FINAL

Deliverable D9.3

Common Data Set and disease-, treatment and other specific modules.

III-Proposal for a Platform set of Common Data Elements

Luciano Vittozzi, Emanuela Mollo, Sabina Gainotti, Domenica Taruscio

National Centre for Rare Diseases, National Institute of Health, Rome (Italy)

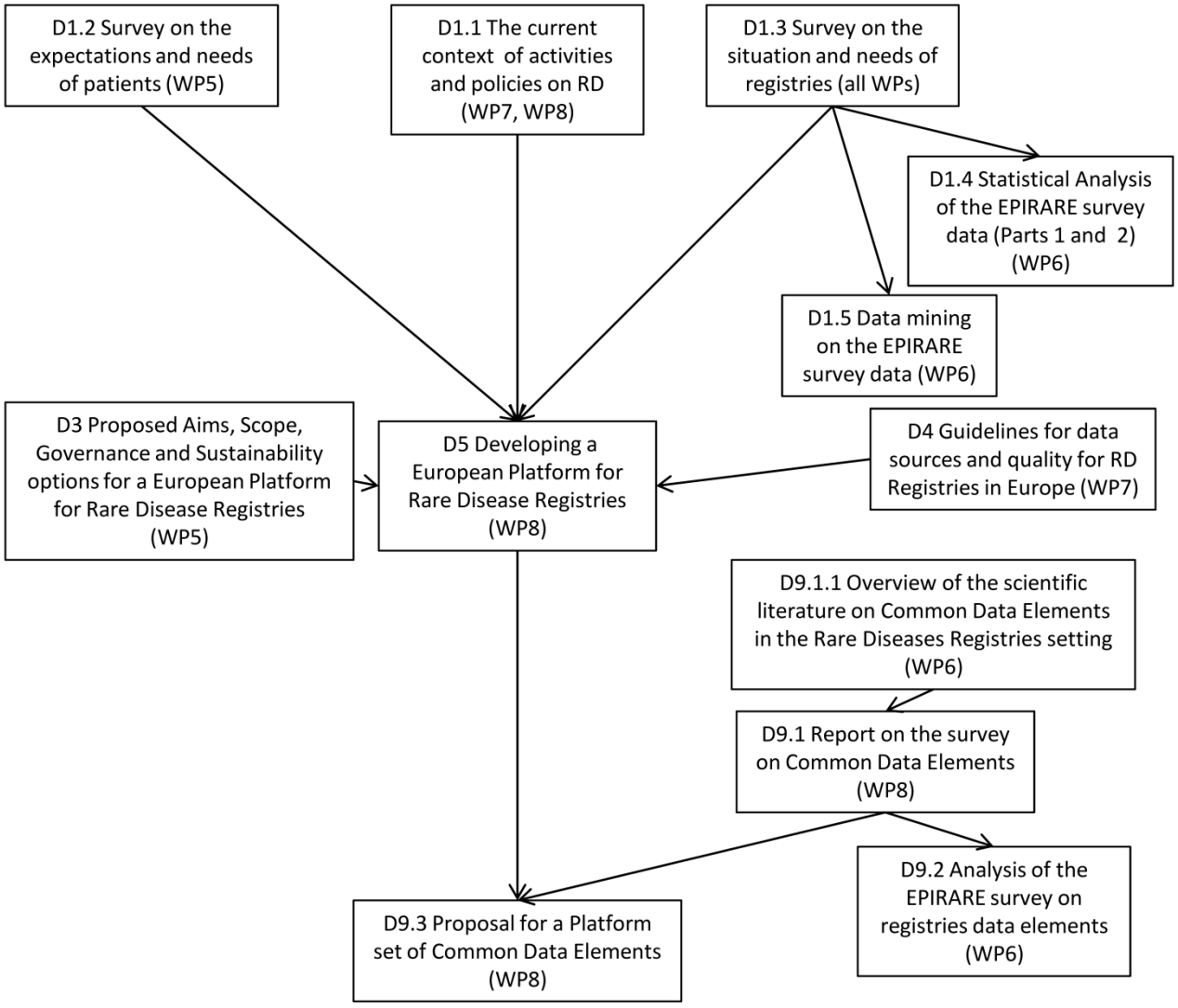
CONTENTS

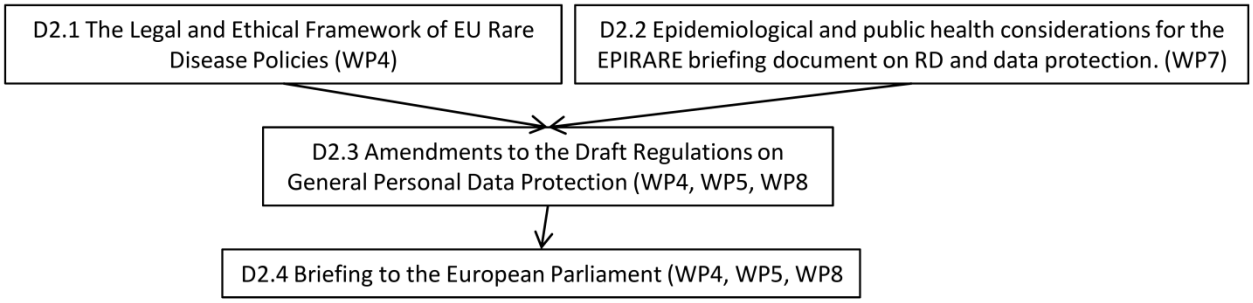
Acknowledgements

The EPIRARE documents have been circulated to extensive networks of experts with different interests and expertise. We are gratefully indebted with all those who have kindly contributed with their observations, comments and amendments. A special thank is deserved for all experts that have replied to the surveys, which have provided the factual evidence which allowed the formulation of the concept of the platform, the definition of the common data set and the assessment of their feasibility.

In particular for this document, the Authors wish to express their gratitude to Pilar Soler Crespo (General Directorate of Public Health, Quality and Innovation, Ministry of Health, Social Services and Equity – Madrid, Spain), Rosalia D'angelo (University of Messina Policlinic Hospital – Messina, Italy), Persephone Doupi (THL - National Institute for Health and Welfare – Helsinki, Finland), Wilhelm Kirch (Technical University Carl Gustav Carus, Medical Faculty – Dresden, Germany), Vitaliy Matyushenko (Kharkiv Charitable Foundation "Children with Spinal Muscular Atrophy" (CSMA) - Kharkiv, Ukraine), Matic Meglic (National Institute of Public Health – Ljubljana, Slovenia), Nicholas Nicholson (European Commission - Brussels, Belgium), Yaffa Rubinstein (National Institutes of Health, National Center for Advancing Translational Sciences, Bethesda (MD), USA), Metka Zaletel (National Institute of Public Health – Ljubljana, Slovenia).

Overview of the documents produced by EPIRARE





Disclaimer

The contents of this document is in the sole responsibility of the Authors; The Executive Agency for Health and Consumers is not responsible for any use that may be made of the information contained herein.

