Section		Sections Entries	Not required
Entry	Participations	Clusters Elements	Strict
• Observation • Evaluation • Instruction • Action	= history model, protocol, state = evaluation & summary = order = activity and state model	(Structures) (Structures)	
Cluster		Clusters Elements	Strict

Copyright 2012 Ocean Informatics




Figure 6.19: Types of ENTRY Archetype class (Leslie, Heather & all: OpenEHR archetypes in detail, Ocean Informatics, 2012)

[Figure 6.20](#) demonstrate the different types of ENTRY class. To select the proper archetype class the tool presented by Ocean Informatics can be used ([Figure 6.21](#))

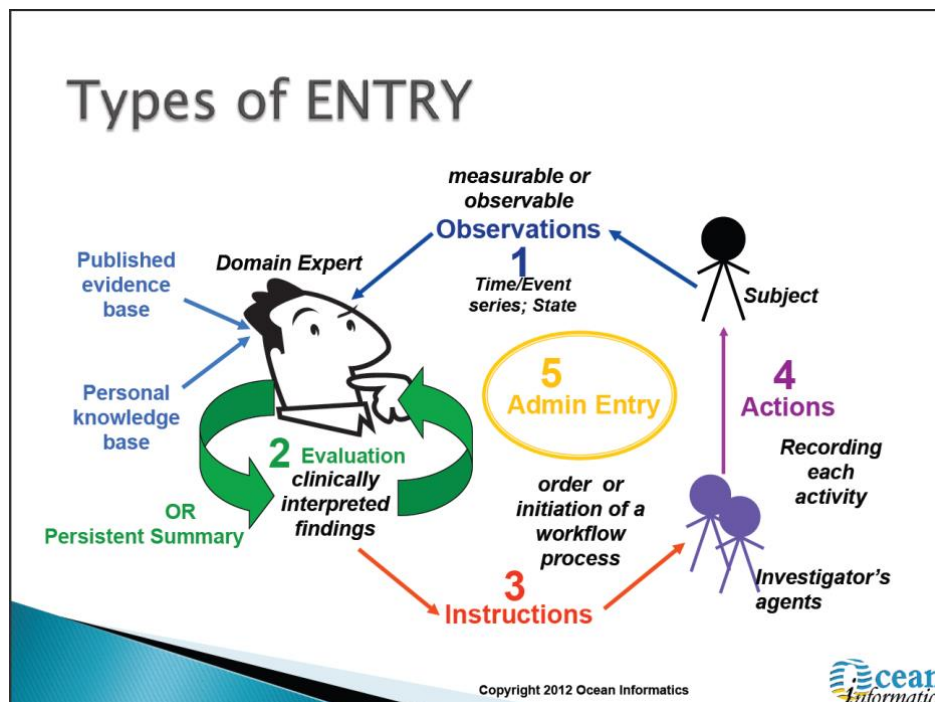


Figure 6.20: Types of ENTRY Archetype class (Leslie, Heather & all: OpenEHR archetypes in detail, Ocean Informatics, 2012)

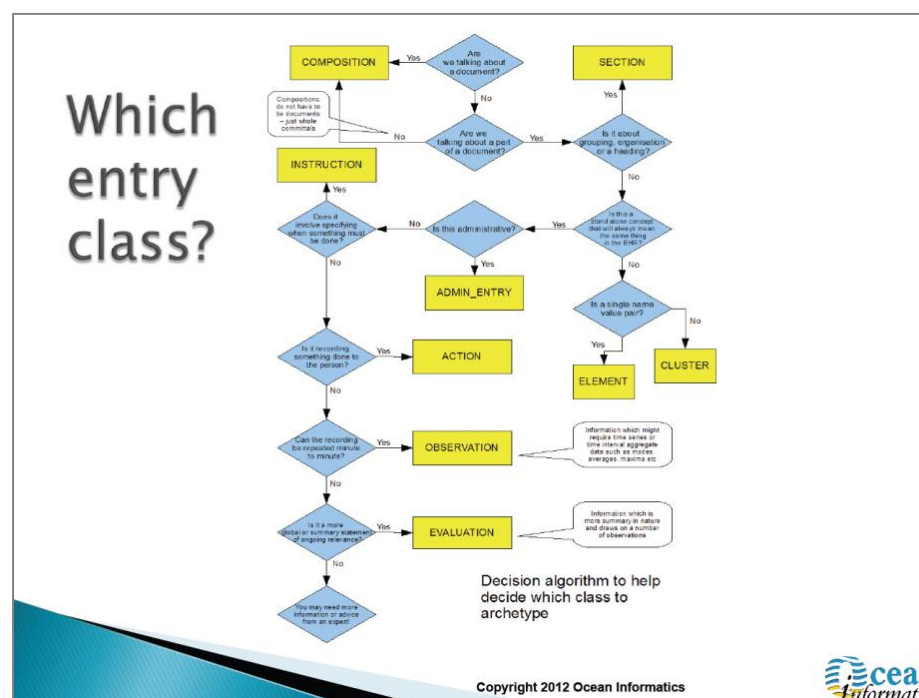
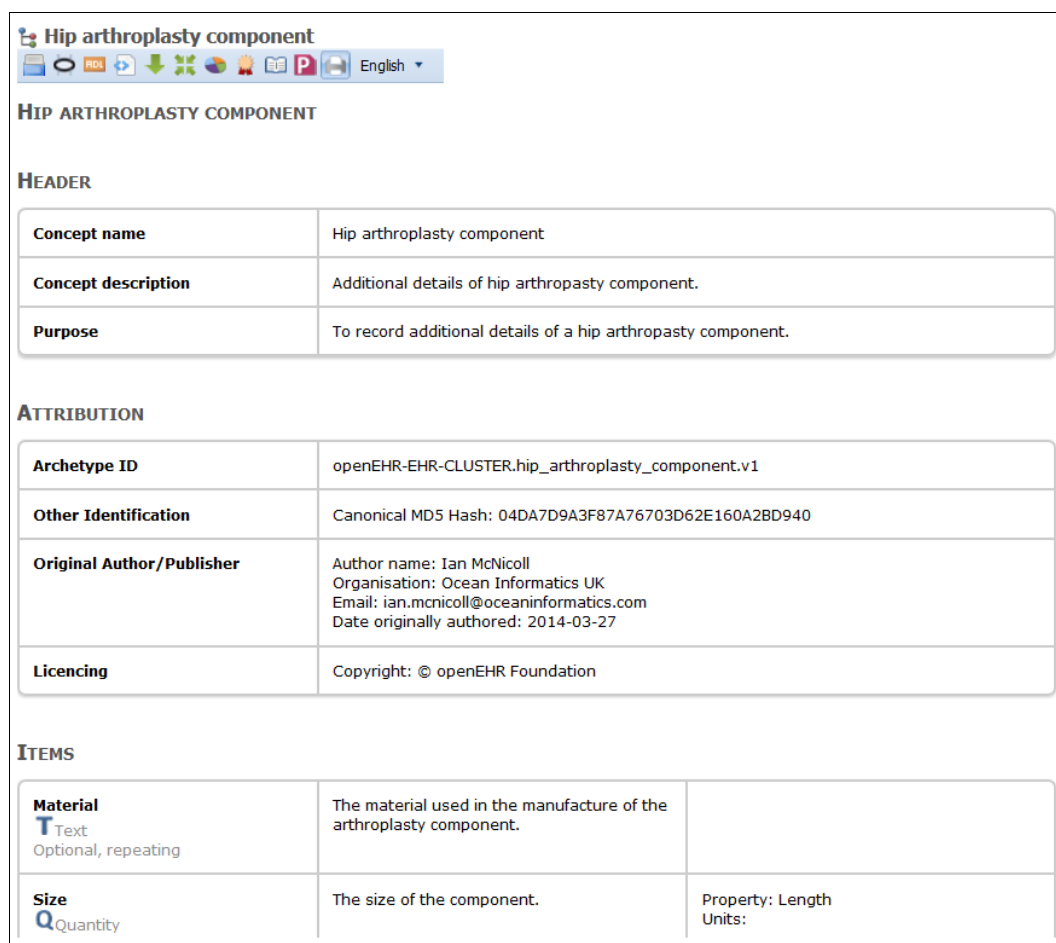


Figure 6.21: Process of class type selection (Leslie, Heather & all: OpenEHR archetypes in detail, Ocean Informatics, 2012)

4. Build the archetype

To build an actual archetype a computer application for archetype design has to be used (for example Archetype Editor). Tools for modelling archetypes can be found on the web page www.openehr.org/downloads/modellingtools.



Hip arthroplasty component

English

HIP ARTHROPLASTY COMPONENT

HEADER

Concept name	Hip arthroplasty component
Concept description	Additional details of hip arthropasty component.
Purpose	To record additional details of a hip arthropasty component.

ATTRIBUTION

Archetype ID	openEHR-EHR-CLUSTER.hip_arthroplasty_component.v1
Other Identification	Canonical MD5 Hash: 04DA7D9A3F87A76703D62E160A2BD940
Original Author/Publisher	Author name: Ian McNicoll Organisation: Ocean Informatics UK Email: ian.mcnicoll@oceaninformatics.com Date originally authored: 2014-03-27
Licencing	Copyright: © openEHR Foundation

ITEMS

Material T Text Optional, repeating	The material used in the manufacture of the arthroplasty component.	
Size Q Quantity	The size of the component.	Property: Length Units:

Figure 6.22: Hip arthroplasty component archetype

5. Collaborate & Publish

Archetypes should be published on openEHR CKM (www.openehr.org/ckm/).

6. Add to a Template

SLOVENIA RES PRIMARY HIP ARTHROPLASTY REPORT

Where not otherwise stated, elements are optional ([0..1]).

Collapse All **Show Annotations** **Show Paths**

- other_context
- Previous procedures
- Current procedure
 - Consent to share information
 - Patient state (Additional details)
 - Primary hip arthroplasty
 - Current procedure [1..1] Total standard hip replacem
 - Reason for procedure [0..*]
 - Primary Osteoarthritis
 - Dysplasia /dysplasia with di
 - Post-Traumatic
 - Aseptic Necrosis of Femoral
 - Rheumatoid Arthritis, other a
 - Other
 - Prosthesis fixation [0..*]
 - Cemented
 - Uncemented
 - Hybrid-cemented stem only
 - Inverse_Hybrid
 - Approach [0..*]
 - Anterior
 - Anterolateral
 - Direct lateral
 - Posterolateral
 - Minimally Invasive
 - Hip device components
 - Device name [1..1] Hip implant (Default: Hip implant)
 - Medical Device [0..*]
 - Hip arthroplasty component manufacturer device name [1..1]
 - Type
 - Part
 - Hip arthroplasty component [0..*]
 - Material
 - Material Sub-group
 - Specific Material
 - Size 0..20 mm
 - Fixation type Cemented
 - Fixation sub-type
 - Tesla level 0..5
 - Antibacterial in cement?
 - Manufacturer
 - Ref number
 - Batch/Lot number

Figure 6.23: Example of a part of template on Slovenia RES primary hip arthroplasty report (tabular view)

6.5.10 The importance of Terminologies and Code lists

The **unambiguous definition**⁶¹ of PR data elements and **proper selection of terminologies and code lists** for defined data elements are **prerequisites for semantic interoperability**. Representing clinical information in standardized ways allow humans and computers to understand clinical information correctly (see also chapters [3.2.5.1](#) ‘Standards, models and tools’ and [10.11.2](#) ‘eHealth standards’).

Therefore it is strongly recommended to use **standardized and internationally recognized terminologies and code lists** if they exist for the given item. For example diagnosis and conditions are most commonly coded using World Health Organisation (WHO) International Classification of Diseases (ICD).

When selecting code lists many things have to be considered and for bigger (especially cross border) projects it is very valuable **to prepare a list of coding system selection criteria** in advance. As a start, coding system selection criteria can be used defined by project European Patients – Smart Open Services (epSOS) (34) and also recommended in Guidelines on minimum/non-exhaustive patient summary dataset for electronic exchange in accordance with the cross-border directive 2011/24/EU (35):

- Be internationally used;
- Be in use by some project participants;
- Have translations in a number of different languages;
- Have a maintenance process;
- Have a number of transcoding systems/services, e.g. mapping facilities;
- Be easy to implement;
- Take account of the cost of licences, implementation and maintenance

⁶¹ See ISO/IEC 11179 – 4 (1995) Good Data Definition

References

1. Chonoles, Michael Jesse and Scardtt, James A.: UML 2 for Dummies, Hiongry Minds 2003.
2. Introduction to OMG's Unified Modeling Language™ (UML®), Web page http://www.omg.org/gettingstarted/what_is_uml.htm (accessed 25th June 2014).
3. Asbjørn Rolstadås (1995). "Business process modeling and reengineering". In: Performance Management: A Business Process Benchmarking Approach. p. 148-150.
4. What is BPM Anyway? Business Process Management Explained, <http://www.bpm-institute.org/resources/articles/what-bpm-anyway-business-process-management-explained> [accessed 10th May 2014].
5. Hammer, M., Champy, J. (1993, 2003). Reengineering the Corporation: A Manifesto for Business Revolution, New York: Harper Business, 1993, New York: Harper Business Essentials, cop. 2003
6. Keen P.G.W., Knapp E.M (1996). Every Manager's Guide to Business Processes, Harvard Business School Press, 1996
7. Keen, P.G.W.: The Process Edge, Creating Value Where It Counts, Harvard Business School Press, 1997
8. Event based process chain (EPC), Aris community web page <http://www.ariscommunity.com/event-driven-process-chain> [accessed 23rd May 2014].
9. Chinosi, Michele, BPMN: An introduction to the standard, Computer Standards & Interfaces, Volume 34, Issue 1, January 2012, Pages 124–134.
10. Davids, R., Brabänder, E. (2007): Aris Design Platform, Getting Started with BPM, Springer- Verlag London Limited 2007
11. White, Stephen A., Introduction to BPMN, IBM Corporation, http://www.omg.org/bpmn/Documents/OMG_BPMN_Tutorial.pdf [accessed 10th May 2014].
12. Gliklich RE, Dreyer NA, eds. Registries for evaluating patient outcomes: A User's Guide. 3rd ed. 2014.
13. Hernandez, Michael J.: Database Design for Mere Mortals™: A Hands-On Guide to Relational Database Design, Second Edition, Addison Wesley Professional, 2003
14. Everest, G.: Database management: objectives, system functions, and administration, New York, McGraw-Hill, 1986
15. Riordan, R.: Designing Effective Database Systems, Addison Wesley Professional, 2005
16. IIBA International Institute of Business Analysis: Business Analysis Body of Knowledge (Babok Guide V2.0), IIBA 2009, ISBN-13: 978-0-9811292-1-1
17. Craig Larman (1998) : Applying UML and Pattern, An Introduction to OO Analysis and Design Prentice-Hall Int London ISBN: 0-13-748880-7
18. Richard Barker: Case Method-Entity Relationship Modelling Oracle Corporation 1990 UK Limited
19. Vojislav Ivetić, Janko Kersnik: Diagnostične preiskave za vsakdanjo uporabo, Združenje zdravnikov družinske medicine 2007, ISBN: 978-961-91889-5-8
20. Rafael J. Curbelo · Estíbaliz Loza · María Jesús García de Yébenes · Loreto Carmona: Databases and registers: useful tools for research, no studies Rheumatol Int (2014) 34:447–452
21. Nadkarni PM¹, Marengo L., Easing the transition between attribute-value databases and conventional databases for medical data. Proc AMIA Symp. 2001:483-7.
22. European Arthroplasty Register: <http://www.ea-efort.org/registers.aspx> access [15.12.2014]
23. OB Valdoltra Register Artroplastike <http://www.ob-valdoltra.si/raziskovalna-dejavnost/publikacije-register/register-artroplastike-ob-valdoltra> access [15.12.2014]
24. Shahar Y¹, Combi C.: Timing is everything. Time-oriented clinical information systems. West J Med. 1998 Feb; 168(2):105-13.

25. Rafael J. Curbelo · Estíbaliz Loza · Maria Jesús García de Yébenes · Loreto Carmona: Databases and registers: useful tools for research, no studies. *Rheumatol Int* (2014) 34:447–452
26. Richard T. Snodgrass: *Developing Time-Oriented Database Applications in SQL*. Morgan Kaufman Publishers, 2000, ISBN: 1-55860-436-7
27. Richard T. Snodgrass: *TSQL2 and SQL3 Interactions*. Association for Computing Machinery / Special interest group on management of data
28. Beal, Thomas (2002), *Archetypes: Constraint-based Domain Models for Future-proof Information Systems*
29. Tapuria, Archana and all (2013): *Contribution of clinical Archetypes, and the Challenges, Towards Achieving Semantic Interoperability for EHR*, Healthcare Informatics Research, December 2013
30. Costa, Catalina Martinez and all (2011): *Clinical data interoperability based on archetype transformation*.
31. Garde S, Knaup P, Hovenga EJS, Heard S (2007): *towards Semantic Interoperability for Electronic Health Records: Domain Knowledge Governance for OpenEHR Archetypes*. *Methods of Information in Medicine* 46(3): 332-343.
32. OpenEHR web page, <http://www.openehr.org/>, access [25. 06. 2014]
33. Leslie, Heather & all: *OpenEHR archetypes in detail*, PPT, Ocean Informatics, 2012
34. *The experience of selecting the code systems for the development of the epSOS master value catalogue (MVC)*, September 2013
35. *Guidelines on minimum/non-exhaustive patient summary dataset for electronic exchange in accordance with the cross-border directive 2011/24/EU*, eHealth Network, November 2013 http://ec.europa.eu/health/ehealth/docs/guidelines_patient_summary_en.pdf access [1st June 2014].
36. Stellman, Andrew; Greene, Jennifer (2005). *Applied Software Project Management*. O'Reilly Media. ISBN 978-0-596-00948-9.

7 PATIENT REGISTRY INFORMATION SYSTEM DEVELOPMENT AND IMPLEMENTATION

Vesna Lešnik Štefotič

Computer based patient registry information systems are nowadays typically web-based and allow end users to access the study database through the Web using a web browser. Many of them also operate on Software as a Service model (SaaS).

Key principles:

- computer based information system development is an engineering approach following the so called Software Developing Life Cycle with its phases:
 1. Planning and Requirement Analysis
 2. Defining software (SW) Requirements
 3. Designing the product architecture
 4. Building or Developing the Product
 5. Testing the Product
 6. Deployment in the Market and Maintenance
- Active user involvement in the computer based PR information system development is crucial for project success.
- users are typically involved in all stages of SDLC except in the product building stage,
- users are especially needed as a key members in system analysis and system design process, as user requirements drive the entire system development
- users involvement in the SW development increases user acceptance of the system
- proper user training (use of SW combined with PR content and rules) is essential for successful use of a computer based PR information system

Healthcare is **information-intensive**, generating huge amounts of data every day. It is estimated that up to 30% of the total health budget may be spent on handling information (1).

Also patient registries are dealing with data and information, collecting it, looking for it, storing it, 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 this chapter the basics of patient registry information system development and implementation will be described by presenting a typical software development lifecycle and emphasizing the importance of the (end)user in this process. The different possibilities to obtain patient registry software (SW): in-house development, buying/using PR SW product or outsource the development of PR SW will be addressed. Training in PR software is very essential to the proper and efficient use of the application, therefore it will be covered as a special topic.

