EUCERD Joint Action

THESAURUS OF REGISTRY TERMINOLOGY

EUCERD JOINT ACTION WP8
TASK 8.1: INTEGRATION OF REGISTRY ACTIVITIES

JANUARY 2015

Co-funded by
the Health Programme
of the European Union

Thesaurus of Registry Terminology and Types

Abstract

Many stakeholders in the field of rare diseases use registries, talk about registries or plan to establish these powerful tools for research and/or clinical care or quality management purposes. Unfortunately there is no common definition of many of the terms and items involved. This often causes misunderstandings because stakeholders each presume that their own interpretation of the meaning of a term is unequivocal. Since this is not the case, but a common understanding of the subject matter is essential, WP8 Task 1 has started to collect terms and phrases that arise when discussing and developing RD registries and has proposed definitions or specific meanings to these.

The *Thesaurus* features three core areas of definitions: registry types, data access models, and registry-related terminology. This is complemented by a matrix of registry and documentation types to help determine which type of data scheme or access model is necessary/required/advisable for each type of registry, and by an overview of the most commonly used abbreviations.

The benefits of such a collection of terms and definitions are manifold. In particular, the Thesaurus will:

- provide clarity on terminology, purpose and strategy
- aid in developing new registries with optimum use of resources
- facilitate interoperability of registries
- promote information exchange on RD patients between stakeholder groups
- optimize data collection and exploitation

We consider this document a “work in progress” eventually becoming a reference tool of relevant terminology which all registry stakeholders can use to avoid confusion and misunderstandings. Through the “open approach” of actively inviting suggestions for additions, the *Thesaurus* can eventually provide a widely accepted set of terms and definitions and thus form a comprehensive basis for smooth and profitable registry collaboration on a European and international level.

Methodology

Terms and phrases with the potential to be included in the *Thesaurus* could and can continue to be proposed together with a tentative explanation or definition of the term by any stakeholder. After discussion within the Frankfurt part of EUCERD JA WP 8 in collaboration with the OSSE (Open Source Registry System for Rare Diseases) working group, the *Thesaurus* was made available on the EUCERD website, again inviting further feedback.

A draft version of the *Thesaurus* has been submitted to the JRC and it is proposed this be used as the basis for common terminology in all RD registry communications from and with the European platform for RD patient registration.

This work was enabled by the EUCERD Joint Action (DG SANCO No. 2011 2201).

PLEASE NOTE:

This document is seen only as a starting point for further discussions within the EUCERD Joint Action and among further stakeholders from the area of European RD registration. Your feedback is greatly appreciated – please send your comments and suggestions to

Thomas Wagner
EUCERD Joint Action WP 8
Goethe University Hospital Frankfurt
t.wagner@em.uni-frankfurt.de

"Thesaurus of Registry Terminology and Types"
Abstract

Patient registries are a crucial factor in advancing research and hence therapeutic options for rare disease (RD) patients. With numerous often heterogeneous registry efforts existing on the national level, the new European RD registry platform will be a leap forward for a European and international collaboration of registry activities. Since a common understanding of the subject matter is essential to fully seize the potential of such an enterprise, the *Thesaurus of Registry Types and Terminology* was created as a reference document containing definitions of various registry-related terms and concepts in order to avoid confusion and help harmonize collaboration efforts.

Emanating from work carried out within the EUCERD Joint Action on Rare Diseases, the *Thesaurus* features three core areas of definition: registry types, data access models, and registry-related terminology. This is complemented by a matrix of registry and documentation types to help determine which type of data scheme or access technology is necessary/required/advisable for each type of registry, and by an overview of the most commonly used abbreviations.

The benefits of such a collection of terms and definitions are manifold. In particular, it will

- provide clarity on terminology, purpose and strategy
- aid in developing new registries with optimum use of resources
- facilitate interoperability of registries
- promote information exchange on RD patients between stakeholder groups
- optimize data collection and exploitation

The *Thesaurus* is considered a work in progress which can function as a reservoir to collect items that arise when discussing and developing RD registries. Comments and suggestions for additions are welcome and encouraged. Thus providing a widely accepted set of terms and definitions, the *Thesaurus* can serve as a comprehensive basis for smooth and profitable registry collaboration on a European and international level.