ACRONYMS

Acronym	Meaning
CDE	Common Data Elements
CT	Clinical Trial
EU	European Union
EUCERD	EU Committee of Experts on Rare Diseases
EuroCAT	Network for the epidemiological surveillance of congenital anomalies
EUROPLAN	European Project for Rare Diseases National Plans Development
GRDR	Global Rare Disease Registry
GUID	EU Global Unique Identifier
HGNC	HUGO Gene Nomenclature Committee
HGVS	Human Genome Variation Society
HRQoL	Health Related Quality of Life
ICD	International Classification of Diseases
ID	Identity
OD	Orphan Drug
PPV	Positive Predictive Value
RD	Rare Diseases
RDTF	Rare Diseases Task Force

Executive Summary

This document analyzes the data on the basis of the main features to be addressed in the data collection for the production of sound results regarding the predetermined platform indicators, which have been studied in the EPIRARE Deliverable 9.1, and presents a proposal of Common Data Set to be complied with by the registries which intend to be connected with the platform data repository.

I. Background

The EPIRARE "Survey on CDE"[1] showed that data necessary to compute indicators of potential interest to the platform stakeholders are collected with different frequencies by registries responding to the survey. The feasibility of each indicator was also assessed in that document on the basis of the frequency of occurrence of the set of data necessary for its computation.

The actual possibility to provide a sound and suitable evidence base, especially for the most complex indicators, depends not only on common definitions and detail and extent and formal control of the data collection, but may depend also on the use of longitudinal collections, completeness of case ascertainment and expert validation. Here we analyze the data on the basis of these additional requirements be addressed in the data collection for the production of sound results regarding the predetermined platform indicators and elaborate a proposal for the organization of the Common Data Set.