After reading this chapter the reader will:

- understand the basics of SW development lifecycle and different SW development models
- understand the importance of user involvement in SW development
- be aware of different options to obtain PR SW

7.1 Computer based Patient Registry Information System

Laudon (2) defines an **information system technically** as a **set of interrelated components** that collect (or retrieve), process, store and distribute information to support decision making and control, helping people analyse the problems and visualize complex subjects, and create new products.

There are three main activities in each information system (see [Figure 7.1](#)): input, processing, and output which produce the needed information. Feedback in an information system is output returned to the information system with the aim to evaluate and refine the input.

Also a Patient Registry can be seen as an information system. Input is the data on patient health issues, processing is done by classifying, arranging or calculating the data, outputs are relevant reports, alerts etc. Feedback can be, for example, a quality report on data collected which will initiate better data quality control.

Computer based patient registry information systems used for clinical data collections in research and operational settings are usually called **Electronic data capture (EDC)** applications or SW. They are used to collect clinical data in electronic forms. Modern EDC software applications are typically web-based and utilize a thin client. This allows end users to access the study database through the Web through a web browser without the need for the installation of an application on the local computer. Many of them also operate on Software as a Service model (SaaS).

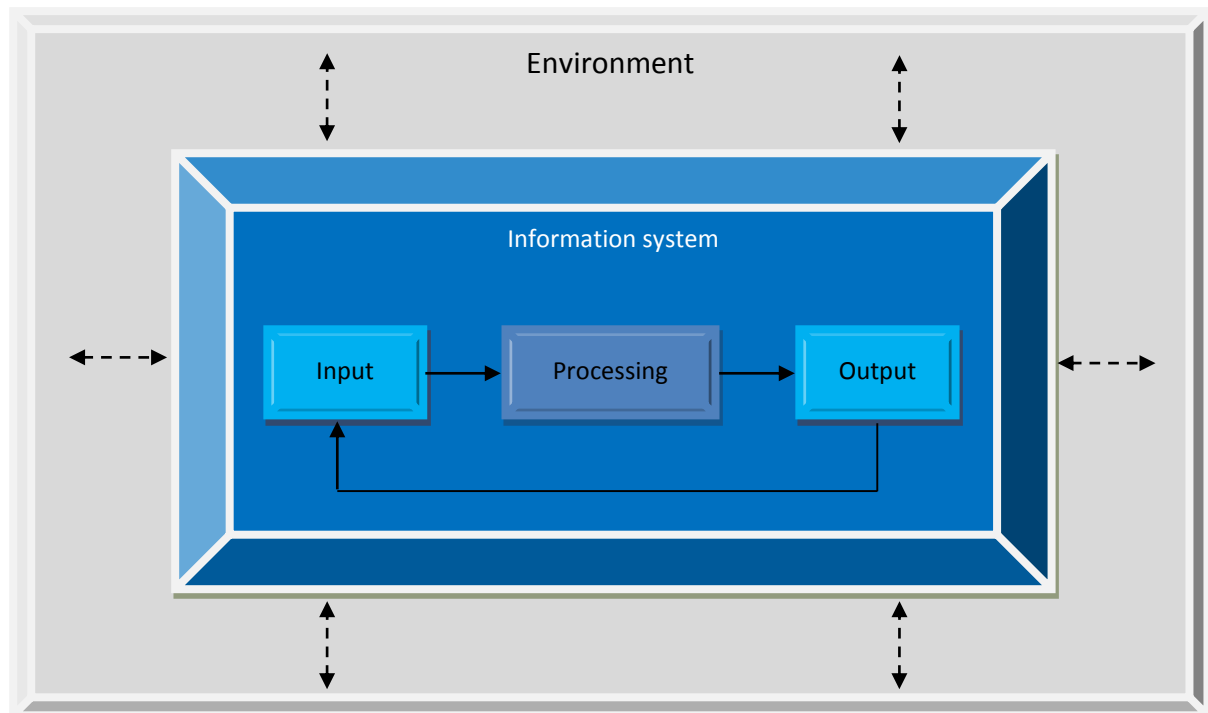


Figure 7.1: Main components of an information system

7.2 Development of Registry Information System

Computer based information system development (short software development) is an engineering approach following the so called Software Developing Life Cycle (SDLC). Typical stages of SDLC are (3):

1. Planning and Requirement Analysis (requirements gathering and system analysis, feasibility study (economical, operational & technical), planning of basic project approach)
2. Defining software (SW) Requirements (preparing and approving - by customer/user - of Software requirements specification document)
3. Designing the product architecture (preparing and approving the product architecture)
4. Building or Developing the Product (building a product, code generation)
5. Testing the Product
6. Deployment in the Market and Maintenance (formally release of the product, includes user training, regularly maintenance is required)

There is also an international standard for software life-cycle processes - ISO/IEC 12207. Standard that defines all the tasks required for developing and maintaining software.

There are various software development life cycle models prescribing a series of steps to ensure success in SW development. The most important and popular SDLC models followed in the industry are (3):

- Waterfall Model
- Iterative Model
- Spiral Model
- V-Model
- Big Bang Model

The other related methodologies are Agile Model, RAD Model – Rapid Application Development and Prototyping Models. More on SW development process models can be read in Tutorial on Software Development Life Cycle (www.tutorialspoint.com/sdlc/index.htm).

7.2.1 Important role of users

Regardless of the selected model of SW development the **active user involvement** in the development cycle **is crucial for project success**. Users have the domain knowledge and have their expectation towards the functionalities of the SW application. **User requirements drive the entire system-building effort**. Users must have sufficient control over the design process to ensure that a system reflects their priorities and information needs, not the biases of the technical staff. **Working on the design also increases users' understanding and acceptance of the system**. Insufficient user involvement in the development effort is a major cause of system failure. The required degree of users' involvement is dependent on the nature of the system built and also on the selected SW development model. For example Agile model and prototyping require more intense user involvement than traditional approaches.

Users are typically involved in all stages of SDLC except in the product building stage. They are especially needed as key members in **system analysis and system design process** (see chapter [6.5](#) 'The role of information system methodologies and techniques in the phase of patient registry creation' where some modelling techniques are presented), in **testing** (user tests should be performed e.g. acceptance test) and especially in the deployment phase where **training** of the end users is very important.

7.2.2 Software testing

"Software testing is a process of analysing a software item to detect the differences between existing and required conditions (that is defects/errors/bugs) and to evaluate the features of the software item" (ANSI/IEEE 1059 standard).

Testing is executing a SW system or its components with the intent to assess if SW satisfies agreed and specified requirements / functionalities or not.

In the process of testing usually software tester, software developer, project lead/manager and end user are involved (4). In SDLC testing can be started from the requirements gathering phase and lasts till the deployment of the software. However it also depends on the development model (see chapter [7.2](#) 'Development of Registry Information System') that is being used.

Testing is performed in different forms like for example during a *Requirements gathering phase* where analysis and verification of requirements is also considered as testing or code testing executed by the developer (Unit type testing).

Proper testing is undoubtedly a very important task in SW development. Therefore a lot of standards dealing with SW testing and quality assurance are used by SW developers (e.g. ISO).

Two testing types can be distinguished: **manual testing** where a system is tested manually without using any automated tools and **automation testing** where system is tested using special tools (for

example scripts or another SW). Automation testing is used to re-run the test scenarios that were performed manually, quickly and repeatedly.

There exist different methods which can be used for SW testing: black box testing, white box testing, grey box testing.

There are two main levels of SW testing: **functional testing** and **non- functional testing**.

Functional testing assesses functionalities of the system. Examples of functional testing are:

- Unit testing – testing functional requirements of a unit;
- Integration testing – testing if different components work together correctly;
- System testing – testing the system as a whole in an environment close to the production environment;
- Regression testing – if any change is made to SW, then we have to test SW again, and
- **Acceptance testing** – testing the SW system in production environment by end user, this testing is also a legal and contractual requirement for acceptance of the system.

Non- functional testing:

- Performance testing – testing speed, capacity, stability, scalability; load testing, stress testing;
- Usability testing – testing efficiency of use, learnability, memorability etc.;
- Security testing – testing security and vulnerability of the system (confidentiality, authentication etc.);
- Portability testing – testing SW when it is moved in another environment (another computer with operating system).

All the tests should be well documented. Documentation usually includes: Test Plan, Test Scenario, Test Case and Traceability Matrix.

Further reading on SW testing: Software testing tutorial from tutorialspoint.com
http://actoolkit.unprme.org/wp-content/resourcepdf/software_testing.pdf

7.2.3 Training

Proper user training is probably one of the most important aspects of successfully rolling out a new SW, but is often the most poorly executed task. The training has to be planned in advance, tailored to the audience's needs, and executed by lectures with knowledge on adult learning theory and experience in adult lecturing (not just IT experts). Nowadays training could be executed also online as distance learning modules, when the audience is from a different location or it is difficult to ensemble users in one place at the same time.

A training program related to PR should be a combination of learning PR content and PR SW usage. When training program is being prepared the following questions have to be asked (5):

- Who is the audience?
- What are the learning objectives?
- What are the best mechanisms for disseminating the information?
- What is the best approach to ensure that learning has occurred?

7.3 Different options to obtain the Registry system

Information systems for PR can be built **in-house**, **outsourcing of** the development of PR application can take place or some **product packages can be bought** (this option includes also buying SW as a service) from PR, which has usually then to be tailored to in-house needs or by an external partner. All of the possibilities have their pros and cons. The differences will be described below.

The first step is to determine if there are any viable products on the market that will meet business needs and it can be bought as an **off the shelf SW package**. If so, a careful analysis of all identified off-the-shelf products should be made. The features, functions, benefits and costs of the products have to be analysed. Usually the product will not fit PR requirements completely and it will require customization. In cost estimation the time and cost of this task must also be included. Another important cost factor is ongoing licensing and maintenance costs in the product lifecycle – it can make off-the shelf SW a very expensive option. The benefits of buying a SW package are: lower initial costs, reduced time to deployment, higher success rates, availability of training support, access to user manuals and documentation.

Interesting reading about Electronic data Capture SW is available on www.ncbi.nlm.nih.gov/pmc/articles/PMC3049639.

PR can be built **in-house** when the organisation possesses enough internal technical capacity. The major benefits in building SW in-house are: organisation has overall control of the development process, the IT experts can be involved already in planning PR (see chapter 6.5), the organisation has a clear understanding on how the SW works, the future development is in control of organization. There are also some common challenges to build PR in-house (6): Unrealistic deadline, vague definition of project deliverables, inadequate time allotted for SW design, little or no testing, lack of quality assurance process, lack of proper project management, insufficient resources for ongoing maintenance and support, documentation that is overlooked or avoided.

Outsourcing the PR SW development is usually chosen when there is no in-house development staff or they do not possess enough technical capacity. Outsourcing means that there is a dependence on an external company to complete SW for PR. Therefore a strong business relationship will be critical and effective communication between the organizations will be a critical success factor. To select a proper outsourcing SW company focus should be placed on its (7): experience, approach, infrastructure, quality, reputation, stability, culture.

The major benefits of outsourcing are: reduced project and financial risks, clearly defined requirements and deliverables, use of the most up-to-date design capabilities, reduced project timeline and budget and SW is easy to maintain and enhance.

Quick guide on Outsourcing SW projects guideline can be found on www.bhmi.com/pdf/Outsourcing%20Guidelines.pdf.

References

1. Catalogue of National Health Information Sources in Ireland, HIQA, July 2010
<http://www.hiqa.ie/system/files/HI-Catalogue.pdf> [accessed 5th June 2014]
2. Laudon, C. Kenneth in Jane P. Laudon. 2012. Management in Information Systems – Managing the Digital Firm. 12th ed. New Jersey: Pearson Prentice Ltd.
3. Software Development Life Cycle Tutorial, tutorialspoint.com,
<http://www.tutorialspoint.com/sdlc/index.htm> [accessed 5th June 2014]
4. Software Testing Tutorial, tutorialspoint.com, http://actoolkit.unprme.org/wp-content/resourcepdf/software_testing.pdf [accessed 22nd October 2014]
5. Shinder, Deb, Plan your end-user training strategy before software roll-out, March 2006
<http://www.techrepublic.com/article/plan-your-end-user-training-strategy-before-software-roll-out/> [accessed 5th June 2014]
6. Buy, Build or BHMI, BHMI; http://www.bhmi.com/buy_build_or_bhmi.html [accessed 5th June 2014]
7. Outsourcing guidelines, Best practices for outsourcing SW development projects, BHMI,
<http://www.bhmi.com/pdf/Outsourcing%20Guidelines.pdf> [accessed 5th June 2014]

8 RUNNING A REGISTRY

Óscar Zurriaga, Carmen López Briones, Miguel A. Martínez-Beneito, Clara Cervero-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

Running a registry is not a simple procedure. It requires technical knowledge, scientific aptitudes, and a rigorous execution of the previous plan. A multitude of aspects have to be considered. Sequential and overarching processes have to follow for running a patient registry.

Key principles:

- The way of collecting data and the case report form (CRF) are crucial. Electronic based methods are preferable.
- A plan to review each data source must be established and the processes to control and cleaning data do then have to be systematized.
- Storing data regards to technical and legal aspects, especially for cross-border use (security, access permission, anonymization of personal data stored).
- A data analysis plan has to exist and to be executed, considering the characteristics of the registry data.
- The process of data dissemination has to be considered thinking in all the interested public and stakeholders.
- What is to be measured and controlled has to be defined in order to assure and to assess the data quality.
- The structure (steering committee, scientific advisory board) and the establishment of responsibilities, duties, roles of the people in charge of the registry have to be established for the registry governance.
- A plan for audit (internal or external) is necessary to validate all the processes.
- A registry is always in a continuous process of actualization. The development of the registry has to be continuously and periodically tested.
- Technical problems have to be considered regarding to the information system management: access management, security, backing-up, archiving.

8.1 Sequential Processes

8.1.1 Collecting data

Data collection is defined as the ongoing systematic collection, analysis, and interpretation of health data necessary for the patient registry. Data collection can be considered as regards two major domains; data source and data provider (see chapter [6.2.4](#) 'Data collection procedure').

The AMIA (the American Medical Informatics Association) has summarized the “Guiding principles for clinical data capture and documentation” that can be used to orient the implementation for collecting clinical data in a registry.

8.1.1.1 Modes of Data Collection

The way of collecting data for a registry is a crucial part, because it determines its feasibility. Regarding the data sources there are two main sources: paper based and electronic.

In the past, paper based models were predominant but nowadays the electronic based methods are the main. However, paper can still play a core role in a registry.

Different paper based methods are listed and discussed in chapter [6.1.2.1.3](#). Their important characteristic is that they are inexpensive and easy to create and develop, but in the registry's whole process they imply a substantial cost because they need to be recorded in an electronic way and there is no easy and cheap way to do that. The existing paper based processes are being adapted to an electronic environment, with the risk that the paradigm for electronic data capture would be determined by the historical model of paper based documentation.

The electronic based methods are the present and most probably the future ones (but almost half of the EU registries are still based on paper-and-pen mode). Electronic based methods can be computer based or mobile devices based (smart phones or tablets), but the main focus has to be that the data captured would be accurate, relevant, confidential, reliable, valid, and complete. Sometimes, the electronic based methods are focused on integrating several clinical data sources and to produce a new electronic form with the outcomes of the integration (see chapter [8.1.2](#) 'Data Linkage').

In the past, traditionally, a distinction was made between “passive” collection of data and “active” methods, and the difference was that the passive way is based on the notification and in the active one based on the personnel of the registry visiting the various sources to identify and collect the data. Nowadays the registries use a mixture of methods.

8.1.1.2 Case Report Form

A case report form (or CRF) is a paper or electronic questionnaire on which individual patient data, required by the registry, are recorded. The terminology is widely used in clinical trial research.

The CRF must include the common data elements planned in the design phase and it has to use standard definitions of items and variables (according to international recommendations).

The principles of a good CRF are: easy and friendly use, standard based, short, understandable and connected (if it is possible) with other potential sources.

Obviously the CRF paper based are less flexible and usable. The electronic CRF allows a higher functionality: data entry control, coherence validity, automatic error correcting system or help to the user.

An example of data to be included in a CRF can be accessed in the book “Cancer Registration: Principles and Methods” (Available from www.iarc.fr/en/publications/pdfs-online/epi/sp95/). The EPIRARE project has worked to identify the common data elements for rare diseases registries across Europe and a questionnaire about it can be accessed in www.epirare.eu/del.html.

8.1.1.3 Data entry/import

The data flow in a registry may include either the data entry (both paper or electronic based) or the capture, or it may import patients’ data from clinical databases.

In both cases it is important to establish the next items:

- Who will enter the data?
- Does the data entry program allow certain data items to be entered automatically, or is the data recorder able to make any changes?
- Does the data entry program effectively validate the data?

Paper based:

If the CRF is paper based, a direct data entry can be used. A computer keyboard is used to enter data from the paper CRF into the registry database. It is the easiest way, however, it requires personnel specifically dedicated to record data. Another option is to capture the data from the paper CRF, by using a scanner as well as special software to extract the data from it. In this case, it requires specific CRF forms to avoid errors.

Electronic based:

The data entry can be carried out in a local computerized database, though usually this is an option only for localized registries with a few patients. It is more common to use central database servers using web based data entry forms. In this way the data entry for the registry can be shared in several places.

Mobile devices (smart phones, tablets) can be also used as data entry tools, and it is specially indicated when the registry personnel goes to the clinical source.

Finally, a registry can get the data directly from the clinical databases. In this case, the data are captured or imported and require a data linkage process (see chapter [8.1.2](#)) with specific decision algorithms.

8.1.1.4 Patient/Data Provider Recruitment and Retention

A patient registry does not search completeness as a main goal, however, it is important to get enough patients to reach its objectives. In this way, there is a need to develop a source study to know where the data about patients are and which type of data could be used. A plan to review each data source must be established (periodicity of review, type of data source, way to get the data, permissions needed...). Sometimes, it will be necessary to contact the patients face to face and offer them the opportunity to include their data in the registry. An informed consent form has to be ready to use.

There are some incentives to recruit patients to the registry, but the most effective is the prestige and outcomes of the registry. If a registry is scientifically well considered, that patient will be more willing to participate. If there are some advantages, like the access to some specific health care processes or the increasing of the visibility of some diseases (especially important in the rare diseases field), patients will be willing to collaborate with the registry.

The transparency and the reputation of the registry are especially important: any problem regarding data protection vulnerability, for example, will imply the loss of patients' confidence and will entail problems for their recruitment and retention.

If the cases are regularly followed up, it will be possible to produce outcomes like remission or survival. For this reason, a registry has to prepare strategies to get the patients' status data regularly. It will be important to maintain an updated registry database with the date of each review. An active follow up process may be established by scanning different sources (mortality, treatment or drug prescriptions).

8.1.2 Data Linkage

The data linkage, or record linkage, process is referred to the task of identifying records in one or several datasets that correspond to the same individual or entity. This process may seem trivial if an identification code (ID) or a similar variable, unique for every entity, is available in the dataset(s) to be linked. Nevertheless, this setting is less usual in practice than could be expected or less usual than would be desirable.