Registry-related Terminology

A-Z	Term	Definition
A	**Aggregation** Cohort	**Aggregation** means the consolidation/compaction of information. Example: The data set {John Doe, born March 1st, 1981; Jane Roe, born May 26th, 1980} is aggregated by a specified aggregate function to {two patients aged 30-40 years}. Of course, various other aggregate functions can be constructed. These can be used to build a **cohort**, which is a group of humans with common time-specific demographic attributes; in the above example the birth cohort 1972-1982.
B	**Basic Data Set (BDS)**	A **basic data set** is the selection of items in a registry that allows to answer the standard set of questions the registry was established for.
	Basic research registry	A **basic research registry** is a data repository serving the needs of researchers (medical sciences, biochemistry, pharmacology, etc.) working in academia or industry, mainly adding and connecting relevant information to biomaterial bank samples (some overlap of data with epidemiological and clinical care registries).
C	**Clinical care registry**	A **clinical care registry** is a data gathering activity of care and best practices which documents the main elements of care (structures, processes, and outcomes) for the support of clinicians of one specialization or of all relevant specializations concerned with (a) specific disorder(s). It will also serve the needs of other care team members (nurses, physiotherapists, psychologists, social workers, dieticians, speech therapists, etc.) or support the interaction with local health care (attending physicians / rehabilitation centres etc.) and could smoothen the transition from paediatric to adult care (significant overlap of data with epidemiological and clinical research/trial registries).
	Clinical research / clinical trials registry (CTN)	A **clinical research/clinical trial registry (CTN)** is a data gathering activity concerning information about availability of patients suited for participation in clinical research/trials. The data are closely related to those of a clinical care network, since the same participants may be involved but some information may be different (significant overlap of data with epidemiological and clinical care registries).
	Cohort	see **"aggregation"**
	Common Data Set (CDS)	A **common data set (CDS)** is the overlap of all items of all registries that are to interoperate and which are sufficient to answer the questions planned to be addressed by the platform
D	**Data scheme**	see **"registry"**
	Data set	see **"registry"**
E	**Epidemiological registry**	An **epidemiological registry** is a data gathering activity concerned with (a) specific disorder(s) and aimed at delivering epidemiological data (incidence and prevalence of the diseases or group of diseases). It has to cover as many patients as possible to provide representative data for epidemiological purposes (almost complete coverage of all possible cases to specific disorder(s)).
	European Centralized RD Registry	A **European Centralized RD Registry** is one central registry for all patients of a given rare disease, group of rare diseases or all rare diseases in Europe.