II. Specific features of groups of data elements

Besides the use of data elements for the computation of sound platform indicators and other information outputs, the results of the EPIRARE survey on CDE showed that some data elements have a particular importance for the best use of registry data. These comprise a) all the data elements necessary for the elaboration of an unambiguous universal patient coding, which, as discussed, could be: given name(s), family name, date of birth, city of birth, sex, country of birth and the national unique identification code; b) the data elements allowing indicator analysis by diagnosis, geographic area and setting: disease code (or disease name according to a reference list), longitudinal collections of patient city and country of residence and of treatment centre ID code (or name), city and country; and c) data for the ethical processing of patient data, including the informed consent, patient's willingness to participate in clinical trials and to donate biospecimens, and his/her contact data. This data, which altogether provide an important characterization of the patient, should receive a special priority in the platform set of common data elements. It should also be noted that registries collecting data of suspect patients (i.e. patients for whom a diagnosis of suspect RD has been formulated) provide important information regarding the operation of the health service in the confirmatory stage of the diagnosis.

The data elements related to patient exposure encompass information regarding very different fields, from risk factors, such as genetic variants and familial factors, environmental exposures, lifestyle and nutrition habits, to the provision of a variety of health services, from drug treatments to other treatments and procedures, embedded in different policy settings. Common (non disease-specific) data elements and indicators can be identified in this domain. However, the selection, by the registries, of data elements related to this heterogeneous domain is related to the features of the disease, depends on the aims of the registry and relies on the possibility for the registry holders to use the available data sources and collection techniques. Therefore it is very difficult to define a set of common data elements covering all possible data which can happen to be of interest for the description of RD patient exposures. The EPIRARE survey on CDEs addressed with some detail the provision of treatments by the health services and neglected the field of environmental and nutritional risk factors, which are of known relevance for registries studying congenital malformations, but are not usually recorded for most other rare diseases. On the other hand, the EuroCAT and cancer networks have been developing specialized registries for many years to monitor environmental exposures and are the reference for these types of data.

[1] EPIRARE Deliverable 9.1

Together with age at death, non disease-specific common data elements measured with generic questionnaires can be used to calculate the disability profile and the health-related quality of life (HRQoL) index score. Other data associated to outcomes are disease-specific clinical parameters, for which the possibility of defining common data elements for sensible comparisons across diseases is very limited, even within groups of related diseases. Some disease group-specific common outcome data may be collected by some of the registries participating in networks of related diseases, such as EuroCAT, Treat-NMD and cancer networks.

Information on the use of common coding systems, reference terminologies and questionnaires has been collected for very few data elements: the results obtained extend previous evidence obtained by EPIRARE, outlining a rather fragmented picture and suggesting that the adoption of common reference instruments will affect most registries.

III. Requirements of platform indicators

An analysis of the features that have to be fulfilled by registries aiming at different goals, ranging from population surveillance to service monitoring, health care, research, health promotion and regulatory assessments, has been carried out recently by Richesson and Vehik[2], with reference to the following requirements: Completeness of Case Ascertainment; Clinical data (beyond diagnosis or procedure); Expert Verification of Data Validity; Follow up Data (Longitudinal collections). The analysis of the indicators studied in this report using these criteria (Tab. 1), shows that the completeness of collection of cases is a requirement for almost all indicators studied: the only exceptions are the indicators of familiarity and the combination of data elements (exemplifying one of many possible criteria) for cohort selection and patient recruitment. Longitudinal observations are required for all those indicators that are related to the patient experience during her/his life. Validity control is necessary for indicators relying on clinical and genetic data, including those from disease-specific disability questionnaires. Finally, there might be many other indicators, here indicated with generic definitions, which require clinical data. These indicators will be of use, e.g., for research on the natural history of disease, healthcare quality or drug effectiveness.

This analysis shows the basic need for the platform to collect data on all cases belonging to a population in a defined geographical area with a longitudinal design. Most registries participating in the EPIRARE surveys indicated to be population-based and performing longitudinal observations. It would be important, however, that the platform undertakes actions to receive data from additional and independent sources to improve or assess the completeness of case registration. Clinical data and data validation processes play a central role to achieve more sophisticated objectives in the area of healthcare and research.