Although it may seem obvious, it is worth mentioning the importance of a cleaning/purging phase on the dataset(s) of interest before proceeding to link them. This process should be done with particular attention to the variables used to link the databases. Dates in different formats or categorical variables with different codifications, such as {Male, Female} and {Man, Woman} for sex, are simple examples where this kind of issue may produce record linkage methods to fail dramatically. The pre-processing phase will also have to pay attention to string variables where different naming variants or nicknames could be used, such as Jim and James, and unify those variants into a single term.

When dealing with just a single dataset, it is very frequent that in case of having an ID variable included in the database, this is empty for a considerable number of registries. This is particularly frequent in health care registries, where sometimes urgent attention is required and not having access to the ID of the patient is not enough reason to deny the attention requested. This problem would be particularly prevalent in foreigners who do not have an ID of the corresponding health system because either they require attention during a temporary visit to that country or they are still in the process of getting their ID. In that case the ID corresponding to that record is forced to remain empty, with the problems that it may cause for identifying records corresponding to unique entities. A second problem when dealing with just a single dataset may come from records corresponding to children. In some health systems children do not have their own ID and they are recorded either with a missing ID or with the ID of one of their parents. This may cause some records corresponding to children to be linked to records of some of their parents (sometimes to one of them and sometimes to the other) altering the results of analyses that could be subsequently made from that dataset. This may be particularly frequent in new born babies, where administrative delays, as regards getting an ID, may make this setting as the general rule for this collective. Lastly, there is a possibility that the ID code of some of the records in the database were wrongly introduced due to typeset errors or to some other reasons.

All these circumstances will make naïve record linkage methods fail and will make the use of more sophisticated methods necessary.

When dealing with more than one dataset this problem is even worse. Besides the already mentioned problems which will also be present in this case, the record linkage of two or more databases has some new particularities that should also be borne in mind.

Special care should be taken to ensure that the linking fields of the databases are of exactly the same type and of the same length, since otherwise the linking process of the datasets could miss some records that should be matched. This is a particularly frequent setting when the databases to be linked come from different providers or institutions.

It is also a very common occurrence that the databases used in the linking process were not specifically devised to be linked and were designed for very different aims. Therefore, it is not rare to find that both databases do not share a common ID field that allows linking of their records. This is also a very frequent situation when linking databases of different administrations, such as the health and economic authorities, since the identification codes used for any of them are usually different. Specific record linkage methods have been developed for these settings making use of several fields in the database instead of just one.

Record linkage methods can be divided into two sets: deterministic and probabilistic methods. Deterministic methods are used when the databases to be linked lack a common ID field univocally identifying their individuals. However, if the datasets to be linked contain a set of variables whose combination could be an approximate ID, that combination could be used to link them. For example, the set of variables: name, surname, date of birth and city of residence, could be merged as a unique code univocally identifying any individual in the dataset. In that case, record linkage could be made attending to that code. Nevertheless, errors in the information recorded on these fields or simply because some of them contain missing values, would make this procedure fail to detect some matching records. To make deterministic record linkage methods more robust against these scenarios, it is usual to include as many fields as possible in the linking process, and match only those records where the percentage of matching fields is above some threshold.

The second set of record linkage methods are those relying on probabilistic decision rules. Thus, not every field in the deterministic methods, such as sex on one hand and date of birth on the other hand, has the same probability of containing two matching records. Probabilistic methods take into account those probabilities to decide if two records belong, or not, to the same entity. It is common in probabilistic methods to build, for every pair of records, a score summarizing the probability of observing as many matching fields as they have, and compare them with a fixed threshold that separates those scores resulting just from chance, from those coming from records of a common entity.

Data linkage can be done with two main purposes: merging the records of several datasets of different providers (e. g. hospitals) in a unique dataset, or enhancing the information of the records in a dataset with those fields coming from a second dataset. In the first case, record linkage will identify records in the different databases to be merged with those that correspond to the same individual. This will avoid accounting for those individuals more times than it should, making it possible to derive reliable rates that otherwise would not be reliable at all if repeated records were not excluded from the analysis. In the second case, inaccurate record linkage methods will make the resulting database into a riddle of

missing data coming from unlinked records, making posterior analyses of that database either unreliable, or more difficult to be done.

Data linkage is one of the most important topics regarding the anonymization legal aspect, because an ID is needed, which is an obvious piece of personal data. The individual right to integrity and protection of personal data has to be matched with the possibility of doing data linkage. There are several options to achieve it from a legal point of view, and currently a new regulation is in discussion in Europe. The perspectives of the new regulation in Europe are mentioned in chapter [6.1.4](#).

8.1.3 Controlling and Cleaning the Data

Data control and cleaning on patient registries involve the process by which erroneous data are removed or fixed and missing data are filled.

Three different phases in the cleaning process can be distinguished: screening, diagnosis and editing. All of them shall be applied not just as an independent step of the process, but also during the collection, linkage and analysis of the data.

The screening phase involves any action carried out to detect anomalies in the data. Several types of oddities can be found when screening data and each of them should be taken into account.

- Lack of data can be disguised when data sources use internal codes to declare a missing value, like filling a date field with '99/99/9999' or even literals like 'missing' or 'unknown'. A chart of these internal codes must be built and used as a filter.
- Duplicates can be detected by a redundant identification code of the patient or by a match in other identification variables such as name, date of birth, sex or external identification codes. Algorithms of approximate matching can detect non exact duplicates.
- Format incoherence shall be scanned, detecting values that are incompatible with the preset format of the variable (if Sex is defined as 'M' or 'F', a field filled with 'Male' is erroneous, and shall be recoded).
- The nature of variables offer ranges of values that are improbable or impossible (Age must be a non- negative number and is unlikely to be greater than 100). Thresholds must be defined to screen inconsistent and outlier values.
- Joint distributions of variables present different and more restrictive improbable or impossible joint values, like some pathologies combined with sex or age (Sex='Female' and Disease='Testicle Cancer' are incompatible, though each value is coherent by itself). A particular case of this screening is the chronological coherence that dates and ages must have.

The diagnosis phase can classify each oddity detected as erroneous, correct or dubious. A 'hard cutoff' leaves outside logically or biologically impossible values of data that will automatically be classified as erroneous. Improbable but not impossible values are filtered through 'soft cutoffs', and declared as dubious. They should be crossed with external databases (like censuses or other registries) or checked with the primary sources.

Modification of the database in the editing phase can be done automatically or manually over erroneous data. Redundant data shall be merged or deleted. Erroneous values can be corrected or deleted. Linking more external databases provides a source to fill or correct missing or erroneous values. Special codes or flag variables can be set to distinguish corrected fields.

Proper documentation and transparency is required for good practice in data management. Procedures, criteria and actual modifications shall be documented. A good way to keep track of the modifications is to record in a different database the original entries of data before modification.

A cleaning process can provide feedback to collection and linkage processes, so that future errors are prevented. It is important to encourage data users to report any anomalies they may find in the data, to improve the controlling and cleaning process.

8.1.4 Storing Data

Storing and retrieval of data are among the IT services giving support to registry operations. In addition to the general considerations about running these types of services, some specific remarks are worth mentioning here (Refer to [8.2.7 Information System Management](#) to complete the picture).

Data privacy is a major concern in European countries. At the time of writing (August, 2015) the legal framework of reference in this subject is still the Directive 95/46/EC, on the protection of individuals with regard to the processing of personal data and on the free movement of such data, and their different national implementations. In 2012 the European Commission announced a reform of this legal framework. After a lot of work and discussion, that reform is about to be completed.

Personal data about health are among the most sensitive issues. Accordingly, ethics, good corporate governance (transparency, responsibility, accountability, due diligence...) and regulations pose important restrictions on the processing and free movement of these data. Some restrictions have a direct impact on data storage and retrieval. Fines for noncompliance with regulatory requirements may be very important.

Access control (before). Procedures for proper user identification and authentication, as well as for granting and revoking access privileges have to be established. This also includes technical staff.

Access control (after). Logging procedures must keep track of every single access, even if it is only an attempt. Access logs must be kept safely, as they may become evidence, and be periodically examined. Any irregular event must be further investigated.

Data input/output. Any data input/output operation involving systems or facilities not under direct control of the registry owner must be previously approved and then recorded. Once again, these records must be periodically examined. These operations range from copying data to external devices to provide some sort of mobility, to data exports (or backups) to external facilities in order to provide data or operations recovery in case of disaster.

Cloud storage. Even when IT services, based on cloud computing, look interesting, they might be not appropriate at all. The registry owner and any potential provider of IT services (cloud based or not) must previously sign a detailed agreement. The following parts must be present in this document (among others):

- the provider has to declare and assure his knowledge, will, and ability to fulfill all requirements posed by the aforementioned legal framework;
- what the service provider has to do, what it is not allowed to, and what it must do when the engagement with the registry comes to an end;

- the procedures or evidences available to the registry owner to reassure it that the service provider is running everything according to the terms of the agreement.

Many cloud services are provided outside of the EU, where the legal framework mentioned above cannot be enforced (See also the Safe Harbor framework developed by the U.S. Department of Commerce). Besides, most big providers of cloud services have their own set of terms of service and operate on the basis of “take it or leave it”. Any of these two handicaps may be determinant to discard a provider.

Data integrity and availability. Power shortages, disk crashes, roof leaks, floods, fires, human errors... These things happen. Whether it is acceptable that they have an impact on the registry operations (or rather how much impact can be acceptable) is something to be determined by the registry owner, who will have to enable adequate countermeasures. Backup procedures should be conducted according to data recovery objectives and business continuity plans. The ability to recover from the backups is not something to take for granted, but to be tested on a regular basis.

Anonymization. For those purposes (i.e. research) where patient identity is not of primary relevance, dissociation of health data from identity data must be done. Privacy restrictions do not apply to data that cannot be traced back to the identity of the patient. Therefore, adequate dissociation processes should be made available as an option for data retrieval. These processes may be either one way dissociation (=anonymization *sensu stricto*), or two way dissociation (=reversible dissociation). The difference is that, in the former one, it is virtually impossible to trace back to the identity of the patient. In the case of reversible dissociation, the keys and procedures to unveil patient identity must be kept under strict control.

8.1.5 Analysis of Registry Data

The analysis of registry data presents as much variety as can be found in the purpose and objectives of registries. 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. Situations that call for unplanned analysis will often arise under two different circumstances: first, to address unexpected findings that can lead to new research questions, and second, to give answers to special requests set up by stakeholders. A planned analysis meets researchers’ objectives, whereas the foundation of a study based on unexpected findings is developed after making the observation; on the other hand, ad hoc analyses are directed to satisfy a registry user’s specific needs.

Closely linked to the data analysis plan, statistical methods should be stated in as much detail as possible. Researchers need to be cautious when interpreting registry data, which often has inherent biases. Potential sources of bias should be addressed in advance and, to the extent that it is possible, also the procedures for handling missing data and controlling any confounding.

8.1.5.1 Data Analysis Plan

The data analysis plan depends on the registry objectives, but registry planners should be aware that some relevant research questions could arise over time and may not be defined a priori.

Registry-based studies can be descriptive or analytical, but most of the times registries have aims that are primarily descriptive. Descriptive studies focus on disease frequency, distribution patterns (by

examining the person, place, and time in relationship to health events), clinical features of patients and natural history of diseases; descriptive studies can suggest risk factors and can help to generate all kinds of hypotheses that could be later tested by analytical studies.

In the case of rare diseases, patient registries are often a first step to try to understand the number of people affected and the characteristics of the disease and the patients, though the scope of these registries may evolve over time.

Disease-specific health indicators (morbidity, mortality and disability indicators) should be made available for the total studied population and for age and sex subgroups. Absolute numbers, as well as crude and age-standardised rates should be calculated. To ensure comparability, standardization should be based on the European standard population⁶².

The main measures of disease frequency are: incidence rate, cumulative incidence, point prevalence, period prevalence, lifetime prevalence and (for congenital diseases) prevalence at birth.

Incidence, often considered the most important measure in epidemiology, is usually expressed as incidence rate, which provides a measure of the occurrence of new disease cases per person-time unit; when incidence rate refers to one year, the denominator is the number of persons under surveillance. High mortality rate diseases, such as some cancers, are better measured in terms of incidence.

Point prevalence can be practically defined as the proportion of the population that has any given disease at some specific point in time, while period prevalence is the probability that an individual in a population will be a case, anytime during a given period of duration, often one year. Prevalence indicators are crucial in rare diseases, as prevalence itself constitutes the main criterion to define a disease as rare.

Mortality indicators, such as mortality rate and case fatality rate, provide a good measure of the burden of disease. Other health status indicators include premature mortality, measured by Years of Potential Life Lost (YPLL); disability-adjusted life year (DALY), a time-based measure that combines years of life lost due to premature mortality and years of life lost due to disability; and quality-adjusted life year (QALY), based on the number of quality years of life that would be added by an intervention. Analytical studies, such as cohort studies and case-control studies, focus on examining causal associations between exposures and outcomes, or between characteristics of patients and treatment, and health outcomes of interest. Data quality requirements in analytical studies are much higher than in descriptive studies.

For analytical studies, the association between a risk factor and outcome may be expressed as attributable risk, relative risk, odds ratio, or hazard ratio, depending on the nature of the data collected, the duration of the study, and the frequency of the outcome. Attributable risk is defined as the proportion of disease incidence that can be attributed to a specific exposure, and it may be used to indicate the impact of a particular exposure at a population level.

For economic analysis, although not very common in registry-based studies, the analytic approaches encountered are cost-effectiveness analysis and cost-utility studies.

⁶² Available at: <http://ec.europa.eu/eurostat/documents/3859598/5926869/KS-RA-13-028-EN.PDF/e713fa79-1add-44e8-b23d-5e8fa09b3f8f>. Also WHO's World Standard Population is defined; more on that can be found at: www.who.int/healthinfo/paper31.pdf.

8.1.5.2 Statistical Analysis

Statistical analysis is used to summarize and transform the data stored in the registries into knowledge. This knowledge is the ending result of the registries, since it allows us to know the population covered by the registry and, if appropriate, to compare it with the general population. Besides this aim, registries have just an accounting aim for performing an administrative control of the registered people.

It is not easy to summarize a particular set of statistical tools of particular use in health registries, since these are devised for very different purposes and, depending on them, some statistical tools or some others will be needed. The first set of statistical tools to be used in the analysis of health registries are descriptive tools. Descriptive tools summarize the, sometimes overwhelming, information stored in these registries. For this aim, graphical tools, either depicting the distribution of the values of a single variable or relating the values of two or more of them, are of particular use. Descriptive statistics are also often used for summarizing information in the databases, thus, the mean, median and standard deviation are typical statistics used to summarize variables. If instead, we are pursuing some measure measuring the amount of dependence between two variables in one dataset, Pearson's correlation coefficient is frequently the most widespread tool.

In addition to the descriptive aims above, it will often be interesting to make inference (learn) about some features of the population covered by the registry. In that case it would be firstly interesting to contrast some specific hypothesis in one's own dataset. In that case, one should resort to statistical tests. There is a huge amount of statistical tests available for very different purposes and it is not within the scope of this section to make even a brief description of their use. Nevertheless, it is convenient to highlight chi-square and t-test as the most common tests for making data analysis. The t-test is usually an appropriate choice for comparing the mean of two different groups in the population, although it requires the variable to be studied to be Normal-shaped. If this condition is not achieved, some alternative non-parametric test should be used, such as the Wilcoxon's signed rank test. On the other hand the chi-square test is used to assess dependency between two categorical variables.

Instead of testing some particular hypothesis in one's dataset, it would be interesting to assume a statistical model for one's dataset and to learn about the parameters ruling that model. Thus, as an example, a linear shaped relationship could be assumed between two variables and one could try to learn about the parameters defining that relationship. There are also several tools for achieving this goal. Thus, linear models (assuming a Normal outcome) are usually used for continuous variables, but if the outcome variable cannot be assumed to be Normal, Generalized Linear Models are the most usual tools to model this kind of settings. Logistic Regression and Poisson Regression models are just particular cases of Generalized Linear Models.

Survival analysis should also be mentioned as a statistical technique of particular use in health registries. Survival analysis is devised to study the time taken for an individual to develop an event of interest, such as the time survived before dying or developing a metastasis. The particularity of this kind of analysis is that many individuals in the dataset do not show the event of interest, maybe because they are not going to develop it in the future, because they have not developed it yet (although they will in the future) or because they have simply left out the study. This makes the variable of interest in these studies to be only partially known sometimes, and the analysis of this kind of data requires a particular treatment. If a more descriptive tool is wanted in this context Kaplan-Meier curves

are the most usual tools, meanwhile, if one prefers to model the effect of some covariates on the time of survival, usually Cox regression models are the most widespread tools.

Finally, it is convenient to mention some available tools for carrying out this statistical analysis. Although this list of statistical packages is not intended to be comprehensive, SAS, Stata or SPSS are highlighted as the commercial packages of more frequent use in the health sciences in general. Any of those packages could be perfectly suitable to carry out the above-mentioned analysis in the context of health registries. Nevertheless, R is nowadays an open source alternative with widespread use well beyond health science. R is usually blamed for being a bit rough for non-statistical users. Nevertheless, some specific R packages, like Rcmdr, are intended to make the use of R for non-statistical users easier, so that they make R available for a wide community of users. The main advantage of R is that it is likely to have specific state-of-the-art packages for hardly any task that could be desired in a registry, such as record linkage, dealing with “confounding by indication”, missing values, ...

There are lots of textbooks covering the statistical methods mentioned above; in fact, even specific monographs for any one of most of those methods have been published. Certainly the most appropriate book for any user will be that which illustrates their examples with the software habitually used for making the current statistical analysis. Thus, depending on the software used these could be the appropriate textbook choices: Le (2003) for SAS users, Cleophas and Zwinderman (2010) for SPSS users, Hills & De Stavola (2012) for Stata users and Lewis (2009) for R users. Once again this is not intended to be a comprehensive list of possibilities, but just a collection of useful textbooks.

8.1.5.3 Analytical considerations

When undertaking the analysis of the information stored in health registries there are a series of issues that deserve particular attention and that should be always borne in mind. Below are some of those issues to make the reader aware of their existence and their effects.

8.1.5.3.1 Potential sources of bias

There are numerous potential sources of bias when dealing with data-providing from health registries. Four specific sources of bias in observational studies in general are listed here: selection bias, non-response bias, information bias and recall bias.

One of these sources is selection bias, which is the result of the selection mechanism in the inclusion of individuals in one registry. Thus, as an example, assume that a diabetic patient registry is composed of patients recruited from their visits to a hospital. By definition, only those patients who have visited the hospital have the opportunity to be included in the registry. Regrettably patients visiting the hospital are not a random sample of the diabetic patients out of the whole number of patients but, on the contrary, they are patients with severe problems who have possibly had a complication related to his/her disease. This will make the results drawn from the registry to be non-representative of the diabetic patients within the whole population.

A related bias would be the non-response bias, in which all candidates to be included in the registry may have been previously recorded, however, some of them show missing values for some specific fields. These missing values can be rather innocuous if they are produced at random. Regrettably, quite often, the presence of missing values responds to a non-random mechanism, making those fields in

the database not being representative of the population and, therefore, biasing the results if this potential bias is not taken into account during the statistical analysis.