M	**Meta content**	see **"metadata"**
	Meta registry for rare diseases	A **meta registry for rare diseases** is a registry containing meta content of existing disease-specific registries of rare diseases. Its data sets are the "lowest common denominator" of all contributing registries and are therefore called minimum data sets. They correspond to a minimal data schema * much narrower than that of a disease-specific registry. As a result, a meta registry is unsuitable for concrete disease-specific evaluations but will provide interested researchers with a first overview of existing cases along with contact details for data holders having further information or even corresponding biomaterial available. ** The registry provides the minimal dataset (content) according to the minimal data schema (structure). So when we talk about searching a minimal data set, we actually mean to search a minimal data schema .*
	Metadata Meta content	**Metadata** are data describing the structure of data. For example, a well-documented data schema describes the meaning (semantics) and the possible range of values of compatible data sets. Metadata must not be confused with data about data sets/content. The latter are called **meta content**, which is an aggregation of data sets as seen under **"aggregation"**.
	Metadata registry/repository (MDR)	A **metadata registry/repository (MDR)** is a registry about elements of data schemas. These are stored in a structured way – the data schema can, for example, orient itself by the ISO/IEC standard 11179. An MDR allows comparing data schemas and, as a consequence, data sets from different collections (registries, studies, databases, etc.).
	Minimal Data Set (MDS)	A **minimal data set** is the smallest possible selection of items in a registry that allows to answer a single question.
R	**Registry** Data scheme Data set	A **registry** is a (not necessarily electronic) database containing a systematic collection of structured information describing "objects" of an arbitrary but constant type. This structure (i.e. the valid values to be entered) constitutes the **data schema**. The information items (user-entered data) themselves are called **data sets** or content. Example: in the management of data sets in a table, the data schema corresponds to the table heading ("Surname, Name") and the data set to the table row ("Doe, John"). Registries can differ in terms of diversity, complexity, reach or organization. A classification can be found in the NAMSE concept paper.
	Registry toolbox	A **registry toolbox** is a collection of composite tools for establishing and running a disease-specific registry for rare diseases. Recent unfinished discussions propose the following content: • quality criteria and standards • technical aid for building a database structure • legal requirements • means for long-term sustainability
	Registry of Registries (RoR)	A **Registry of Registries (RoR)** is the most basic type of documentation – a comprehensive listing of available registries on a given rare disease, including some information on the type of the respective registries, number of patients enrolled, etc. plus connecting links to relevant registries. Example: if a user would like to know the weight of all Prader Willi patients aged 25, the RoR will yield the information which registry could provide this kind of data, and also provide contact information for the specific registries.
	Registry Data Hub (RDH)	A **Registry Data Hub (RDH)** is an automated interface allowing direct inquiries for specific data on a given rare disease with all existing registries for this rare disease.
W	**Web portal for rare diseases**	A **web portal** forms a "registry of registries" making existing registries for rare diseases visible to patients and researchers alike. Further information can be provided, e.g. on current or planned clinical studies. In contrast to the meta registry, the web portal and its contents are freely available over the internet. Registries are described not on a case basis but as a whole, similar to the TMF Biobank registry (www.biobanken.de), which does not collect data on a biosample level but solely from questionnaires filled in by the participating biobanks.