International classifications and coding systems should be used as reference as far as possible to facilitate the integration of the platform into global initiatives and, where applicable, those related to billing should be used in order to facilitate the economic analyses. Recommendations in favor of any one of these reference systems, which have to be agreed in view of international collaborations and of the many goals to be fulfilled by the European RDR Platform, should consider the availability of mapping tools to reduce the workload and loss of data associated with the conversion of already collected data. Table 2 reports an overview of international coding systems of relevance for the European RDR Platform. Moreover, linking the registry/EPIRARE platform data with other platforms, such as those for genomic and phenomic studies, as well as with the European Infrastructure for Spatial Information in the European Community[3], should be pursued as far as possible.

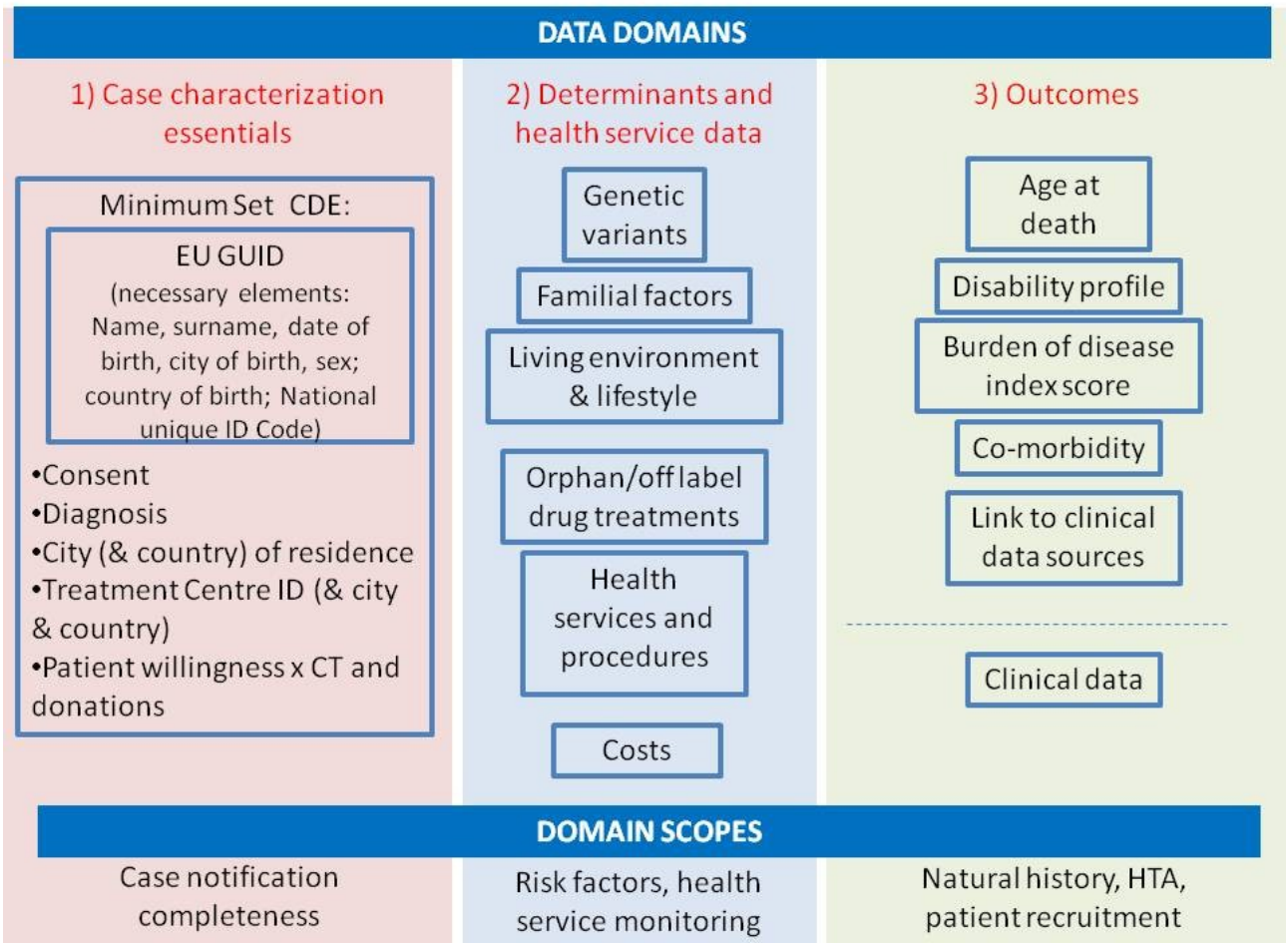
[2] Richesson R. and Vehik K. (2012) Patient Registries: Utility, Validity and Inference. In: Posada de la Paz M., Groft S. C. (Eds.), Rare Diseases Epidemiology. Adv. Exp. Med.Biol. vol 686. Pp. 87-104. Springer Science+Business Media