The second source of bias that it is thought convenient to mention is information bias. Information bias is the bias coming from inconsistencies in the way that information is introduced into the registry. Some artefact in the process of retrieving or coding the information into the database could make that information not reflect the reality, but rather a biased and distorted image of that reality. This could be the case of a variable reflecting the vaccination status of the individuals in the database. By default, this variable could be set to “non-vaccinated” and changed to “vaccinated” if appropriate. Nevertheless, as the default value goes always in the same sense, it can often happen that vaccinated individuals are registered in the database as non-vaccinated for the reason that the person who administered the vaccine did not record it into the database. This systematic bias could introduce problems and further bias into later analyses of the information in the database.

Finally, the recall bias should also be borne in mind when working with health registries, mainly when part or the entire database is retrieved from interviews or questionnaires. This source of bias is produced by differences in the accuracy of the information of the people included in the registry coming from their past. People who have family members with a history of cancer may be more prone to develop cancer than people without such connections. So, the information of both kinds of people could be systematically biased towards different directions, simply by their particular circumstances.

These biases are usually incorporated in the database from the very moment of introducing information. Registries professionals should be very aware of them, so that, even from the design phase of the registries they are prevented and, when possible, these biases are avoided by means of appropriate statistical analysis.

8.1.5.3.2 Confounding by indication

When analysing data coming from health registries it is quite common to study a variable as a function of some covariates. Nevertheless, the distribution of the values of the covariates in data coming from registries is not done at random or following a specific and controlled design. On the contrary, these values in observational studies in general, and health registries in particular, are the result of some factors not registered and out of the control of the study. For example, the decision of administering a medicine to a patient may be taken by a practitioner as the result of a general assessment of his/her health. As a consequence those patients with a worse general health status will take the medicine and those who are better will not. When assessing the effect of the medicine on a final outcome, such as dying in the following year, we could conclude that taking the medicine could increase the probability of dying, when this would be an effect of the previous health status of the patients. This effect is known as confounding by indication, and it may lead us to draw wrong conclusions on the effect of a variable because it is simply confused with other uncontrolled variable(s).

When interpreting the results of health-registry based analysis this potential problem should be very much borne in mind. If it is suspected that it could have had an influence on the estimation of the effect of a variable in the study, resort should be had to statistical techniques designed to control that effect, as for example, the inclusion of propensity scores into the analysis. Propensity scores will be auxiliary variables to be included into the analysis controlling the non-random mechanism that has generated the missing values in the dataset of the study.

8.1.5.3.3 Missing data

Health registries quite often contain missing data for some of their variables. These missing data are a real problem for data analysis and should be treated with care to avoid the potential bias introduction.

It is very convenient to know the reason why the missing data are produced. The best, although the less likely, scenario is that missing values occur at random. That is, no relationship can be found between their occurrence and any known variable. In that case, missing data are not very harmful, although they introduce some difficulties in data analysis. If the dataset at hand is large, those individuals containing missing values could be simply removed from the analysis and big changes in the new results would not be expected. If, on the contrary, there is knowledge or evidence from data, that missing data have not been produced at random, this should be borne in mind because they could be much more harmful in the data analysis phase. In that case, removing these individuals from the analysis would mean removing a particular part of the whole population, that could produce little or large biases depending on the degree of particularity of that sample. Therefore, in this case, a naïve removal of these individuals from the analysis does not seem to be an option. In this case imputing the missing values is the main option, although that imputation should be made taking into account the mechanism generating the missing data, e. g. if those individuals with particularly large (or low) values of a variable tend to show missing values in a second variable we should take this into account. If these two variables showed some correlation, the value of the first variable should be considered in order to impute the values in the second one, instead of doing it completely at random.

8.1.6 Data Dissemination

Well established, multicenter or population-based registries that have held large data collections can be a rich source of information with many different users, while small locally-held registries have a limited number of potential users, but in both cases data should be made accessible to ensure that all information is used to the maximum benefit of the population it serves.

Data should be disseminated in different ways, depending of the addressee of the data. Thus, three different points of view should be taken into account concerning registry data dissemination: 1) registry holders or owners, 2) patients and general public, and 3) decision makers and researchers.

Patients and service users, researchers, health professionals and policymakers, as well as other stakeholders and even the general public, should have access to valid and properly presented information in order to make choices and decisions. By making outcome data transparent to stakeholders, well-managed registries enable medical professionals to engage in continuous learning and to identify and share best clinical practices. To identify potential stakeholders, it is important to consider to whom the research questions matter. It is useful to identify these stakeholders at an early stage of the registry planning process, as they may have a key role in disseminating the results of the registry.

Registry-based information can be made available in many different ways, such as periodical reports, extracts on request and specific tools provided to allow users to access the data themselves via online portals. The principles of good dissemination of data have to be considered. An example would be the United Nations Good Practices on National Official Statistics⁶³.

⁶³ <http://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx>

Writing reports, presentations, tables, graphs or maps can be used to show the registry outcomes. Understanding is the main principle and it is very important to use the right type of tool for presenting the information. However, if a particular dissemination tool (represented, for example, by a table, graph or map) does not add to or support the analysis, it should be left out.

The dissemination reports should contain only information, only data or both data and information (data with a text explaining those data).

According to the addressee of the data, a good dissemination strategy (fixed in a dissemination plan) would have to consider the next features:

1. Registry holders or owners: dissemination requires actions to reinforce the acknowledgment of the people implied in the registry process, such as data providers, clinicians or managers
2. Patients and general public: it will be necessary to disseminate basic indicators, mainly in the form of basic tables, as they are more easily understood by them. On the other hand, graphics, maps and other sort of representations are also needed.
3. Decision makers and researchers: the dissemination has to be done in the form of aggregated data, but it is important to prepare individual anonymized data for researchers.

Every original finding and all scientifically significant information generated by disease registries should be communicated to the scientific community and finally rendered as scientific publications (on paper, online or both). Indeed publishing of results is inherently linked to the purpose of most if not all patient registries, as proper publishing can be considered an integral part of the scientific method.

Long-term population-based registries, an essential tool for public health surveillance, typically produce periodical descriptive analysis of data to be distributed to all potential users and especially to health professionals providing the data, as this feed-back enhances subsequent cooperation.

In clinical registries, data on disease progression or other long-term patient outcomes may not be available for many years, but safety data could be examined periodically over time. Studies based on patient registries, even short-term registries, may conduct intermediate analysis before all patients have been enrolled or all data collection has been completed, in order to document and monitor the progress of the project.

As the paradigm about health information goes, registry data should be collected once and used many times. Timeliness will be the key.

8.2 Overarching Processes

8.2.1 Data Quality Assurance

Electronic Health Records are generally designed for their primary use. As a consequence, when their data are collected with secondary, reuse purposes, such as for the construction of research repositories, their Data Quality (DQ) may not be optimal. Research repositories generally count with higher levels of DQ as specific, mostly due to manual curation and data profiling processes. However, DQ problems are still present. These can lead to suboptimal research processes, or even to inaccurate or wrong hypotheses. With the purpose of ensuring the highest levels of quality, continuously improve DQ processes, and avoid further DQ problems, organizational DQ Assurance protocols should be established.

DQ Assurance protocols combine activities at different levels, from the design of the information system, the user training in DQ, to a continuous DQ control. To this end, many research and industrial DQ Assurance proposals have been related to the Total Quality Management Six Sigma process improvement methodology. Concretely, the DMAIC model can be used to improve the DQ and their related processes, involving the following cycle of steps: Define, Measure, Analyse, Improve and Control. Defining what to measure and how to do it is the basis for the DQ Assurance, being the initial steps to any DQ improvement. These steps, along with the DQ control, can be defined under a DQ Assessment framework.

8.2.2 Data Quality Assessment

The Data Quality (DQ) Assessment is managed according to DQ dimensions: attributes that represent a single aspect or construct of DQ. Dimensions can conform to data definitions or to user expectations. Thus, DQ Assessment concepts and methods can be defined according to specific domains or problems. A set of DQ dimensions can be established to assess the DQ of cross-border patient registries based on different studies (see chapter 4 'Quality dimensions of Registries').

There exist other dimensions which, rather than measured on data by themselves, can be measured on their related stakeholders. Data *Availability* refers to their degree of accessibility to users. Data *Security* refers to their degree of privacy and confidentiality. Finally, data *Reliability* refers to the degree of reputation and trust of the stakeholders and institutions involved in its acquisition.

DQ problems may affect single or combined variables within an individual patient registry, e.g. inconsistent combination of variable values. Otherwise, DQ problems may affect a sample composed of a set of registries, e.g., a biased sample mean. For that reason, according to the purpose, methods should be considered to be applicable to large-scale big data repositories.

To conclude, it is of upmost importance for the DQ Assessment to formally define what is to be measured and controlled according to the aforementioned dimensions. Based on that, strategies can be defined to correct or prevent DQ problems. DQ processes can be applied to off-line research datasets. However, continuously controlling (based on on-line methods or multi-site audits) DQ indicators within a DQ Assurance cycle, from which to obtain a feedback to improve processes, is a recommended strategy to continuously reduce the DQ problems and optimize resources.

8.2.3 Evaluation and Improvement of Registry Service

Quality assessment of a registry should be a continuous process integrated in the registry's running. The dimensions needed to measure it (completeness, validity, opportunity,...) are common to different type of registries, but methods and indicators are related to the type of registry. The population cancer registries are one of the most advanced examples.

Regarding their complexity and cost, some methods can be routinely implemented, while others - contrasting it with an independent series of cases, which is one of the most used methods to assess completeness - should only to be used in a sporadic way.

However, using an external audit for the registry is a good idea, though external audit and accreditation - used in the health sector since decades, and considered useful to promote high quality products and

services with efficacy and reliability - are less developed in the registry field, except in the United States of America.

For those registries in which the health administration is both the data supplier and data user (client), there is a need to incorporate in an active way the opinion of health planner and health management professionals.

An example could be the REDEPICAN (Latin America Network for Cancer Information Systems) Guide for the External Evaluation of Population-based Cancer Registries, used in several Spanish and Latin America cancer registries. It is a new tool inspired in the accreditation principles: voluntary process, standard and defined criteria, self-assessment, external verifying process, and independent organism report. The Guidelines assesses 7 dimensions (Structure, Procedures Manual, Registry Method, Comparability, Completeness, Validity, Outcome Dissemination, and Confidentiality and Ethical Aspects) through 68 criteria with three standard levels, allowing to assess the traditional indicators and procedures needed to make the necessary changes, in order to offer the maximum efficiency. The final score, and the criteria with a low score, identify problems to be solved in the registry with concrete objectives for improvement. An external audit with a homogeneous measurement tool is useful as the starting point to measure quality improvements and to compare between registries.

8.2.4 Governance

Patient registries' 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.

Good governance must include:

- Accomplishment of the normative (regional, national, international). In some countries prior approval for the operation of a registry, by professional or health authorities is needed. The support and approval of the institution in which the registry is located is fundamental. The ethics committee's approval is also needed.
- Principles in which the registry action is based. Some of them are: transparency, participation, accuracy, security and data protection.
- Operating rules definition. This is a document that specifies the rules, case definitions used, codes and classification used (assuring the semantic interoperability). All the operating procedures have to be elaborated and released to all the participants in the registry. The way in which the data of the registry may be accessed has to be clearly defined. A document for the consent and its procedures has to exist (see chapter [4.1](#) 'Governance').
- The structure of the governance board (and its role and responsibilities). According to the governance plan (see chapter [6.1.9](#) 'Governance, Oversight and Registry Teams') the governance board can be structured in several ways:
 - The prioritization is to have a project management team, and scientific committee and a quality assurance committee.
 - A scientific committee or expert group can be formed to guide the development of the registry and to ensure the scientific basis. Its role is as a consultant group.

- The project management team can be developed also as a steering committee. It has to ensure that the registry is running according to the principles and objectives marked and planned. Its composition has to be done taking into account the institution in which the registry is based, the organism which funds the registry, the professionals implied, the health authorities, the academic or scientific institutions related to the subject of the registry and the patients and their families affected. Its role is to assume the responsibility of the registry. The Chair of the steering committee assumes final responsibility.

An example of Registry Governance Document is the National Cancer Registry of Ireland (16).

8.2.5 Auditing

According to the Dictionary of Epidemiology, an audit is an examination or review that establishes the extent to which a condition, process or performance conforms to predetermined standards or criteria. In a registry, audits may be carried out on the quality of data or completeness of records. Depending on the purpose of the registry, several types of audit can be performed. The audit can assess: enrolment of eligible patients, data completeness, selection bias, or data quality. An example of quality assessment is shown in chapter 8.2.3. The audit can be conducted either on the whole set of data of the registry, or just for a selected (random or systematic) sample of patients, using sampling techniques.

For example, the Spanish National Rare Diseases Registry has performed an audit in a Spanish region to assess the validity of diagnosis of aplastic anaemia by the International Classification of Diseases codes in hospital discharge data and the mortality registry, in order to detect cases to be included in the rare diseases registry. After getting the data from both databases the patient medical records were reviewed to confirm true aplastic anaemia cases. Only 15% of the cases were confirmed⁶⁴.

The audit can be internal or external. Internal audit is carried out by the registry staff, using a concrete plan and specific indicators to assess the most significant sources of error as regards the purpose of the registry. External audit is performed by external personnel, in accordance with pre-established criteria.

8.2.6 Continuous Development

A registry is always in a continuous process of actualization. For example, the way of collecting data can experience changes due to technological innovations, organizational modifications or new legal rules.

For that reason, a registry should be flexible and adaptive in all the facets of the registry process:

- For paper-based registries, it is crucial to move on to electronic based ones.
- New data elements could be added (new treatments or new disease stage for example).
- Definitions can be modified according to improved knowledge.
- Revisions of the classification systems happen and the registry has to be ready to be adapted to new ones.

⁶⁴ Ruiz E, Ramalle-Gómara E, Quiñones C, Rabasa P, Pisón C. Validation of diagnosis of aplastic anaemia in La Rioja (Spain) by International Classification of Diseases codes for case ascertainment for the Spanish National Rare Diseases Registry. *Eur J Haematol*. 2014 Aug 18. doi: 10.1111/ejh.12432. [Epub ahead of print]

- It is necessary to foresee any legal modifications regarding to ethical and protection data rules, as the personal identification number can also change or may need to be encrypted.
- The technological innovations affect the way in which a registry operates.
- The methods of data quality processes should be adapted to the results achieved.
- The reports and the diffusion mechanisms need to be flexible, because new data users can be incorporated and the stakeholders' concerns may change.

The development of the registry has to be continuously and periodically tested, in order to progress and adapt to the potential changes.

All the modifications have to be done ensuring the quality and integrity of the data and planning the date of the beginning.

8.2.7 Information System Management

Running a registry requires dealing with a certain number of stakeholders (patients, providers, clients, partners, regulatory authorities...). Running a registry also takes a good deal of IT. A registry owner will therefore be interested in raising trust among the stakeholders, as well as in getting most value from his information systems. IT governance can provide both.

None of the IT activities within the registry should take place on an improvised, contingent or ad-hoc basis, but within an adequate governance and management framework⁶⁵. This is the best way to:

- Maintain high-quality information to support business decisions.
- Achieve strategic goals and realize business benefits through the effective and innovative use of IT.
- Achieve operational excellence through reliable, efficient application of technology.
- Maintain IT-related risk at an acceptable level.
- Optimize the cost of IT services and technology.
- Support compliance with relevant laws, regulations, contractual agreements and policies.
- Provide trust to all stakeholders.

Among IT activities there should be management processes related to deliver, service and support. In that area, the following processes have to be considered:

- Manage operations.
- Manage service requests and incidents.
- Manage problems.
- Manage continuity.
- Manage security services.
- Manage business process controls.

⁶⁵ ISACA's COBIT 5 is the most comprehensive business framework for the governance and management of enterprise IT. This framework provides with good practice and guidance from the knowledge and experience of a large (>115000 members) community of IT audit, security, risk and governance professionals worldwide. Alternatively, different sets of ISO standards address some of the main issues (e.g. ISO/IEC 38500, ISO/IEC 20000, ISO/IEC 27000 ...). More at www.isaca.org/cobit/pages/default.aspx and www.iso.org.

Other processes should be run to monitor, evaluate and assess performance and conformance, the system of internal controls, and compliance with external requirements. Key indicators are essential in this context, as they are a main source of knowledge and allow measuring variables like cost, risk, disruption, improvement, and others.

All of the above should discard/discourage anyone pretending to “take care of all IT stuff” around the registry without adequate knowledge or tools. Being proficient at making scrambled eggs at home does not qualify one as a chef.

References

1. National Health Information Management Group. Minimum Guidelines for Health Registers for Statistical and Research Purposes. Australian Institute of Health and Welfare. 2001.
2. Cruz-Correia RJ, Pereira Rodrigues P, Freitas A, Canario Almeida F, Chen R, Costa-Pereira A. Data Quality and Integration Issues in Electronic Health Records. In: Information Discovery On Electronic Health Records. V. Hristidis (ed.); 2010. p.55–96.
3. Cusack CM, Hripcsak G, Bloomrosen M, Rosenbloom ST, Weaver CA, Wright A, Vawdrey DK, Walker J, Mamykina L. The future state of clinical data capture and documentation: a report from AMIA's 2011 Policy Meeting J Am Med Inform Assoc. 2013;20(1):134-40.
4. ENERCA (European Network for Rare and Congenital Anaemias): <http://www.enerca.org>
5. EPIRARE (Deliverable 1.4 Statistical Analysis of the EPIRARE survey data): http://www.epirare.eu/download/D1.4_StatisticalAnalysisofRegistrySurveyData.pdf
6. EUROCISS (Cardiovascular Indicators Surveillance Set): http://ec.europa.eu/health/ph_projects/2000/monitoring/fp_monitoring_2000_frep_10_en.pdf
7. Gliklich R, Dreyer N, editors. Registries for Evaluating Patient Outcomes: A User's Guide. 3rd ed. Rockville, MD: Agency for Healthcare Research and Quality; 2012 (Draft released for peer review).
8. Health information and Quality Authority. Guiding Principles for National Health and Social Care Data Collections. Dublin: Health information and Quality Authority; 2013.
9. Karr AF, Sanil AP, Banks DL. Data quality: A statistical perspective. Statistical Methodology. 2006, 3, 137:173.
10. Larsson S, Lawyer P, Garellick G, Lindahl B, Lundström M. Use Of 13 Disease Registries In 5 Countries Demonstrates The Potential To Use Outcome Data To Improve Health Care's Value. Health Aff. 2012;31:1220-227.
11. Lee YW, Strong DM, Kahn BK, Wang RY. AIMQ: a methodology for information quality assessment. Information & Management. 2002. 40, 133-146.
12. MacLennan R (1991). Items of patient information which may be collected by registries. In: Jensen OM, Parkin DM, MacLennan R, Muir CS, Skeet RG, editors. *Cancer Registration: Principles and Methods*. Lyon: International Agency for Research on Cancer (IARC Scientific Publications, No. 95); pp. 43–63. Available from <http://www.iarc.fr/en/publications/pdfs-online/epi/sp95/index.php>.
13. McMurphy AJ, Murphy SN, MacFadden D, Weber G, Simons WW, Orechia J, et al. SHRINE: Enabling Nationally Scalable Multi-Site Disease Studies. PLoS ONE. 2013 03;8(3):e55811.
14. National Cancer Registry of Ireland: <http://www.ncri.ie/sites/ncri/files/documents/GovernanceFrameworkfortheNationalCancerRegistry24September2010.pdf>
15. Navarro C, Molina JA, Barrios E, Izarzugaza I, Loria D, Cueva P, Sánchez MJ, Chirlaque MD, Fernández L. [External evaluation of population-based cancer registries: the REDEPICAN Guide for Latin America]. Rev Panam Salud Publica. 2013 Nov;34(5):336-42
16. Newton J, Gardner S. Disease registers in England. Institute of Health Sciences. University of Oxford. 2002.
17. Posada de la Paz M, Villaverde-Hueso A, Alonso V, János S, Zurriaga O, Pollán M, Abaitua-Borda I. Rare Diseases Epidemiology Research. In: Posada de la Paz M, Groft S, editors. Rare Diseases Epidemiology. Heidelberg, London, New York: Springer Science+Business Media B.V.; 2010.
18. Richesson R, Vehik K. Patient Registries: Utility, Validity and Inference. In: Posada de la Paz M, Groft S, editors. Rare Diseases Epidemiology. Heidelberg, London, New York: Springer Science+Business Media B.V.; 2010.