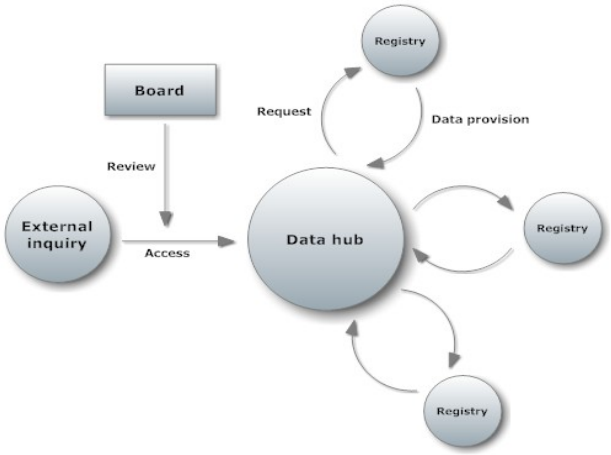
Registry Types

Type of Registry	Definition	Objectives
Epidemiological registry	a data gathering activity concerned with (a) specific disorder(s) aimed at delivering epidemiological data (incidence and prevalence of the diseases or group of diseases). The registry has to cover as many patients as possible to provide representative data for epidemiological purposes (almost complete coverage of all possible cases to specific disorder(s)).	• Incidence of the given RD or group of RDs across Europe • Prevalence of the given RD or group of RDs across Europe • Regional differences of I and/or P across Europe
Clinical care registry	a data gathering activity of care and best practices by documenting the main elements of care (structures, processes, and outcomes) for the support of clinicians of one specialization or of all relevant specializations concerned with (a) specific disorder(s). The registry work will also serve the needs of other care team members (nurses, physiotherapists, psychologists, social workers, dieticians, speech therapists, etc.) or support the interaction with local health care (attending physicians / rehabilitation centres etc.) and could smooth the transition from paediatric to adult care (significant overlap of data with epidemiological and clinical research/trial registries).	• Making care (resources, processes, outcomes) comparable across Europe • Gaining knowledge about medical care structures • Cross-border quality assessment incl. benchmarking processes • Implementation of cross-border referral incl. a system to avoid unnecessary cross-border health care (e.g. stepwise approach)
Clinical research / clinical trials registry (CTN)	a data gathering concerning information about availability of patients suited for participation in clinical research/trials. The data are closely related to those of a clinical care network, since the same participants may be involved but some information may be different (significant overlap of data with epidemiological and clinical care registries).	• Facilitated feasibility assessment across Europe to assure sufficient study populations • Facilitated access to patients/patient data for clinical trials across Europe • Cross-border multi-centre trials with SOPs in all participating centres/sites
Basic research registry	a data repository serving the needs of researchers (medical sciences, biochemistry, pharmacology, etc.) working in academia or industry, mainly adding and connecting relevant information to biomaterial bank samples (some overlap of data with epidemiological and clinical care registries).	• Facilitation of basic research in Europe by shared expertise and large cohorts of patients or patient materials. • Facilitated access to tissue/genome samples and patient data • Sharing of resources (antibodies, cell lines, DNA constructs, etc.) and human specimens with knowledge about the purpose of the study/trial

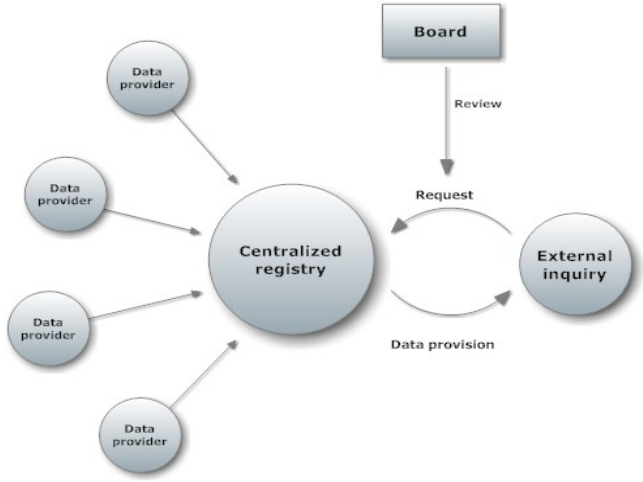
Type of Documentation / Data Access Models

Type of documentation	Definition	Data Collection and Administration	Access
Registry of Registries (RoR)	the most basic type of documentation – a comprehensive listing of available registries on a given RD incl. some information on the type of the respective registries, number of patients enrolled, etc. plus connecting links to relevant registries.	no actual patient data collected or provided	unrestricted/public
Registry Data Hub (RDH)	an automated interface allowing direct inquiries for specific data on a given RD with all existing registries for this RD	decentralized – individual registries collect and manage their own data; the Data Hub could be considered to be the automated version of the metadata repository and the RoR: a search engine is capable and allowed to search all European, national, local registries connected to the Data Hub	restricted/non-public; registration required; inquiries to be reviewed and approved/rejected by a Board
European Centralized Registry	one central registry for all patients of a given RD, group of RDs or all RDs in Europe	centralized – data provided by point of care and directly delivered to the centralized registry; all data collected and managed in one large registry	restricted/non-public; inquiries to be reviewed and approved/rejected by a Board

Registry Data Hub



European Centralized Registry



Matrix of Registry and Documentation Types / Access Models

Which type of data scheme or access technology is necessary/required/advisable for each respective type of registry? (Yes/No)

	Epidemiological Registry	Clinical Care Registry	Clinical Research / Clinical Trials Registry (CTN)	Basic Research Registry
Registry of Registries (RoR)				
Registry Data Hub (RDH)				
European Centralized Registry				

Useful Registry-related Abbreviations

Taken and adapted from: Agency for Healthcare Research and Quality (AHRQ), *Registries for Evaluating Patient Outcomes: A User's Guide*