[3] Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)

IV. The platform data repository organization

The overall organization of the platform data repository is depicted in Fig. 1. The data elements are organized in three different data domains, which are characterized mainly by the data contents and sources, but are also functional to specific platform scopes: 1) Case notification completeness; 2) Risk factors detection and service monitoring; 3) Any application of outcomes analysis, such as natural history of disease, healthcare quality assessment and patient recruitment. In principle, the combination of data within the same domain supports the specific purpose of the domain; however, the combined analysis of data and indicators across domains may serve, as exemplified in the Report on the Survey on CDE5, the needs of different platform stakeholders, from basic epidemiological information to the monitoring and quantification of health service delivery; decision-making for services, marketing and research; cohort selection; planning of clinical trials; the natural history of the disease and clinical benchmarks.

Figure 1 – The platform data repository organization



The first domain (Tab. 3) aims mainly at facilitating the completeness of case notification, also ensuring the case identification, the geographical location of the patient and of the services involved in the patient treatment and informing on the patient position regarding participation in research. This data provides information on the patient distribution and problem dimension, and is of use for health services and clinical trial planning, for the prioritization of product development and for patient advocacy. **This is the minimum information necessary to characterize the case; therefore, it makes up the mandatory set of data**

elements. With very few exceptions, these data elements are collected by at least two-thirds of the registries in our survey. It is made of data which are in the knowledge of the patient (or their family) and which can be entered without the involvement of physicians or the health services which follow the patient. Therefore, this data set can support a notification process that is fully independent of any other source based on patient records or the active notification by physicians. Actually this data can be entered directly by the patient. Drop down menus and automatic checks of data entry correctness should suffice to ensure data accuracy. However, the assistance of patient association would be desirable to help and guide the patients in filling this data as well as to promote the notification of patients to the platform or to registries connected with the platform. While it is appropriate that notifications entered directly by patients are validated by experts with a known position within the network of the platform and the connected registries before they are included in the platform database, the validation process itself activates a process which streamlines patients to competent centres, which may complete their records with information belonging to the other domains of the platform repository and requiring specific expertise for its production and processing.

Combining data from multiple independent data sources, including direct patient notification, in spite of the need for its validation, and besides the possibility to extend the registered population, gives the important opportunity to estimate the degree of underreporting of the sources based on health services and get better estimates of true patient prevalence and distribution.

With regard to this domain, two important remarks should be made. The identifiers have been selected to facilitate the univocal identification following the results of Johnson et al[4]. However, as explained in the report of the Survey on CDE[5] it is considered necessary for EU and global registration, that EU registry sources collect two additional elements as further patient specification: the country of birth, as already done in the US-GRDR, and the national unique identification code. The other point is that the correct indication of the diagnosis requires the adoption of an agreed reference coding system or disease list, but this is not achieved at present. EPIRARE suggests the use of the ORPHANET list of diseases as the reference list, in the wait that the ICD11 is published and the ORPHA Codes are adopted internationally.

The second domain of the platform data elements (Tab. 3) aims at characterizing the patient risk factors, at monitoring and planning the operation of the health services and to quantify the associated costs. It extends the patient characterization with genetic data and with data regarding his/her health status and familial information. Moreover, this domain includes data regarding the history and status of diagnosis and treatments. This information can be collected with a variety of methods and requires specific methodological expertise for the data handling and use for health service monitoring and planning and for health service research. However, most of the data collected and of the purposes of collection do not require expert verification of the data validity.