19. Röthlin M. Management of Data Quality in Enterprise Resource Planning Systems. Josef Eul Verlag GmbH (ed.). 2010.
20. Sáez C, Martínez-Miranda J, Robles M, García-Gómez JM. Organizing data quality assessment of shifting biomedical data. *Stud Health Technol Inform*. 2012; 180:721–725.
21. Sáez C, Robles M, García-Gómez JM. Stability metrics for multi-source biomedical data based on simplicial projections from probability distribution distances. Under review. 2014a.
22. Sáez C, Robles M, García-Gómez JM. Probabilistic change detection and visualization methods for the assessment of temporal stability in biomedical data quality. Under review. 2014b.
23. Sáez C, Martínez-Miranda J, Robles M, García-Gómez JM. Organizing data quality assessment of shifting biomedical data. *Stud Health Technol Inform*. 2012; 180:721–725.
24. Sebastian-Coleman L. Measuring Data Quality for Ongoing Improvement: A Data Quality Assessment Framework. Morgan Kaufmann (ed.). 2013.
25. Van den Broeck J, Argeseanu Cunningham S, Eeckels R, Herbst K (2005) Data cleaning: Detecting, diagnosing, and editing data abnormalities. *PLoS Med* 2(10): e267.
26. Wang RY, Strong DM. Beyond accuracy: what data quality means to data consumers? *J Manage Inf Syst*. 1996;12(4):5–33.
27. Wang RY. A Product Perspective on Total Data Quality Managements. *Communications of the ACM*. 1998; 41(2):58–65.
28. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *J Am Med Inform Assoc*. 2013 Jan; 20(1):144–151.
29. Zurriaga O, Bosch A, García-Blasco MJ, Clèries M, Martínez-Benito MA, Vela E. [Methodological aspects of the registries for renal patients in replacement therapy]. *Nefrología*. 2000; 20 Suppl 5:S23-31.

9 CHANGING AND STOPPING REGISTRIES

Marcel Kralj

A registry sometimes needs to be modified in order to continue to successfully accomplish its goals. The activity of changing a registry should not be underestimated, there are number of points that have to be taken into account. Stopping a registry may not be seen as a crucial task, however there are still factors that contribute to the successful shutdown of a registry.

Key principles:

- A clear rationale for the registry change should exist. Modifications should be done only if essentially needed.
- The scope of the registry modification should be carefully determined and understood; the feasibility of a registry modification has to be assessed.
- It is necessary to carefully elaborate the consequences before implementing a registry modification.
- It is essential to try to retain the comparability over time as much as possible; too ground-breaking changes might hurt time series.
- All relevant stakeholders have to be informed about the registry change.
- It should be considered whether modification to a registry requires any training or other support for a successful application of a change.
- The effect of any modification should be monitored.
- Clear decision on stopping a registry should exist; it is recommended to set the criteria for a registry stop in advance.
- It has to be decided what will happen to the registry data when a registry is stopped (e.g. dissemination of results, archiving)
- When storing/archiving a registry data the retention period, security, costs, monitoring and evaluation of the archiving process should be taken into account.
- Final report and other documentation that supports future (re)use of registry data should be prepared.

9.1 Changing an existing registry

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

There are various reasons for a registry to undertake a 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, for example, are some of them.

In connection with this a registry can undergo many different modifications. Maybe a registry needs to change its general purpose, goals or outputs, change the mode of data collection (e.g. moving from paper-based data collection to electronic data collection), introduce a new technology (e.g. web-based data entry), modify the target population or cohort (e.g. geographical expansion, the expansion of the age range, additional excluding criteria), adapt the outcomes or exposures (e.g. inclusion of knee implants as an exposure in addition to hip implants), modify any of the data elements (e.g. removing a redundant data element, adapting the outdated one, or adding a new one), change the case report form (e.g. to develop a more user-friendly form that is less subject to human error), modify the data collection protocol (e.g. different time points for follow-up), improve data analysis or data dissemination (e.g. more appropriate analytical techniques, different graphical representation of data, or different frequency of the dissemination), adapt a registry team or governing board, change the funding source or find new stakeholders (e.g. a move to public-private partnership), improve the overarching processes (e.g. quality assessment, auditing) etc. Some minor changes in a registry can be implemented more easily and quickly, but modifying a registry can be also a complex task that requires more effort, time and money. Therefore, it is 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.

This chapter does not provide guidance for each and every change that can occur and be implemented within a registry, and does not cover in detail every step of the registry modification. The modification of existing registry is in many ways similar to a process of establishing a new one, the latter already being covered in other chapters. Thus, a reader is encouraged to read other parts of the guidelines for additional information. For example, when reconsidering legal and ethical obligations during the modification process a reader is encouraged to see chapter [5](#) 'General requirements for cross-border use of patient registries'; or when modifying data elements a reader should read the chapters [6.3](#) 'Registry dataset', and [6.5.6](#) 'Data modelling', on how to adequately develop data elements for a registry. However, as already mentioned, there are various points that should be considered/emphasized while planning and/or implementing the modification of a registry:

1. It is important that there is a **clear rationale** for the registry change since not every change is a good one. A registry team needs to understand why the change is necessary, what exactly needs to be changed, and what the change will bring. Thus, a clear purpose and goals of the change should be developed. If there is no solid reason for a registry change, it should not be implemented.

2. It is essential to know how major and complex the change will be. The fact that changing one element of a registry can lead to the changing of other elements here should be taken into account. For example, changing dataset can lead to the adaptation of case report form, the data collecting process, statistical analysis and data dissemination. When modifying a registry this can be a good opportunity to make some other changes that are necessary as well. Moreover, it is important to be aware that registries are not just the documentation centre, but a network of data providers, so any modification has to be implemented in all satellites/units. Therefore, a registry team needs to **carefully determine and understand the scope** of the registry modification. Furthermore, an **assessment of the feasibility** of registry modification is a crucial part. Costs, time, effort, skills, and other resources are essential factors to consider. It is necessary to be aware of potential limitations and risks (e.g. technical breakdown or incompatibility, delays) as well. As for creating a new registry, **good planning** contributes to the successful implementation of a registry modification/transition. Thus, it is recommended to develop a thorough and realistic action plan and strategy for a registry modification/transition. It is worth mentioning here that piloting and testing are activities that should not be underestimated.
3. Once it is clear what modification in a registry will be implemented, a registry needs to have a **team for transition**. In that case it is important to consider the skills and knowledge that members possess, and how the effort of the specific team member will be increased during the modification process, because a member will probably have to perform his or her regular work simultaneously. It is also important to establish continuous, honest and open communication between all members, because effective collaboration between them can identify some unexpected barriers or risks that can be suitably addressed during the planning phase. The **role of leadership** here cannot be overemphasized since, as in many other areas, it is one of the key factors for a successful implementation of the modification process (1, 2, 5).
4. The **elaboration of the consequences** before implementing a registry modification is a very important task and should be undertaken carefully. Knowing where and what differences will occur with the registry modification, and how this will affect the further registry operation, will help in making the right decisions during the registry modification. When thinking about these consequences a registry team should look at every step and part of the registry operation. Are the changes going to reduce the quality of data (e.g. greater number of errors, new biases, lower statistical power, incompatibility etc.), increase the burden on data providers or registry participants, cause delays in reporting of results, increase the operating costs etc. are just a few examples that need to be considered. Similarly, it is crucial to **monitor the effect** of any modification after the implementation phase. The evaluation of differences in data prior to and after a change would be an example.
5. The registry team should **use experiences that were acquired with the existing registry**. Which things worked well and which did not (bearing in mind every component of the registry operation) represents an important feedback that can be used as an advantage when undertaking a registry change.
6. The registry needs to develop a **good notification protocol** for informing key stakeholders about the registry change. If the stakeholders are not engaged in a decision-making process they certainly have to be adequately informed to understand the rationale for the change, and its benefits. This is especially true for the participating sites/data providers which must be kept informed also about the timeline and implications that registry adaptation will have on the users. For any additional clarifications a registry team should be available, knowing that the change can take people out of their comfort zone and raise their stress and anxiety levels (1,

- 3, 5). If it proves necessary (e.g. in case of the additional follow-up), patients have to be provided with information about the change as well.
7. It is crucial to **reconsider the ethical and legal requirements**. A registry holder needs to be aware of how the changes in a registry affect the privacy, confidentiality and data access. It is necessary to consider whether the modifications require a new (or first) review from the ethics committee, if the inclusion of informed consent or change in informed consent form is needed, or if the re-consenting is required. In case of changes in stakeholder composition a registry holder must also determine whether the previous stakeholders should have access to data and if so, to which one (1). It is important to look from the other point of view as well. A registry can be modified as a result of (new) legal requirements. It is therefore important that registry holders actively follow the potential changes in this area (e.g. regulation updates) and comply with them if necessary.
 8. When implementing changes to a registry dataset (e.g. removing redundant data element, adding a new category/permmissible value or modifying a whole value set, introducing a brand new data element, adapting data element's definition, changing a relationship between data elements etc.) a registry team should be aware that comparability over time (i.e. longitudinal comparability) can be a great advantage in obtaining new information and knowledge. Therefore, it is advisable to try to **retain the comparability over time** as much as possible. If a registry team is changing a value set/categories of a specific data element, a **mapping** between the old and new value sets usually needs to be done and a so-called conversion table designed to clearly show the link between the prior and new value set. It is important that the conversion table is accessible and understandable to every user. The mapping may be a lengthy and intensive process (e.g. problems with the equivalence of prior and new categories) which needs to involve well qualified personnel. Certain changes may make it difficult to match a prior value set with the new value set which can have the result that missing ("unknown") data for subjects, on which data collection has already been done, can appear. In that case, these subjects can be reviewed/re-evaluated to update the missing value with the valid one. When this is not feasible it means that longitudinal comparability is not preserved. This is especially the case when significantly changing a definition on one of the key data elements, where the reality often is that everything must start again, meaning there is no comparability with the previous registry period, unless some well-established and validated conversions exist that enable making approximate comparisons.
 9. As a result of continuous development in technology, and also due to some other reasons (e.g. moving from one database vendor to another) a registry may go through the process of **data migration** which is a process of transferring data between storage types, formats, or computer systems (4). Data migration is a complex process that should be carefully managed as, due to its iterative nature, it can easily lead to schedule and cost overruns. First, data on the old system needs to be mapped to the new system. Next, data are extracted from the old system, and, at this point, thorough data cleansing is recommended. If there are any redundant data, they should be removed. When the data are loaded/imported into the new system a data validation needs to be performed to check whether data were accurately transferred causing no errors or data loss. As already mentioned, mapping, loading, and validation steps will probably need to be repeated several times (6-8). Last but not least, a registry holder must ensure that the data migration process complies with the legal requirements.
 10. **Appropriately documenting** the registry modification will allow registry users (and other stakeholders) to understand changes that have been implemented in the registry, provide insights into the history of changes and increase the transparency of the registry. Rationale for a registry change, a description of a change and its practical implications, were there any

unexpected problems and how they were solved, are there any other changes that need to be done in a future as a result of a recent change are important items that should be documented.

11. A registry holder should think about whether a modification to a registry requires any **training or other support** for a successful implementation/application of this change. Changing software for data entry or changing the analytical approach, for example, will probably require more comprehensive training than some other change, such as changing the data element's value set. A registry holder therefore needs to carefully consider how extensive the training should be, who has to be trained (e.g. data providers, registry's staff), what is the most appropriate way of training, and if any supporting material is needed.

9.2 Time to stop? - Stopping a registry

Partly also due to the fact that registries are often open-ended, the activity of stopping a registry does not seem so crucial, time and resource consuming as planning and setting up a registry. However, this does not mean that this activity should be neglected and that there are no important points that contribute to the correct and successful stopping of the registry.

When stopping a registry (with this we mean stopping data collection and ending all other sequential and overarching processes of a registry), first there must be a **clear decision on stopping** it. Setting the tangible and measurable goals/criteria for a registry stop in advance (in a registry planning phase) will help the registry holder to decide on whether the registry should continue with the operation, or if it is time to stop. Such criteria/goals might be, for example, to obtain a certain number of cases in the registry, achieve the desired precision of estimates and/or simply to fulfil the general purpose of the registry⁶⁶. However, the registry is not stopped only when certain goals are accomplished but it should be looked at from the opposite side as well - failure to meet a registry's predefined objectives or the fact that a registry appears to be unable to meet them in a reasonable time, poor operating results, loss of registry's relevance, lack of a purpose for the continuation, or other serious problems (e.g. discontinued funding, lack of personnel, poor data quality, low patient accrual or significant withdrawal of the registry's participants, ethical issues) could also represent the rationale for ending a registry (1, 9).

When a registry holder together with other stakeholders involved in the decision-making process decides to stop the registry he or she should **establish the communication** with the data providers, and inform also registry users, personnel and, if necessary, any other stakeholders (e.g. patients that are enrolled in the registry and regulators) about the registry stop. It is important that the key stakeholders understand the rationale for the registry stop and the consequences that this decision will bring. In case of regulatory issue that may arise when a registry is stopped before the regulatory question has been answered, the regulators might be even involved in discussions about stopping a registry, if applicable.

Furthermore, a registry team has to decide on **what will happen to the registry data**. Will the registry aggregate and disseminate the collected data (as a kind of final report) and/or will archive the data, meaning the data will continue to be available in the future? If preserving registry data brings important benefits (e.g. to have insights into the historical data; possibility to perform additional analyses and address the questions that were not covered in a prior registry's reports) then archiving

⁶⁶ Planning and consideration of the registry's anticipated size and duration is covered in subchapter [6.2.2](#)

might be the right decision. However, it is recommended that the decision about data archiving is discussed in a registry planning phase and not only when it comes to the registry stop.

When storing and archiving registry data a registry holder should take into account several points of data preservation:

- **Retention period** (how long the registry will retain the data, considering regulatory requirements if they exist)
- **Security** (following the norms of data protection and confidentiality of information a registry should establish policies and procedures to safeguard all data against loss, destruction, unauthorized use, or inappropriate alteration, and if necessary, also policies for proper and secure destruction of data. Some practical procedures for the above issues are authentication of system users, firewalls, back-ups, use of appropriate technology/storage media, policies that address copying data, disaster preparedness, emergency response, disaster recovery and training (10).)
- **Data for archiving** (in addition to the main registry data that are usually obtained by the case report form, the registry should preserve also a data entry log that tracks changes and users who made them, allowing registry to find the sources of the potential errors easier. To ensure that data can be correctly (re)used in the future, especially by others, data that are selected to be preserved must be packed with sufficient metadata. According to ICPSR (10) preservation metadata include all the information that is required by an organisation to preserve data, namely descriptive, structural, administrative and technical metadata.)
- **Monitoring and evaluation** (monitoring and assessment of the quality and effectiveness of the data maintaining/archiving process enables controlling of the process, finding out if everything is going according to the plan, whether any system errors are occurring, and enabling the adaptation or improvement of internal operations themselves.)
- **Costs** (data preservation requires financial, human and IT resources; registry holder should assess whether the funding is and will be available for the long-term maintenance of the registry data).

Finally, it is recommended that a registry prepares a **final report** in which its work, achievements, any encountered obstacles, rationale for the stop, and any implications for future work are clearly described. Along with the report, a registry should provide all the necessary documentation that supports the potential future (re)use of collected data.

References

1. Gliklich R, Dreyer N, Leavy M, eds. Registries for Evaluating Patient Outcomes: A User's Guide. 3rd edition. 2014. Two volumes. Available from: <http://www.effectivehealthcare.ahrq.gov/registries-guide-3.cfm>.
2. Newton R. Managing Change Step by Step: All You Need to Build a Plan and Make It Happen. Pearson Education, 2007.
3. Harding P., Pooley J. Resource Efficiency and Corporate Responsibility – Managing Change. 2004. available from <http://www.oursouthwest.com/SusBus/mggchange.pdf>.
4. Bal Gupta S., Mittal A. Introduction to Database Management System. Laxmi Publications, 2009.
5. Queensland Government. Change Management Best Practices Guide: Five key factors common to success in managing organisational change. Available from: <http://www.psc.qld.gov.au/publications/subject-specific-publications/assets/change-management-best-practice-guide.pdf>.
6. AHIMA. "Data Mapping Best Practices." *Journal of AHIMA* 82, no.4 (April 2011): 46-52. Available from: http://library.ahima.org/xpedio/groups/public/documents/ahima/bok1_048788.hcsp?dDocName=bok1_048788#Notes.
7. Computer Economics. Ensuring Success of Data Migration (April 2008). Available from: <http://www.computereconomics.com/article.cfm?id=1329>
8. SAS Institute Inc. Enhancing Your Chance for Successful Data Migration - Critical steps for creating data migration solutions that balance cost and rapid delivery. 2009. Available from: http://www.sas.com/resources/whitepaper/wp_5969.pdf
9. Rothman J. K., Haas J. When Should a Patient Registry End? Draft White Paper for AHRQ Patient Registries Handbook II. 2009. Available from: http://www.effectivehealthcare.ahrq.gov/repFiles/draftDocuments/2009_0817StoppingARegistry.pdf
10. Inter-university Consortium for Political and Social Research (ICPSR). 2009. "Principles and Good Practice for Preserving Data", International Household Survey Network, IHSN Working Paper No 003, December 2009. Available from: <http://www.ihsn.org/home/sites/default/files/resources/IHSN-WP003.pdf>
11. NISO. Understanding Metadata. Bethesda, MD: NISO Press; 2004. p. 1-20. Available from: www.niso.org/publications/press/UnderstandingMetadata.pdf.

10 RE-USE OF REGISTRY DATA

György Surján, Károly Fogarassy

Re-use of information means that some information collected for a given purpose is to be used for another one. Re-use of clinical data for registry (and other public health) purposes is usually and typically an abstraction process based on some sort of knowledge.

Key principles:

- There are several types of re-use of the data: internal re-use, international comparison (same purpose, different context), cross-registry comparison, comparison with information outside the health domain.
- Both aggregated and micro data can be re-used, but the first are much easier to apply.
- Cross-border use of data for public health is well-known and used for cross-country data comparison and surveillance, outbreak alerting and communication, bioterrorism threat, identification of best and cost-effective practices and public health research.
- There are many different applications in the field of cross-border use for research purposes (risk factor studies, genetic research, clinical and therapeutic research).
- The issues like compatibility, comparability and interoperability do need to be taken into account.
- Before planning to re-use of data, the legal background and various policies need to be studied carefully at EU level, national level and institutional level.