Standard	Acronym	Description and Web Site	Web Site	Developer
Billing related				
Current Procedural Terminology	CPT	Medical service and procedure codes commonly used in public and private health insurance plans and claims processing.	http://www.ama-assn.org/ama/pub/category/3113.html	American Medial Association
International Classification of Diseases	ICD, ICD-O ICECI, ICF, ICPC	International standard for classifying diseases and other health problems recorded on health and vital records. ICD-9-CM, a modified version of the ICD-9 standard, is used for billing and claims data in the United States, which will transition to ICD-10-C	http://www.who.int/classifications/icd/en	World Health Organization
Clinical				
Systemized Nomenclature of Medicine	SNOMED CT	Clinical health care terminology that maps clinical concepts with standard descriptive terms. Formerly SNOMED RT and SNOP.	http://www.ihtsdo.org/snomed-ct	International Health Terminology Standards Development Organization
Unified Medical Language System	UMLS	Database of 100 medical terminologies with concept mapping tools.	http://www.nlm.nih.gov/research/umls/	National Library of Medicine
Classification of Interventions and Procedures	OPCS-4	Code for operations, surgical procedures, and interventions. Mandatory for use in National Health Service (England).	http://www.datadictionary.nhs.uk/web_site_content/supporting_information/clinical_coding/opcs_classification_of_interventions_and_procedures.asp	Office of Population, Censuses, and Surveys
Diagnostic and Statistical Manual	DSM	The standard classification of mental disorders used in the United States by a wide range of health and mental health professionals. The version currently in use is the DSM-IV.	http://www.psych.org/MainMenu/Research/DSMIV.aspx	American Psychiatric Association

Standard	Acronym	Description and Web Site	Web Site	Developer
Drugs				
Medical Dictionary for Regulatory Activities	MeDRA	Terminology covering all phases of drug development, excluding animal toxicology. Also covers health effects and malfunctions of devices. Replaced COSTART (Coding Symbols for a Thesaurus of Adverse Reaction Terms).	http://www.meddramsso.com	International Conference on Harmonisation (ICH)
VA National Drug File Reference Terminology	NDF-RT	Extension of the VA National Drug File; used for modeling drug characteristics, including ingredients, chemical structure, dose form, physiologic effect, mechanism of action, pharmacokinetics, and related diseases.	Not available	U.S. Department of Veterans Affairs
National Drug Code	NDC	Unique 3-segment number used as the universal identifier for human drugs.	http://www.fda.gov/cder/ndc/	U.S. Food and Drug Administration
RxNorm	RxNorm	Standardized nomenclature for clinical drugs. The name of a drug combines its ingredients, strengths, and/or form. Links to many of the drug vocabularies commonly used in pharmacy management and drug interaction software.		National Library of Medicine
World Health Organization Drug Dictionary	WHODRUG	International drug dictionary.	http://www.who.int/druginformation/index.shtml	World Health Organization
Lab specific				
Logical Observation Identifiers Names and Codes	LOINC®	Concept-based terminology for lab orders and results.	http://www.regenstrief.org/loinc/	Regenstrief Institute for Health Care
Other				
HUGO Gene Nomenclature Committee	HGNC	Recognized standard for human gene nomenclature.	http://www.genenames.org/aboutHGNC.html	Human Genome Organization
Dietary Reference Intakes	DRIs	Nutrient reference values developed by the Institute of Medicine to provide the scientific basis for the development of food guidelines in Canada and the United States	http://www.iom.edu/CMS/54133.aspx	Institute of Medicine Food and Nutrition Board
Substance Registry Services	SRS	The central system for standards identification of, and information about, all substances tracked or regulated by the Environmental Protection Agency.	http://iaspub.epa.gov/sor_internet/registry/substreg/home/overview/home.do	Environmental Protection Agency
Online Mendelian Inheritance in Man	OMIM code	Six-digit number referencing each disease and gene; first digit classifies the method of inheritance	www.omim.org	Johns Hopkins School of Medicine
Orphanet Classification of Rare Diseases	Orpha number	Unique identifying number assigned by Orphanet to a given disease (crosslinked with OMIM number and ICD	www.orpha.net	