Data pertaining to this domain would not be collected by all registries in all these fields. Indeed each registry should select the data elements which are relevant to the scope of its observations, adopt the definitions and formats proposed and collect the corresponding data, even if they are going to collect further detailed or specialized information. The results of our survey indicate that longitudinal data in any field of this domain are collected, according to proposed specifications, by about one-third of the registries, except for the genetic data of the patient and most data of the history of diagnosis, which are collected by more than two-thirds of the registries. Information regarding the history of diagnosis of suspect patients also is collected longitudinally by about one-third of the registries. These relatively low frequencies indicate essentially that the currently existing registries have specialized interests, but that, altogether, they ensure a rather homogeneous monitoring of different sections of the health service.

[4] S B Johnson, G Whitney, M McAuliffe, H Wang, E McCreedy, L Rozenblit, C C Evans (2010): Using global unique identifiers to link autism collections. J Am Med Inform Assoc 2010;17:689-695. Available on line at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3000750/pdf/amiajnl2063.pdf (accessed on 9 August, 2012)
[5] EPIRARE Deliverable D9.1

The standards and terminologies to be used in the platform should be agreed with clinical, epidemiological experts and, possibly, involving representatives of EU national information systems; the selection process should take into account the existing EU legislation and agreements, the guidelines developed by EUCERD as well as the interoperability with already established initiatives, the availability in EU languages, the licensing conditions, and the existence of tools allowing mapping among different standards and terminologies, such as those reported in Table 2 for laboratory test, procedures drugs and devices.

The third domain (Tab. 3) aims at supporting outcome analysis, with either providing connections to sources of disease-specific data, which is functional to patient cohort selection, or supporting disease-specific modules for diseases which cannot be followed with dedicated registries, or collecting data of disability and HRQoL for integrated assessments across diseases. According to the survey results, with the exception of the date of death, which can be collected by more than three-fourths of the registries, longitudinal disability and HRQoL data can be collected by one-fourth or less of the registries with heterogeneous tools. In comparison, the longitudinal collection of outcome data based on disease-specific clinical parameters can be practiced by about two-thirds of the registries. Therefore, it appears that, at present, outcome information can rely mostly on disease-specific clinical parameters, which need, as shown in the first EPIRARE survey of registries, a process of data validation and quality assessment. However, validation and other actions aiming at data comparability might have been already done by some networks of registries. It is expected that, in the future, disease specific data, although varied from disease to disease and unique to each registry, can be collected and organized under more defined health related domains, because rare diseases are often affecting multi organs and many of them share the same symptoms.

The assessment of disability and HRQoL are extremely important since many RD are not impacting on the lifetime and can serve many purposes, from patient-centered description of the disease course, to monitoring the impact of policies and best practices, and to equity decisions based on assessments cutting across all diseases. These achievements are hampered by heterogeneity of the instruments used for these measures and by the experience that generic disability and HRQoL measures are subject to many bias or are not suitable to capture relevant changes. However much work has been carried out recently by the international networks to validate questionnaires for patient reported outcomes in these fields[6]. Therefore, the extensive collection of outcome measures common to all diseases, based on disability and HRQoL, requires a substantial effort consisting mainly of expert reviewing the validation studies available, for promoting the agreement on reference tools, and of extending the actual collection practice.

The arrangement of data elements in Table 3 is indicative of the information details that can be covered by the Common Data Elements and refer to the different domains depicted in Fig.1. However, these data elements are to be arranged in different ways according to the structural organization of the databases designed for longitudinal data collections or according to the case report form used for the first collection of data of patients, which may be made of different modules depending on the need for separate data inputs from different specialists. These arrangements are not developed here.

V. Conclusions

In conclusion, this report indicates the important role that can be played by population-based and longitudinal data collections. Since not all these features may be present in the same registry, an important aspect of the platform is to facilitate interoperability and data merging among the different registries, promoting the use of common tools and standard terminologies and the collection of comparable data, including some data, such as on the patient willingness to participate in clinical trials or to donate biospecimens for research, which is currently neglected. Another important goal of the platform is to

[6] Rajmil L., Perestelo-Pérez L. and Herdman M. (2012) Quality of Life and Rare Diseases in Rare Diseases Epidemiology (Eds.: M. Posada de la Paz, S.C. Groft), Advances in Experimental Medicine and Biology 686, pp. 251-272

contribute to the completeness of case ascertainment and to liaise with different sources and, in particular, with patient associations, which have direct contact with patients and may have knowledge of patients who are not registered, thus representing an additional independent information source.