10.1 Background

Re-use of information in general is a current issue in informatics and for health informatics in particular. In 2012 the International Medical Informatics Association organised a summit in Brussels with the title “Trustworthy re-use of health data”. The title in itself points out that re-use of health data is a sensitive issue and it is important to find ways, where re-use can be done in a trusted manner. The conclusion of the summit has been published in the article referenced above. The participants considered various scenarios of re-use, with a focus on re-use of EHR data. In the following sections it is shown that re-use can be done at different levels, and all registries re-use clinical data in a certain sense, but on a higher level. Data stored in the registries can be re-used again for further purposes.

But first, in order to avoid confusion it is necessary to clearly define what is meant by re-use of registry data.

10.1.1 Definition of re-use

According to information theory, information is “something about something” i.e. a series of symbols that represents something else (1). For our purpose it is important to understand from this, that all information is only an abstraction of the thing (event or phenomenon) that is represented. No representation can completely describe the represented entity. Due to the abstraction, some features of reality are neglected, and only the relevant attributes of the real world entity are expressed. The best example for this is when we use identifiers to denote human beings. A “social security number” refers unanimously to a real person but nothing or only a very few attributes (e.g. gender, birthdate etc.) can be expressed by such a series of digits or characters.

As a consequence: **all reasonable representations are purpose dependent**. For a given purpose some features are relevant while others are not. The effective use of the information depends on appropriate selection of relevant features. Naturally, the relevancy depends on the purpose. A very good example of this is the different kinds of maps about the same territory. Maps for touristic purposes will be totally different, for example, from maps for public administration and these differences explain why a map created for some purpose is difficult or even impossible to use for another.

Re-use of information means cases where some information recorded for a given purpose is to be used for another one.

10.1.2 Re-use in the context of patient registries

The fact that all information is purpose dependent generates serious limitation of re-use, which of course does not mean that no information can be used for any other purpose but that for which it was originally recorded.

Sometimes there is a temptation for purposeless data collection: i.e. trying to store everything without defining the goals and future usage of data. As data acquisition and storage costs decrease, this temptation could become greater and greater. In the case of patient registries the privacy concerns prevent us from yielding to the temptation (moreover in most European countries legislation makes it impossible). But it is also important that purposeless data gathering is not a good way: it often leads to bad quality of the collected information.

Registries – often and preferably – are realisations of information re-use. Perhaps with the exception of registries created for public health purposes, it may be difficult to justify collecting data just for registry purposes, that are not relevant or not needed for clinical purposes (this is especially true for especially hospital-based registries). In this case the primary reason for storing some patient data is the clinical need, and **registries should store extracts and abstraction of clinical information**. This requirement will be addressed in section [10.4](#).

Summing up these considerations:

- 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 originally planned for

The next two sections provide brief answers to the emerging questions on why to re-use data and whether re-use is possible.

10.2 Why to re-use?

One could think, that if all data collection is purpose-dependent, then any reuse of collected data can be inappropriate. Sometimes it really is the case. For example, using ICD coded data in clinical context can be a misuse, since ICD coded data are neither sufficiently detailed nor reliable enough to directly serve care of individual patients. (The reproducibility of ICD codes is around 30%) This does not mean, that in certain cases such a solution cannot be helpful.

Theoretically, while all representations of reality (all data about something) are abstractions (some part of details is always lost) the remaining details still can convey many useful information that was not in the mind of the designers of the data collection. Beyond that in the most exciting cases of re-use data collected for a given specific purpose are merged and analysed with other data (see sections [10.5.3](#) and [10.5.4](#)) that always gives added value to our data.

Practically, in many health systems a vast amount of information is collected and poorly utilised. If re-use is possible it is more advantageous than separate data collections for all different purposes. Re-use is a much more cost effective and straightforward way.

10.3 Is re-use possible?

In spite of the above mentioned concerns or limitation in many cases it is possible, however care is always needed. For example, data, collected originally for health care reimbursement often can be used for quality assessment or capacity planning. But it is important to note that using some data for financial purposes always induces some distortion. Indeed, all observations distort somewhat the phenomenon that we want to observe. (It is a basic law in quantum physics, but also applies for many social phenomena). It is important to measure or at least estimate how large the distortion is, in order to draw correct conclusions from noisy data.

10.4 Re-use of data

10.4.1 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. The proper way is to extract all relevant data for a registry from the clinical documentation without much human workload. But this “extraction” is not always so straightforward. In Hungary there is a registry for premature new born babies, and this registry stores information on administration of surfactants. In the data model of the registry this is just a YES/NO rubric. Naturally, there is no such rubric in the patient records, but of course all drug administrations (including surfactants) are recorded. In order to automate the data submission to the registry, an abstraction process has to be implemented that is able to extract the information regarding which drugs are surfactants.

Therefore re-use of clinical data for registry (and other public health) purposes is usually and typically an abstraction process based on some sort of knowledge.

10.4.2 Re-use of spatial data

Using geographic data in different application domains has resulted in large amounts of data stored in spatial databases and these spatial data can be re-used for health purposes, sharing accurate geographic references to track communicable diseases by place and time, link various geo-referenced environmental factors such as air pollution, traffic, and built environment with geo-referenced health outcome data to analyse potential associations and identify risk factors. Such spatial data have been extensively used in the health domain in recent years. However, re-use of spatial data collected outside of the health domain has still an enormous potential for re-use related to the health domain.

10.5 Types of re-use of registry data

10.5.1 Internal re-use

Whenever an authority establishes a patient registry, the tasks, roles and goals of the registry are defined. The data-model of the registry is ideally designed based on these tasks. It may happen however that later the collected data are used for further purposes. For example, if the original task of a cancer registry is to measure cancer incidence, but later on the same data are used within the registry to estimate cancer prevalence, then this is a case of internal re-use. The term ‘internal’ refers to the fact that the re-use happens in the same organisation operating the registry. Sometimes such internal re-use requires additional data from different sources. In the mentioned example this could be cancer mortality data.

10.5.2 International comparison (same purpose, different context)

Patient registries for the same disease (same purpose) have been set up in different countries. Obviously, a cancer registry is the best example, since most of the countries operate some sort of cancer registry. Evidently there is a benefit in cross-country comparison of their data. Due to lack of standardisation it is often not so easy. This applies not only for the standardisation of their data structure, but also for the aim, scope and organisation of the registries. For example, data of

population-based registries are difficult to compare with hospital-based registries. Comparison of national (one single registry for the whole country) with country level data aggregated from regional registries may raise methodological problems.

10.5.3 Cross-registry comparison (correlation between diseases)

Morbidity patterns are evergreen research topics. Correlation between disease incidences either from a genetic or a geographic aspect is a subject of tremendous number of studies. Using patient registry data for this purpose can be done on individual or aggregated level.

Cross-registry comparison of registry data at individual level implies the possibility to merge data about the same person from different registries. However, this does not necessarily require the use of personal data. Such investigations can be performed also on pseudo-anonymous data as well. Different scenarios are possible. Consider two registries for two different diseases. If a comparison is to be made among them, the following options emerge:

- a. When two registries use personalised data based on the same identifier (e.g. social security number), to make a comparison without infringing privacy, one possibility is to have both registries remove the IDs from the records, and replace it with an artificial identifier, or pseudonym and merge them by this artificial identifier.
- b. Another option is to aggregate the data in the two registries separately and compare them at aggregate level. This method necessarily has some limitations.
- c. Datasets with common identifiers can be merged on a secure server with encrypted data transfers, and a de-identified dataset is generated on the server and provided back to researchers.

10.5.4 Comparison with information outside the health domain (e.g. environmental, economic, social etc. data)

With large amount of environmental, economic, social, spatial data generated and available in different databases and registries, these data can be linked with health data and secondary data analysis and comparisons are possible. Interactions of these factors could provide useful information for researchers, policy makers in both health and non-health domains.

10.6 Re-use of aggregations vs. re-use of elementary data

Patient registries typically store data about individual patients and create statistics from the individual data. Such statistical data can be used in many research or policy planning activities, and it can be integrated into other statistical data (e.g. comparing morbidity data with economical or social data etc.). Detailed studies, however, need to process the elementary data, when matching data from various sources is not possible on an aggregated level. Re-use of elementary (individual) data is, of course, much more sensitive and problematic from the privacy perspective. Therefore, it is absolutely important to understand the nature of various kinds of elementary and aggregated information.

10.7 Definition of Possible Types of Data

10.7.1 Aggregated Data (Indicator Compilation)

Data about a single entity (legal or natural person, institution, etc.) is called individual data. Data aggregation is a process where data and information is searched, gathered and presented in a report-based, summarised format that is meaningful and useful for the end user or application. In statistics, aggregated data denotes data combined from several measurements. When data are aggregated, groups of observations are replaced with summary statistics based on those observations. Data aggregation may be performed manually or through specialised software.

Aggregated data are usually calculated from individual data by summing or averaging values of some data-type attribute of a set of individuals (population). For example, “body weight of John Smith is 76 kg” is an individual piece of data. “The average body weight of adult citizens of London is 76 kg” is an aggregated data.

Health indicators such as community, public health, or occupational health indicators are typically aggregated data. Using aggregated data, various reports can be generated containing a compilation of selected indicators measuring health status, non-medical determinants of health, health system performance, and finally community and health system characteristics. Patient registries can serve as a valuable source for health indicators such as morbidity and mortality rates.

Aggregated data are generally considered harmless from a privacy perspective and hence can be used without any legal restriction in most cases, providing that appropriate data disclosure control techniques have been used. The normal way that most statistics work, is that a total amount of some phenomenon is counted and then divided along some attributes. For example, first, the total number of deaths is counted in a country then it is divided according to gender, age group, geography or cause of death. By combining of divisions along different attributes we often get very small numbers and run into the risk of possible identification of some individuals. For that reason in most countries legislation restricts the publication of aggregated data where there are less individuals than a certain limit behind each number. This limit varies typically between three and five.

It is reasonable, however, to make a distinction between publication (making data available to everybody, without any control of further use) and use of such kind of data, for example, for research purposes. In the latter case it is possible to control the proper use of data, for example, by supervision of an ethical committee.

An increasing number of global patient registries have been established in recent years, which could especially be valuable for rare health conditions to help biomedical research. One example for a global patient registry coming from the US National Institute of Health (NIH), National Center for Advancing Translational Sciences (NCATS):

“The goal of the NIH/NCATS Global Rare Diseases Patient Registry Data Repository (GRDR) program is to serve as a central web-based global data repository that aggregates coded patient information and clinical data to be available to investigators to conduct various biomedical studies, including clinical trials. The aim of the program is to advance research for many rare diseases and apply to common diseases as well.

Data are collected and aggregated from rare disease registries in a standardized manner, linking the registry data to Common Data Elements (CDEs) using nationally accepted standards and standard terminologies. The aim is that through standardization, registries will be interoperable to enable exchange and sharing of data. Each registry will be free to develop its own survey questions according to patient preference and the nature of the disease.”(2)

10.7.2 Anonymised Data

Anonymisation is a procedure to completely remove any information from the data that could lead to an individual being identified. Oxford Redcliff Hospitals Confidentiality Guidelines states: “[Anonymous] data concerning an individual from which the identity of the individual cannot be determined” (3).

A Bristol University ethical document defines the following anonymous data types:

“Anonymised data are data prepared from personal information but from which the person cannot be identified by the recipient of the information.

'Linked anonymised data' are anonymous to the people who receive and hold it (e.g. a research team) but contain information or codes that would allow the suppliers of the data, such as Social Services, to identify people from it.

'Unlinked anonymised data' contain no information that could reasonably be used by anyone to identify people. The link to individuals must be irreversibly broken. As a minimum, unlinked anonymised data must not contain any of the following, or codes traceable by you for the following (4):

- name, address, phone/fax number, email address, full postcode
- NHS number, any other identifying reference number
- photograph, names of relatives”

The main difference between anonymous and pseudo-anonymous data is that the former does not contain any key to merge or collect different data about the same individuals. Both data are individual, i.e. contain information about a single person. For example, if all personal identifiers are stripped out from a death certificate (name, birth date, home address, social security number etc.) it is still about a single individual. However, such a document cannot be merged anymore with other (either anonymous, pseudo-anonymous or personal) data about the same person.

Using fully anonymised data is relatively safe from a privacy perspective, however, if one is in possession of additional personal data that allows joining anonymous and personal data with reasonable effort, then privacy concerns emerge.

On the other hand, usability of anonymous data is limited if multiple recording and counting is possible. If there is any chance of having more than one record about the same individual, then calculations will be incorrect (e.g., if we have salary data without personal identifier and one person can have multiple employments, then average incomes cannot be calculated). This is the main reason to use pseudo-anonymised (pseudonymised) data.

10.7.3 Pseudo-anonymised Data

Generally speaking, pseudoanonymisation (or pseudonymisation) is a procedure to break the link to the data subject by replacing the most identifying fields within a data record by one or more artificial identifiers, or pseudonyms. Pseudonymisation is not a method of anonymisation. It merely reduces the linkability of a dataset with the original identity of a data subject, and is accordingly a useful security measure (29).

Pseudo-anonymised (or pseudonymised) data means that information is represented in a way that allows collecting all data corresponding to the same person without the possibility to identify the real person. Such data cannot include personal identifiers such as names and addresses of the person.

However, there is a disagreement regarding the interpretation of what ‘possibility’ means. For example, according to recent Hungarian legislation, the possibility of re-identification exists if the handler of the data is in the possession of the technical tools necessary to re-identify the person. There are much stronger interpretations in some European countries that say if there is any chance to re-identify (e.g. by using additional information) then the data should be treated as personal. Again, other regulation considers the effort necessary to recognize the real persons, saying that data should be treated as personal only if reasonable effort is enough to re-identify.

There are other definitions of pseudo-anonymisation. For example, the National Health Service (NHS) in UK uses the following definition:

“The technical process of replacing person identifiers in a dataset with other values (pseudonyms) available to the data user, from which the identities of individuals cannot be intrinsically inferred, for example replacing an NHS number with another random number, replacing a name with a code or replacing an address with a location code.”(5)

This definition interprets the possibility of re-identification again in another way. It says that the data are pseudonymous if the real individuals cannot be "intrinsically" inferred, i.e. just by using the data. If data need to be merged with any other (extrinsic) information in order to refer to real persons, then it is not personal data.

Independently from which definition is worth accepting, it is clear that the use of such data is extremely important and unavoidable for health research and evidence-based health policy.

On the other hand, it is clear that using such data requires special regulation. For example, current Hungarian legislation says that any data handled by governmental bodies are either public or personal. Pseudo-anonymous data are not mentioned in the legislation. Only the law of statistics mentions that statistical bodies must not publish data with less than three entities in any given cell presented. However, publication of data (i.e., making data available for everybody) and using data for research purposes are different.

A European directive on using pseudonymous data that defines this type of data and the conditions of use of them would be welcome.

10.7.4 Personal Data

Several laws and regulations exist around the world, which include a definition for personal data. Personal data are defined in EU directive 95/46/EC, for the purposes of the directive, as the following:

Article 2a: 'personal data' shall mean any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity (6).

Personal data, personally identifiable information may be categorised into two main groups:

- 1) Personal data, which are often used to identify the individual such as full name, home address, date of birth, birth place, national identification number, genetic information, telephone number, e-mail address, vehicle registration plate number, credit card numbers, biometric records, etc.
- 2) Personal data, which may be shared by many people and may identify the individual. Examples include city, county, state, country of residence, age, race/ethnicity, gender, salary, job position, etc.

However, it is important to keep in mind that sometimes multiple pieces of information, none sufficient by itself to uniquely identify an individual, may uniquely identify a person when combined.

Because a very rare disease itself could be personally identifiable information, collecting and publishing information about rare diseases in patient registries requires careful consideration.

10.8 Cross-border Use for Public Health

There are several initiatives and examples for cross-border use for both public health and research purposes of various data including patient registries' data. Sharing information, data exchange across the borders could serve several purposes.

10.8.1 Cross-country Data Comparison, Surveillance

Data exchange and information-sharing across borders would allow cross-country surveillance, monitoring, and comparison of data. For example, disease rates, trends could be compared by various demographic and clinical characteristic. EUROCAT, European Surveillance of Congenital Anomalies, which collects data on birth defects from several regional and national birth defects programs to generate trends, is a good example of that, as well as the European Network of Cancer Registries (ENCR), which collects and regularly disseminates information on incidence and mortality from cancer in the European Union and Europe. The European Surveillance System (TESSy) is a highly flexible metadata-driven system for collection, validation, cleaning, analysis and dissemination of data on communicable diseases. Its key aims are data analysis and production of outputs for public health action (7).

10.8.2 Outbreak Alerting and Communication

Sharing cross-border information on communicable (infectious) diseases has great significance on the EU level or international alerting of outbreaks and potential pandemics. Several infectious diseases spread from human to human and these do not respect country borders. Therefore, effectively tracking and preventing, or at least minimising the consequence of an outbreak, to the extent it is possible, prompt information sharing and data reporting is extremely important. An example of this is the novel H1N1 influenza virus outbreak in recent years. However, these emerging diseases are usually not related to or part of patients' registries. Nevertheless, this information may be linked to special patient's registries (such as vulnerable patient groups) that could help alerting them and also help a better understanding of the course and treatment of disease. In this highly globalised and mobile world, transmission of many diseases is more frequent and more possible than ever before in recorded history.

10.8.3 Bioterrorism Threat

Sharing data among specific patient registries could even be helpful in the case of a bioterrorism threat to inform and protect vulnerable patients and groups in a timely manner (e.g., patients with immune deficiencies). The anthrax threat and infections in the United States a few years ago showed the potential danger and need to set up harmonised reporting systems. Patient registries may also benefit from sharing information if a functional cross-border data exchange system was in place.

10.8.4 Identification of Best and Cost-effective Practices

Data sharing could help searching for and identifying best and cost-effective practices by health care providers such as timely diagnosis and treatment, professional recommendations. For example, identification of best practices for reducing hospital readmissions could lead to the implementation of such practice by other health care providers, which could lead to significant cost reduction, and reduce avoidable hospital readmissions.

10.8.5 Referral to Services, Establishing New Services

Mapping the distribution of patients by well-defined smaller geographical units could help to refer these patients to the available services on a European level. At the same time, lack of services in certain geographical areas can also be identified and a new service may be established. Taking into account travel time and distance is very important from both the service providers' and the patients' point of view. The less time and distance is needed to travel, the better, especially in urgent care, to save life and also costs.

10.8.6 Public Health Research

Data exchange could provide information for basic and applied research, and help also understand various demographic and clinical characteristics, long-term outcomes of specific diseases, comorbidities, and effective prevention and intervention efforts on a European or global level.

It is important to differentiate ad hoc, irregular cross-border data sharing, data communication, which could also have significant public health value, from public health surveillance, which is, by definition, an ongoing, systematic data collection in a timely manner.

10.9 Cross-border Use for Research Purposes

10.9.1 Issues

The use of registry data for public health and research purposes in cross-organisational and cross-border setting is becoming more and more important. For example:

- increasing mobility increases the risk of cross-country infections,
- for rare conditions setting up international databases or exchanging data is crucial to establish large enough cohorts to study a specific population or specific rare conditions such as genetic disorders, congenital malformations, and metabolic conditions.

Harmonisation of registry data could lead to a reduced cost of managing and using these data, and better quality data would be available for analyses and various indicators.

10.9.2 Risk Factor Studies

Registry data could provide valuable information for epidemiologic studies to analyse potential risk factors for diseases. Sociodemographic data such as race/ethnicity, gender, age can help understand whether there is an increased risk among certain groups of people. Data on environmental factors like air pollutants, agricultural activities such as pesticide exposure can be linked and associations can be analysed. Natural disasters, neighbourhood effects on health can also be studied. Data on medication/drug use and adverse outcomes could be valuable information for drug safety studies.

10.9.3 Genetic Research

Registry data may include information on genetic analysis (molecular or cytogenetic), or the registry data may be linked with bio banks, biological samples that allow further genetic analyses. Gene mutations may be identified for rare genetic conditions. Registries could potentially contribute also to gene-environment correlation studies. Several genetic research initiatives are going on in Europe and researchers look for data from different sources including patient registries.

10.9.4 Clinical and Therapeutic Research

Registry data could also help clinical research studies to look at treatment options, and may include data from clinical trials for new medications and medical devices. Using available data researchers can analyse clinical parameters, effectiveness, and outcomes. Inequalities and disparities in health outcomes by country or other factors could drive establishing new or improved clinical guidelines and recommendations, and inform policy makers.

10.9.5 Some additional information

During 2011-2015, major FP7 project “The Data without Boundaries – DwB” took place (www.dwbproject.org/). The project had a mission to support equal and easy access to the rich resources of official microdata for the European Research Area, within a structured framework where responsibilities and liability would be equally shared. During its four-year lifespan the DwB worked towards preparing a comprehensive European service with better and friendly metadata, a more harmonized transnational accreditation and a secure infrastructure that would allow transnational

access to the highly detailed and confidential microdata, both national and European, so that the European Union would be able to continuously produce cutting-edge research and reliable policy evaluations. Most of the results of DwB could be applied also to re-use of patient registries for research purposes.

Several important and relevant issues had been analysed and a few tools had been developed in the life-span of the DwB:

1. What are the researchers' ideas and expectation regarding the re-use of data for research purposes: the most important issues are search strategy, quick overview, good documentation, comparability, information about procedures, user generated context (see www.dwbproject.org/export/sites/default/promotion/dissemination_material/dwb_factsheet_user-requirements-def.pdf).
2. State-of-the-art of the remote access to data systems has been analysed and Database on National Accreditation & Data Access Conditions has been prepared (www.dwbproject.org/access/accreditation_db.html).
3. Analyses of legal frameworks for data re-use for research purposes have been performed and could be browsed via on-line visualization tool: <http://fryford.uk/wp-content/visuals/europe/european.html>, where possibilities to access data according to different types of data and types of access are presented.
4. Several software tools have been developed: Synthetic Data Tools, CTA (Controlled Tabular Adjustment), Enhanced Controlled Tabular Adjustment - ECTA - & Cell Suppression Free Open Solver software, and Record Linkage tool.

Many more results of the project could be found on their website. However, researchers who are planning to re-use registry data, even in cross-border setting, can find a lot of important information and tools.

10.10 Compatibility, comparability and interoperability

10.10.1 Data compatibility

The integration of multiple data sets from different sources requires that they be compatible. Methods used to create the data should be considered early in the process, to avoid problems later during attempts to integrate data sets.

"Compatibility is the capacity for two systems to work together without having to be altered to do so. Compatible software applications use the same data formats. For example, if word processor applications are compatible, the user should be able to open their document files in either product" (13).

"Another factor that should be considered is the compatibility of existing data sets. Frequently, a data search may reveal multiple sources of similar data types, but the metadata may reveal that the individual data sets are not compatible, as the data have not been collected in a consistent manner ..." (14).

For registries it means that data created by one registry can be imported into another, without manual data manipulation. Such a scenario is reasonable and necessary, for example, when in a country data

collection is carried out at regional level, and regional registry data are used to build up a national registry. Similarly if a European registry is built on Member State registries. Data compatibility is usually considered at technical level (same data structure and format, character coding etc.) as in the mentioned example with word processors. In the case of patient registries the issue is more complex however. If we want to compile a national registry from regional ones, this technical compatibility is a prerequisite only, but far from sufficient. Such compilation can be done on the level of elementary data (e.g. data of patients registered in each registry is to be sent to the national registry). But it also can be done at aggregated level, where only sums and (weighted) average numbers are sent. In both cases it is important to be sure, for example, that each patient is registered in one regional registry only, so double counting is excluded. It is also important, that there are no definitional or methodological differences among the regional registries, or at least there should be awareness of such differences.

Summing up, compatibility of registry data has the following requirements:

1. **Technical compatibility** of data (identical or convertible data structures, formats, coding schemes etc.)
2. **Comparability** (see section [10.10.2](#))
3. **Double counting exclusion** (see the problem of populations in section [10.12](#))

10.10.2 Comparability

Comparability is different from compatibility. Colloquially speaking, comparability means that one has to be sure to compare apples with apples and not peaches. Whenever data are compared from different registries it is important to be sure that the observed differences are attributable to real differences in the thing that is being measured, not some artefacts that are consequences of external or irrelevant circumstances. Full comparability occurs exceptionally, i.e. raw data of registries are hardly comparable.

The more common situation is that the differences that make raw data incomparable are known, and ways can be found to resolve them. The most typical example is standardised death rates. Raw mortality figures of different populations are practically never comparable due to the different age structure of different populations. By standardisation raw data can be projected onto a standard age distribution that enables a comparison of mortality data from very different countries.

In other cases comparability problems arise from different definitions and categorisations. Such entities like 'hospital', 'hospital bed', 'long term care', 'community care' are often interpreted differently, and data that are built on such entities are sometimes hard to compare. Contrary to the standardised death rate example, in such cases the problem cannot always be fully resolved. Sometimes relative comparability has to be accepted. For example, if it is known that 'number of hospital beds' in country A covers more kinds (e.g. new-born baby incubators included) than in country B, but even so country A has fewer hospital beds than country B, then it is certain that there is a real difference, but not in the reverse case.

The most important issue is to be aware of comparability issues. To achieve the possible optimum, the following conditions have to be met:

1. **Sufficiently detailed metadata should be available.** Metadata should describe what is counted in a registry, with what exceptions, how the measured entities are defined, what data collection methodology was applied etc.
2. **Additional data necessary for standardisation should be available.** If there are known external or irrelevant factors that influence the thing to be measured (e.g. as age distribution influences mortality) then these data must be available in order to eliminate these effects.

10.10.3 Interoperability

Interoperability has a huge literature and it is not the aim of this study to give a comprehensive overview of the various approaches, definitions and theories behind it. The various definitions often divide interoperability into different layers such as technical, functional, semantic etc. One of these divisions is described in detail in chapters [3](#) and [5](#), where legal, organisational/process, semantic and technical interoperability are considered.

In this section we restrict ourselves to technical (functional) and semantic aspects, because these are the levels of interoperability where IT standards can be used to find solutions. Briefly and generally speaking, interoperability from IT aspect is the ability of systems to work together (section [10.11](#) explains the technical aspects of interoperability in detail).

Semantic interoperability between registries implies that the recipient system is not only able to handle the received information but also able to automatically interpret it. It is possible that two registries that collect data for the same disease use different disease coding systems. Functional interoperability of such registries implies in such a case that the disease codes can be imported, but does not imply that the semantically identical codes are recognised or codes from one scheme is converted to the other one. (See the problem of mapping in subchapter [10.11.3.2](#)).

Semantic interoperability comes into question only if (at least one of) the systems are able to process information semantically: it makes inferences, or actions that depend entirely on the meaning of information, not on its syntax. Such semantic functions are hard to imagine without using some sort of ontology.

10.11 Interoperability Standards and Approaches for Data Exchange

10.11.1 General Concept

The concept of functional interoperability is to permit one system (sender) to transmit data to another system (receiver) to accomplish a specific communication in a precise and unambiguous manner. To achieve this, both systems have to know the format and content, and understand the terminology used. Using standard terminology can help database and system developers, and can facilitate exchange of data among various systems.

The recognition of the need to interconnect health related applications and exchange data led to the development and enforcement of interoperability standards. The following sections explain the standards used for structuring and encoding data.

10.11.2 eHealth standards

Exchanging and interchanging data in the health care domain in a seamless manner is becoming critically important. Lots of efforts have been made in this area to develop standards, which have obvious economic benefits as well. Here are a few examples of current standards developed and used for data exchange (see also chapter in [3.2.5.1](#)).

- **Health Level 7 (HL7):** HL7 and its members provide a framework (and related standards) for the exchange, integration, sharing, and retrieval of electronic health information. These standards define how information is packaged and communicated from one party to another, setting the language, structure and data types required for seamless integration between systems. HL7 standards were originally developed to exchange data among hospital computer systems. HL7 standards support clinical practice and the management, delivery, and evaluation of health services, and are recognized as the most commonly used in the world.
- **The National Council for Prescription Drug Programs:** The US National Council created data-interchange standards such as drug claims for the pharmacy services sector of the health care industry.
- **Data Interchange Standards for Bioinformatics:** These standards were developed to support data exchange among various databases in bioinformatics and have gained popularity.
- **Health Informatics Service Architecture:** The European Committee for Standardization (CEN) Standard Architecture for Healthcare Information Systems (ENV 12967), Health Informatics Service Architecture or HISA is a standard that provides guidance on the development of modular open information technology (IT) systems in the healthcare sector.
- **openEHR:** It is a virtual community working on interoperability and computability in e-health. Its main focus is electronic patient records (EHRs) and systems. The openEHR Foundation has published a set of specifications defining a health information reference model, a language for building 'clinical models', or archetypes, which are separate from the software, and a query language. The architecture is designed to make use of external health terminologies, such as SNOMED CT, LOINC and ICDx. Components and systems conforming to openEHR are 'open' in terms of data (they obey the published openEHR XML Schemas), models (they are driven by archetypes, written in the published ADL formalism) and APIs. They share the key openEHR innovation of adaptability, due to the archetypes being external to the software, and significant parts of the software being machine-derived from the archetypes. The essential outcome is systems and tools for computing with health information at a semantic level, thus enabling true analytic functions like decision support, and research querying.
- **EN/ISO 13606 - Electronic Health Record Communication:** This European and ISO standard defines the means to communicate a part or all of the Electronic Health Record (EHR) of a single subject of care. The standard can be seen as a harmonisation of openEHR and HL7.
- ESRI developed **spatial interoperability standards** for public health and health care delivery (8).
- **Extensible Markup Language (XML)** is the most widespread markup languages used for data exchange. It defines a set of rules for encoding data structures (including documents) in a textual data format which is both human-readable and machine-readable. It is defined by the World Wide Web Consortium's (W3C) XML 1.0 Specification (23).
- **The Resource Description Framework (RDF)** and **RDF-Schema (RDFS)** are W3C recommendations used as a general method for conceptual description or modelling of information in web resources, using a variety of syntax notations and data serialization formats, the most used is XML. It is also used in knowledge management applications (24).

- **The Web Ontology Language (OWL)** is a family of knowledge representation languages for representing ontologies. The OWL languages are extensions of RDF by constructs allowing the representation of formal semantics and. OWL1 has been extended with additional features in 2009, becoming OWL2. Both languages are supported by Protégé and DL reasoners such as FaCT++, HermiT, Pellet and RacerPro. OWL and RDF have attracted significant academic, medical and commercial interest (25).
- **Simple Knowledge Organization System (SKOS)** is a W3C recommendation designed for representation of thesauri, classification schemes, taxonomies, or any other type of structured controlled vocabulary. SKOS is part of the Semantic Web family of standards built upon RDF and RDFS, and its main objective is to enable easy publication and use of such vocabularies as linked data (26).
- **Common Terminology Services, Release 2 (CTS2)** is a Health Level 7 (HL7) and Object Management Group (OMG) specification for representing, accessing and disseminating terminological content (27). It is an extension of HL7 Version 3 Standard: Common Terminology Services, Release 1 (28).

In the United States the “Public Health Data Standards Consortium was invited by the Integrating the Healthcare Enterprise (IHE) to start a Public Health Domain at IHE. IHE is a collaborative of clinicians, administrators, standard development organizations and health information technology (HIT) vendors that drives the adoption of standards to address specific clinical needs through the development of the technical specifications for the software applications. PHDSC and IHE are collaborating to enable interoperability across clinical and public health enterprises.” (9).

10.11.3 Coding schemes, terminologies

The idea of representing certain entities by codes instead of natural language descriptors goes back many centuries. The original cause of using codes was twofold. An important aspect was the need for unambiguity, either across or within languages. The other reason was to represent the entirety of a domain by a limited number of concepts to conduct statistical studies. In the modern age the computational tractability became another point.

Most coding systems are based on some classification: entities of the given domain are arranged into a – usually hierarchical – structure. One of the earliest problems with classification was the problem of multiple hierarchies. For example, diseases can be classified by location (according to the primarily affected organ), by aetiology (infectious, acquired, hereditary etc.), by epidemiology (sporadic, epidemic, etc.), or by pathology (neoplastic, metabolic disorder, etc.). Therefore a certain disease can be a member of many different, partially overlapping classes. The problem of multiple hierarchy is quite ubiquitous, it applies for nearly all large classifications, not only in the healthcare domain.

It depends again on the purpose, which dimension should be considered as the main aspect of classification. This is one of the most important reasons, why more than one classification exists for most of the medical domains. There are other reasons of course, like differences in granularity, content coverage, availability in different languages, etc.

For public health purposes, however, the *International Classification of Diseases* (ICD) is perhaps the most frequently used classification system, although different versions of it are in use.

The terms – terminology, nomenclature, and vocabulary – are often used interchangeably. However, there are differences in these terms. Terminology can be defined as a set of terms representing the system of concepts of a particular subject field. Nomenclature is a system of terms that is elaborated according to pre-established naming rules. Vocabulary refers to a dictionary containing the terminology of a subject field.

10.11.3.1 Most important terminologies

There are various terminologies used in the health domain. Here is a partial list of terminologies widely accepted and used either globally or by many countries.

- **International Classification of Diseases and its clinical modifications:** this is one of the best known terminologies, which was first published in 1893, and has been revised at roughly 10-year intervals, by WHO. The most recent version is the 10th revision (ICD-10). WHO has been working on the 11th revision for a few years. In the United States the National Center for Health Statistics published a clinical modification of ICD-9 and now ICD-10 by adding an extra digit to the codes to provide an extra level of detail (ICD-9-CM; ICD-10-CM). The Royal College of Paediatrics and Child Health (formerly British Paediatric Association) also created a modified and extended version of ICD-9 and ICD-10 codes for birth defects (congenital anomalies).
- **International Classification of Primary Care:** This classification includes over 1000 diagnostic concepts that are partially mapped into ICD.
- **Medical Dictionary for Regulatory Activities (MedDRA)** is an international medical terminology dictionary used by regulatory authorities in the pharmaceutical industry during the regulatory process and also used for adverse event classification. It has been translated into several languages and used in the EU, Japan and the USA.
- **Systematized Nomenclature of Medicine (SNOMED):** Originally called SNOP (Systematized Nomenclature of Pathology), it has been developed by the College of American Pathologists to describe pathological findings using topographic (anatomic), morphologic, etiologic and functional terms. The current version, SNOMED CT (SNOMED Clinical Terms) was created in 1999 by the merger, expansion and restructuring of SNOMED RT (SNOMED Reference Terminology) and the Clinical Terms Version 3 (formerly known as the Read codes), developed by the National Health Service of the United Kingdom. Since 2007, SNOMED CT is maintained by the IHTSDO (International Health Terminology Standards Development Organisation).
- **GALEN and GALEN-In-Use projects in Europe:** the aim was to develop standards for representing coded patient information. The consortium developed the GRAIL concept modelling language, the structure and content of the GALEN Common Reference Model. It also created tools to enable the further development, scaling-up and maintenance of the model.
- **Logical Observations, Identifiers, Names, and Codes (LOINC) in the US, and a similar EUCLIDES work in Europe:** LOINC was created to represent laboratory tests and observations but later included also non-laboratory observations such as vital signs. A similar work (EUCLIDES) has been done in Europe.
- **WHO Drug Dictionary, ATC codes:** The Drug Dictionary is an international classification of drugs by name, ingredient, and chemical substance). It is used by pharmaceutical companies, clinical trial organizations and drug regulatory authorities for identifying drug names in spontaneous ADR reporting (and pharmacovigilance) and in clinical trials. The dictionary was created in 1968 and it is regularly updated. Since 2005 there have been major developments in the form of a WHO Drug Dictionary Enhanced (with considerably more fields and data

entries) and a WHO Herbal Dictionary, which covers traditional and herbal medicines. Drugs are classified according to the Anatomical-Therapeutic-Chemical (ATC) classification.

- **Unified Medical Language System (UMLS):** started by US National Library of Medicine in 1986, it is a quarterly updated compendium (Metathesaurus) of biomedical terminologies, providing a mapping structure among these vocabularies and thus allows the transcoding among various terminology systems. Altogether, it contains over a million concepts and 5 million terms which stem from the over 100 incorporated terminologies. Each concept in the Metathesaurus is assigned one or more semantic types, and they are linked with one another through semantic relationships. The Semantic Network provides these types and relations: there are 135 semantic types and 54 relationships in total. UMLS can be used to enhance or develop applications, such as electronic health records, classification tools, dictionaries and language translators. It can be also used for information retrieval, data mining, public health statistics reporting, and terminology research.

10.11.3.2 Mapping between classification systems

Whenever we are faced with the Babel of classification and coding systems, a trivial idea is the (automated) mapping (conversion) from one to another. At first sight, it can be done easily, for example, by a simple cross-reference table that contains the corresponding code pairs (triplets, etc.) Since coding systems are not just a set of code values, but – as mentioned – most of them are built on a classification, the matter is not so easy. Usually the categories of one classification do not fit entirely in the categories of the other. Unless the underlying classifications are totally identical, no mapping is possible between two coding systems without distortion. Theoretically, a special case is also possible: if one classification is a mere subset of another, then there is an unambiguous mapping from the former to the latter but not vice-versa.

10.11.4 Ontologies and data structures

Computer-based patient records can be improved by the use of ontologies. “An ontology specifies the conceptualization of a domain and is often comprised of definitions of a hierarchy of concepts in the domain and restrictions on the relationships between them.”(10)

An ontology representing the content of an electronic patient record may include (among others) the following:

- Clinical acts (health care flow, surgical and other procedures, etc.)
- Clinical findings
- Disease manifestation, etiology, pathophysiology
- Diagnosis

10.11.5 Mobile health delivery, personalized medicine, and social media applications

Mobile technology, social media, personalised medicine, remote diagnostics could transform health care. The number of e-health applications available for mobile devices steadily increases. Developing communication standards for information and communication technologies to facilitate interoperability among systems and devices, provide privacy and security, and address the needs of the developing world is timely and important.

Personalised medicine, “A form of medicine that uses information about a person’s genes, proteins, and environment to prevent, diagnose, and treat disease”, is a new area of e-health when personalised medical records are generated (11).

Social media applications related to health are on the rise. Patients often consult medical information online, and turn to social media communities for peer-to-peer support and information. Lot of information can be obtained but careful considerations are needed to filter out useful information (12).