With the establishment of this collaborative network of registries and patient associations, it appears that the collection of data pertaining to the domains of case notification and of risk factors and health services can be in the reach of the platform after a short period of operation dedicated mainly to quality control. Support for research and product development studies also can be provided in a short time by utilizing the registries collecting the disease-specific data of interest; on the other hand, support for public health goals and equitable decision-making, which requires mostly the use of outcome data comparable across all diseases, would require longer implementation times. However, the availability of both disease-specific and generic, non disease-specific, outcome data would best fulfill, with their different features and applications, the information needs of research in the public health, individual patient care and basic biology of the rare diseases.

TABLES

Table 1 – The requirements of selected platform indicators

Registry-based Indicators and other measures	Requirements			
(reference of indicators to disease, time period and geographic area of interest are implicit)	**Completeness of Case Ascertainment**	**Clinical data (beyond diagnosis or procedure)**	**Expert Verification of Data Validity**	**Follow up Data (Longitudinal collections)**
Incidence, per disease and global*	y	n	n	n
Age at death**	y	n	n	n
age at disease onset**	y	n	n	n
time from disease onset to confirmed diagnosis**	y	n	n	n
time from 1st report to the health service to confirmed diagnosis**	y	n	n	n
Activity of centres actually making diagnosis (diseases diagnosed and number of diagnoses per year)§	y	n	n	n
Number and directory of centres actually making diagnosis§	y	n	n	n
life expectancy at diagnosis***	y	n	n	n
effectiveness of neonatal screening programs (positive predictive value)§	y	n	n	n
effectiveness of neonatal screening programs (sensitivity)§	y	n	n	n
number of RD actually diagnosed (and recorded) per Country and per Centre§	y	n	n	n
patients' mobility for diagnosis***	y	n	n	n
Prevalence, per disease and global*	y	n	n	y
other cases in the family**	n	y	y	n
healthy carriers in the family**	n	y	y	n
case parents are consanguineous**	n	y	y	n
Hospital admissions*	y	n	n	y
Activity of treatment centres (diseases treated and number of treated patients per year)§	y	n	n	y
Number and directory of treatment centres§	y	n	n	y
number and types of tranplantations**	y	n	n	y
ODs actually used (and recorded)§	y	n	n	y
number of patients treated per OD per year**	y	n	n	y
number of patients treated per ODs§	y	n	n	y
number and type of surgeries recorded***	y	n	n	y
patients' mobility for treatment***	y	n	n	y
disability profile**	y	n	y	y

burden of disease**	y	n	y	y
indicators supporting cohort selection and patient recruitment for CT (one example given)***	n	y	y	y
indicators based on disease specific clinical data (e.g. clinical care benchmarks)	y	y	y	y

*These indicators were considered by the RDTF particularly important for surveillance of status and trends
§ EUROPLAN indicators

**These measures are the proposed alternates to the indicators considered by the RDTF

***These are additional measures of which the registry-based feasibility is studied in this report

Table 2 - International Coding systems and terminologies

Area	System	Author	Web-site	Remarks
Medical Nomenclature	SNOMED	International Health Terminology Standards Development Organization	www.ihtsdo.org/snomed-ct	ORPHA-codes are being integrated in SNOMED.
Diseases	ICD-10-CM ICD-9-CM	WHO	www.who.int/classifications/icd/en	Billing-related. The coding of rare diseases in the next ICD-11 will be based on the ORPHA-codes
Rare Diseases	ORPHA-codes	ORPHANET	www.orpha.net	ORPHA-codes are being integrated in SNOMED and will be the basis for the codification of rare diseases in the next ICD-11.
	UMLS	NIH ORDR	https://grdr.ncats.nih.gov/index.php?option=com_content&view=article&id=91&Itemid=160	This is the system used by the US GRDR and may be useful for interoperability with this platform.
Genes, genetic disorders and traits	Online Mendelian Inheritance in Man (OMIM)	McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD)	http://omim.org/	
Genes	HGNC	Human Genome Organization (HUGO)	www.genenames.org/aboutHGNC.html	
Genomic variations	-	Human Genome Variation Society	www.hgvs.org/mutnomen/	
Laboratory tests and results	LOINC	Regenstrief Institute for Halth Care	www.regenstrief.org/loinc/	
Procedures	ICD-10-PCS ICD-9-CM Vol. 3	WHO	www.who.int/classifications/icd/en	Billing-related
Devices	Global Medical Device Nomenclature (GMDN)	GMDN Maintenance Agency	http://www.gmdnagency.com/	Supports the European Databank for medical devices foreseen by the EU Medical Device Directive. It includes 20 EU languages.
	Universal Medical Device Nomenclature System (UMDNS)	WHO Collaborating Centre ECRI	https://www.ecri.org/Products/Pages/UMDNS.aspx	The National Library of Medicine has included UMDNS in the Unified Medical Language System.
Drugs and Orphan Drugs	ATC/DDD Index	WHO Collaborating Centre for Drug Statistics Methodology	http://www.whocc.no/atc_ddd_index/	