10.12 Problem with populations

10.12.1 Definition of population

Comparability of data of population-based registries requires clear definition of the given population. Without such a clear definition it cannot be certain, for example, that there is no overlap between the populations of the registries. This is especially true within the EU, where free mobility of people increases the probability that the same person is registered in different registries.

The definition of population in general is in itself not without difficulties. Most often, “population” is defined as a group or collection of individuals inhabiting a certain territory or forming an interbreeding community. There is a proposed definition of population especially for public health purposes saying that “A population (in public health) is a group of persons sharing a common resource.”(15)

10.12.2 Inclusion and exclusion criteria

To generate comparable data on a population level, requires having the same set of inclusion and exclusion criteria (i.e., residency status, socio-demographic data, geographic area, etc.), therefore using data from two or more systems or registries could be interpreted in a uniform fashion. For example, the definition of stillbirths (gestational age cut-off point) varies by country and collecting information on the stillbirths population and comparing characteristics and prevalence could lead to false interpretation of data. When comparing rates of population-based registries, the residency status criterion, whether including or excluding non-resident persons living in a defined geographic area, is very important.

10.12.3 Mobility

Free mobility within and across borders makes the establishment of population-based registries (especially in a smaller geographical area) and comparison of data between other registries without the risk of having the same person recorded in two or more databases challenging. National and EU level, or global registries could help eliminate this problem. Communication between systems and linking data on a regular basis could also help in finding duplicate records and make data comparable.

10.12.4 Socio-demographic, genetic factors

Variations and differences in socio-demographic and genetic factors such as ethnicity, genetic mutations in certain populations could make it difficult or even nearly impossible to compare some specific data among populations.

10.13 Examples of legal frameworks for data protection and data sharing

Data exchange is sometimes a complex process, and organisations, registries, and data providers have to ensure compliance with cross-border restrictions, privacy and confidentiality rules. All member countries of the EU impose restrictions on the sharing of personal information outside the EU. Organisations sharing personal information collected in the EU with service providers based outside the EU need to find ways to comply with these laws (16).

Privacy generally applies to people, while confidentiality applies to information. There are many important reasons to protect privacy and confidentiality.

Privacy is the control over the extent, timing, and circumstances of sharing oneself (physically, behaviorally, or intellectually) with others. For example, persons may not want to be seen entering a place that might stigmatize them, such as a pregnancy counselling centre clearly identified by signs on the front of the building. The evaluation of privacy also involves consideration of how the researcher accesses information from or about potential participants.

Confidentiality pertains to the treatment of information that an individual has disclosed in a relationship of trust and with the expectation that it will not be divulged to others in ways that are inconsistent with the understanding of the original disclosure.

Maintaining privacy and confidentiality helps to protect participants from potential harms including psychological harm such as embarrassment or distress; social harms such as loss of employment or damage to one's financial standing; and criminal or civil liability. Especially in social/behavioral research the primary risk to subjects is often an invasion of privacy or a breach of confidentiality.

The next sections present a few examples for data sharing policies and regulations related to health information in Europe and in the United States.

10.13.1 Policy on data submission, access, and use of data within TESSy

The European Centre for Disease Prevention and Control (ECDC) created the European Surveillance System (TESSy) to collect, analyse and disseminate surveillance data on notifiable infectious diseases in Europe. A procedure with a set of rules was developed for data submission, data storage and custody, data use and data access, and data protection. Relevant forms and notes are also available (17):

- Request for TESSy Data for Research Purposes
- Declaration Regarding Confidentiality and Data Use
- ECDC Data Disclaimer
- Conditions for Publishing Note
- Sample Agreement for Agencies and third parties
- Declaration on Data Destruction

10.13.2 European Commission's proposal for a General Data Protection Regulation

The European Patients' Forum, which is a not-for-profit, independent organisation and umbrella representative body for patient organisations throughout Europe, wrote a position statement on

general data protection regulation, and made recommendations to the European Commission, the European Parliament and Member States to:

- 1) Ensure that the Regulation protects patients' rights as data subjects and as owners of their health and genetic data, and contains measures to enable patients to benefit from these rights effectively (e.g. access to data, data portability, right to information and transparency). Any restriction due to the special nature of the data processed or legitimate reasons for processing of such data should be justified and limited to what is necessary for public health, or the patients' vital interests.
- 2) Make the necessary adaptations to the Regulation in order not to hamper provision of care, the conduct of research and public health activities, including patient registries and activities carried out by patient organisations to advance research and patients' rights, with clear and explicit provisions to ensure the good implementation of this Regulation in the health sector.
- 3) Put in place effective cooperation measures between Member States and minimum security requirements to ensure an equivalent level of protection of personal data shared by patients for healthcare and research purposes across the European Union, and facilitate cross-border healthcare and research.
- 4) Involve patient organisations in decision-making and activities at policy and programme level for questions that relate to the processing and sharing of patients' personal data, transparency towards patients and informed consent, to ensure the processing is carried out ethically and in a transparent manner throughout the European Union (18).

10.13.3 European Data Protection Board, General Data Protection Regulation

The European Commission plans to unify data protection within the European Union (EU) with a single law, the General Data Protection Regulation (GDPR). The current EU Data Protection Directive 95/46/EC does not consider important aspects like globalisation and technological developments such as social networks and cloud computing sufficiently. New guidelines for data protection and privacy are required to address these issues. Therefore a proposal for a regulation was released in 2012. Subsequently numerous amendments have been proposed in the European Parliament and the Council of Ministers. The EU's European Council aimed for adoption of the GDPR in late 2014 and the regulation is planned to take effect after a transitional period of two years.

10.13.4 HIPAA Privacy and Security Rules for Public Health Data Exchange

In the United States the Health Insurance Portability and Accountability Act of 1996 (HIPAA) Privacy, Security and Breach Notification Rules were developed.

"The Office for Civil Rights enforces the HIPAA Privacy Rule, which protects the privacy of individually identifiable health information; the HIPAA Security Rule, which sets national standards for the security of electronic protected health information; the HIPAA Breach Notification Rule, which requires covered entities and business associates to provide notification following a breach of unsecured protected health information; and the confidentiality provisions of the Patient Safety Rule, which protect identifiable information being used to analyze patient safety events and improve patient safety."(19, 20, 21)

References

1. <http://web.mit.edu/6.933/www/Fall2001/Shannon2.pdf>
2. <https://grdr.ncats.nih.gov/> last visited 20/07/2014
3. <http://confidential.oxfordradcliffe.net/anondata> last visited 26/05/2014
4. <http://www.bristol.ac.uk/Depts/DeafStudiesTeaching/ethics/resource/anonymise.pdf> last visited 26/05/2014
5. http://www.jiscdigitalmedia.ac.uk/clinical-recordings/storage_anonymisation.html last visited 26/05/2014
6. <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:EN:HTML> last visited 20/07/2014
7. <http://www.ecdc.europa.eu/en/activities/surveillance/tessy/Pages/TESSy.aspx> last visited 20/07/2014
8. <http://www.esri.com/library/whitepapers/pdfs/hl7-spatial-interoperability.pdf> last visited 20/7/2014]
9. http://www.phdsc.org/health_info/ihe-task-force.asp last visited 20/07/2014
10. A framework ontology for computer-based patient record systems <http://ceur-ws.org/Vol-833/paper28.pdf> last visited 20/07/2014
11. <http://www.cancer.gov/dictionary?cdrid=561717> Last visited: 20/07/2014
12. <https://itunews.itu.int/en/2472-E8209health-standards-and-interoperability.note.aspx> last visited 20/07/2014
13. WhatIs.com Tech definitions <http://whatis.techtarget.com/definition/compatibility>]
14. Development of a framework for Mapping European Seabed Habitats (MESH) <http://www.searchmesh.net/default.aspx?page=1826>
15. Surjan G. Ontological definition of population for public health databases. Stud Health Technol Inform. 2005; 116:941-5.
16. <http://media.mofo.com/files/Uploads/Images/130729-BNA-Cross-Border-Information-Sharing-for-Effective-Services.pdf> last visited 20/07/2014
17. <http://www.ecdc.europa.eu/en/activities/surveillance/tessy/documents/tessy-policy-data-submission-access-and-use-of-data-within-tessy-2011%20revision.pdf> last visited 20/07/2014
18. http://www.eu-patient.eu/Documents/Policy/Data-protection/Data-protection_Position-statement_10-12-2012.pdf last visited 20/07/2014
19. http://www.cdc.gov/phinf/resources/standards/data_interchange.html last visited 20/07/2014
20. <http://www.hhs.gov/ocr/privacy/index.html> last visited 20/07/2014
21. http://www.cdc.gov/phinf/resources/standards/data_interchange.html
22. Int J Med Inform. 2013 Jan; 82(1):1-9. doi: 10.1016/j.ijmedinf.2012.11.003. Epub 2012 Nov 20. Trustworthy reuse of health data: a transnational perspective. Geissbuhler A1, Safran C, Buchan I, Bellazzi R, Labkoff S, Eilenberg K, Leese A, Richardson C, Mantas J, Murray P, De
23. <http://www.w3.org/TR/REC-xml/>] which is a free open standard
24. <http://www.w3.org/TR/REC-rdf-syntax/>, <http://www.w3.org/TR/rdf-schema/>
25. www.w3.org/TR/owl-features/, www.w3.org/TR/owl2-overview/]
26. <http://www.w3.org/2004/02/skos/>
27. <http://wiki.hl7.org/index.php?title=CTS2>, <http://www.omg.org/spec/CTS2/>
28. http://www.hl7.org/implement/standards/product_brief.cfm?product_id=10
29. http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_en.pdf

11 APPENDICIES

Appendix A: RoR survey questionnaire respondents

no.	Short registry name	Full registry name (English)
1	AT Stroke Unit	Austrian Stroke Unit Registry
2	AT Adult Heart Surgery	Registry for adult heart surgery
3	AT IVF	Invitro fertilisation registry
4	AT Pacemaker Looprecorder	Austrian registry for pacemakers, ICD and looprecorder
5	AT Cancer	Austrian National Cancer Register
6	HR Birth Medical	Birth Medical Database
7	HR Persons with Disabilities	Croatian Registry of Persons with disabilities
8	HR Vaccination Side-effects	Registry of vaccination side-effects
9	HR CroDiab	Croatian Diabetes Registry
10	HR Health Morbidity	Health Morbidity Database of Croatia
11	HR Cause of Death	Causes of death registry
12	HR Suicides	Croatian Committed Suicides Registry
13	HR Psychoses	Croatian Psychoses Registry
14	HR Cancer	Croatian National Cancer Registry
15	CY Injury	Injury Database
16	FI Induced Abortions	Register on Induced Abortions and sterilisations
17	FI Medical Birth	Medical Birth Register
18	FI Infectious Disease	National Infectious Disease Regsiter
19	FI Congenital Malform	Finnish Register of Congenital Malformations
20	FI Cancer	Finnish Cancer Registry
21	FI Vaccine Events	Vaccine Adverse Events Register
22	FI Kidney Disease	Finnish Registry for Kidney Diseases
23	FI Prescription Registry	KELA
24	FI Visual Impairment	The Finnish Register of Visual Impairment
25	FI Cause of Death	Cause of death, Statistics Finland
26	FI Social Welfare	Care Registers for Social Welfare (SosiaaliHILMO) and Count of regular home-care clients, 30 November
27	FR HEOP	Register of childhood disabilities and perinatal surveillance
28	FR DAHD	Drug allegry and hypersensitivity
29	FR RACT	Alpine Rhône thyroid cancer registry
30	FR ReLSEP	ReLSEP (Multiple sclerosis)
31	FR RHE31	RHE31 - Children with severe neurodevelopmental disabilities
32	DE ESSCA	European Surveillance System on Contact Allergies
33	DE CDA	German Registry of Congenital Dyserythropoietic Anemia (CDA)
34	GR Kidney transplants	Renal Patients waiting for kidney transplatation from diseased donors
35	HU Infarction	Hungarian Infarction Registry
36	HU Healthcare Episode	Healthcare Episode Database
37	HU Cataract	National Cataract Registry

38	HU IVF	In vitro fertilization
39	HU Antibiotic Resistance	National antibiotic resistance surveillance
40	HU Drug Mortality	Drug-related mortality
41	HU Neonatal ICU	Database of the Hungarian Neonatal Intensive Care Units
42	HU TBC Surveillance	TBC Surveillance System
43	HU Pediatric Cancer	Hungarian Pediatric Cancer Registry
44	HU Arthritis HBSJGB	Arthritis Registry of Polycyclic of the Hospitaller Brothers of St. John of God in Budapest
45	HU Arthroplasty	Hungarian Arthroplasty Register
46	HU HCAR	Hungarian congenital anomalies registry
47	HU Pacemaker	Hungarian Online Pacemaker Registry
48	HU Hepatitis	Hepatitis Registry
49	HU Birth registry	Birth Registry
50	IT Coro/Cerebro Events	National Registry of Coronary and Cerebrovascular Events
51	IT Creutzfeldt-Jacob	Italian Registry of Creutzfeldt-Jakob disease and related disorders
52	IT Bone Marrow Donor	Italian Bone Marrow Donor Registry
53	IR OHCAR	OHCAR - National Out-of-Hospital Cardiac Arrest Register
54	LV HIV/AIDS	National HIV/AIDS case register
55	LV Helthcare Users	Registry of State Health Care Users
56	LV Management IS	MaNAgeMENT Information System
57	LV Mental Disorders	The Register of Patients with Particular Diseases, Patients with Mental Disorders
58	LV Injury	Register of the patients who have suffered injuries
59	LV Diabetes Mellitus	Register of Patients with Particular Diseases, Patients with Diabetes Mellitus
60	LV Drug Use Disorders	Register of the patients with drug use disorders
61	LV Occupational Diseases	Register of the patients with occupational diseases
62	LV Cancer	Register of the patients with cancer
63	LV MS	Register of the patients with multiple sclerosis
64	LV Congenital Anomalies	Register of the patients with congenital anomalies
65	MT NOIS	National Obstetrics Information System (NOIS)
66	MT Congenital Anomalies	Malta Congenital Anomalies Registry
67	MT Cancer	Malta National Cancer Registry
68	MT Hospital IS	National Hospitals Information System
69	MT Mortality	National Mortality Registry
70	MT Injury	Injury Database
71	PL Benign Cancer Salivary	National Registry of Benign Cancer of Major Salivary Glands
72	PL Benign Tumor Jaw	Registry of Benign Tumor of Jaw
73	PL Cancer	Polish National Cancer Registry
74	PL ACS	Polish Registry of Acute Coronary Syndromes PL-ACS
75	PL Fetal Cardiac Anomalies	National Registry for Fetal Cardiac Anomalies
76	PL Brain Strokes	Polish Registry of Brain Strokes
77	PL Late Effect Child Cancer Tr	Database of Late Effects Follow-Up after Childhood Cancer Treatment
78	PL Alfa-1 Antitripsin Deficiency	National Registry of Patients with Alfa-1 Antitripsin Deficiency

79	PT PAH	PAH - Portuguese Registry
80	PT Programa de Rastreio de Cancro da Mama	Breast cancer screening database
81	RO TeleDiab	Black Sea TeleDiab; SincoDiab; SincroPAD
82	SK TB	National TB Register
83	SI Cancer	Cancer Registry of Republic Slovenia
84	SI Tuberculosis	Registry for tuberculosis
85	SI Drug Sales	Drug sales register, IVZ64
86	SI Perinatal IS	Perinatal information system
87	SI Hospital Discharge	Hospital discharge registry
88	SI Injuries at work	Injuries at work Registry
89	SI Child Diabetes	Registry for diabetes for children <18years.
90	SI Vaccination Side-effects	Vaccination side-effects registry
91	SI Mortality	mortality database
92	SE Food Anaphylaxis	Food Anaphylaxis
93	SE Airway	Swedish Airway Registry
94	ES REGICOR	Girona Heart Registry
95	ES Williams Beuren	Williams Beuren Syndrome
96	ES RDR	Spanish national Rare Diseases Registry
97	ES Alfa-1 Deficit Patients	Spanish Registry of Patients with Alfa-1 Antitrypsin Deficiency
98	ES Congenital Anomaly Valencia C	Population-based Registry of Congenital Anomalies of Comunitat Valenciana
99	ES NEPAD	Online Pediatric Registry NEPAD
100	ES Cardiac Transplant	Spanish Registry of Cardiac Transplants
101	ES Cancer Girona	Girona Cancer Registry
102	ES RNEETH	The National TSEs Register
103	ES FEETEG	Foundation for the Study and Therapy of Gaucher Disease and Other Lysosomal
104	ES SCH Stroke	Sagrat Cor Hospital of Barcelona Stroke Registry
105	ES VACUSAN	Regional vaccination registry of the Murcia Region Autonomous Community
106	ES SIGIVIH	The Galician Service of Information on HIV
107	ES Mastocytosis	Spanish Red Mastocytosis (REMA)
108	ES SPAVASC	Registry of the Spanish Society of Angiology and Vascular Surgery
109	ES Toxic Oil Syndrome	Patient Registry for Toxic Oil Syndrome
110	ES SIVIH Valencia	Information System of New HIV Infections and AIDS Cases (SIVIH) of Valencia
111	ES REWBA	Spanish Registry of Wolfram, Bardet-Bidl and Alström Syndromes
112	ES SITUB	Galician Tuberculosis Information System
113	ES Asbestos	Asbestos registry
114	ES REPIR II	Spanish Pediatric Registry of Chronic Renal Insufficiency stage 2-5 without substitute treatment
115	ES Asthma Data Bank	Asthma Data Bank
116	ES Neurogenetic HLF	La Fe Hospital Registry of Neurogenetic Illnesses
117	ES Renal Patient Registry	Spanish Registry of Renal Patients
118	ES PII Tuberculosis	Integrated Tuberculosis Investigation Program
119	ES ADS	ADS Registry

120	ES ACAD EoE	ACAD Eosinophilic esophagitis registry
121	ES Infant Tumor Valencia C	Childhood Cancer Registry of C. Valenciana
122	ES Cancer Castellon	Castellón Cancer Registry
123	ES Cancer Tarragona	Tarragona Cancer Registry
124	ES Voluntary Abortions	Voluntary Abortions Registry
125	ES Perinatal Mortality	Perinatal Mortality
126	ES Basic Hospitalization Data	Minimal Basic Hopitalization Data Set
127	ES Cancer Granada	Granada Cancer Registry
128	ES SIDIAP	Information System for the Development of Primary Care Research
129	SC Lanarkshire paediatric allergy	Lanarkshire paediatric allergy database
130	HU National Cancer registry	Hungarian National Cancer Registry
131	DE German Asthma Net GAN	German Asthma Net Registry

Appendix B: Registry contact record template

[illegible]

	International	University 2									
Relevant Experts	National	Database/Data									
		Epidemiology									
	International	Health Economist									
		Registry									
Professional Groups and Societies	National	Royal College of Physicians									
Registry Groups	National	e.g. National cancer registry									
	International (EU)	e.g. EAR									
		e.g. EUBIROD									
	International (non-EU)										
Registry Sponsor Groups	National										
	International										
Development Groups	National										
	International (EU)										
	International (non-EU)										
Other groups	National										
	International	SNOMED-CT (IHTSDO)									
		ICD-11 Working Group									
		International Organization for Standardization									
		Ecosystem Support Groups									
		EU Data Protection									