	MedDRA (Medical Dictionary for Regulatory Activities)	International Conference on Harmonization (ICH)	http://www.meddra.org/	
Adverse Reactions	WHO-ART	WHO, maintained by the Uppsala Monitoring Centre	http://www.umc-products.com/DynPage.aspx?id=73589&mn1=1107&mn2=1664	
	EU SPC ADR database	EMA	http://www.imi-protect.eu/methodsRep.shtml	Database of all adverse drug reactions (ADRs) listed in section 4.8 of the Summary of Product Characteristics (SPC) of medicinal products authorised in the EU according to the centralised procedure. It is based exclusively on MedDRA terminology.
	MedDRA (Medical Dictionary for Regulatory Activities)	International Conference on Harmonization (ICH)	http://www.meddra.org/	
Disability	ICF	WHO	http://apps.who.int/classifications/icfbrowser/	Billing-related. Available in English, French and Spanish. A Children and Youth version is also available in English only

Table 3 – The Platform set of Common Data Elements

		COMMON DATA ELEMENTS collected in the EPIRARE platform **(elements in bold require longitudinal data collection)**	ANNOTATIONS regarding the data elements; Where indicated: DEFINITIONS and FORMATS, the use of which was investigated in the EPIRARE Survey of data elements used by Registries.	REASON
Domain 1) Case characterization essentials	Mandatory data	EU Global Unique Identifier (EU GUID)	This code is elaborated from the following data elements: • Patient given name: DEFINITION: "First name of patient as recorded in birth certificate, passport or identity card"; FORMAT: full name, not initials • Patient family name (at birth): DEFINITION: "Family name of patient as recorded in birth certificate, passport or identity card"; FORMAT: full name, not initials • Patient sex: see definition below • Patient date of birth: see definition below • Patient city of birth: see definition below • National Unique Identification Code	Unambiguous patient coding (to be processed according to legal provisions) is necessary to keep the integrity of the database and avoid duplication of records. The National Unique Identification Code increases the accuracy of the EU GUID in case of names in foreign languages. It could be an optional part of the encrypted code.
		Patient sex	DEFINITION: "Patient's physical sex at birth"; PERMISSIBLE VALUES: male, female, other (in any format)	Allows studies of sex-related differences in the disease epidemiology and clinical features
		Patient date of birth	DEFINITION: "Date of patient's birth recorded in birth certificate, passport or identity card"; FORMAT: complete date (year, month, day) in any format For privacy reasons, depending on the time course of the disease, this data is to be communicated to the platform at the appropriate level of precision (only month and year or complete)	Allows studies of age-related disease features.
		Patient city of birth	DEFINITION: "Name of city/town/village where the patient was born as it appears on the birth certificate, passport or identity card"; FORMAT: full name of city. For privacy reasons, this data is to be communicated to the platform with the appropriate level of precision (e.g. mapped to the province, or to postal	This data may be communicated to the platform only for some specific diseases for studies of health determinants.

			code). Moreover, it is important that geographical names are mapped to the INSPIRE identifiers. This will enable the link with platforms organized around environmental spatial information, such as environmental pollution databases. This may offer an additional opportunity to indicate the place with an appropriate granularity to comply with privacy needs.	
		Patient country of birth	DEFINITION: "Name of country where the patient was born as it appears on the birth certificate, passport or identity card"; FORMAT: full name of country	Increases the discriminatory power of the EU GUID in global registries
		Diagnosis	ORPHANET list of diseases	attribution of a disease to the case
		Patient city of residence	DEFINITION: "Name of city/town where the patient usually lives"; FORMAT: full name of city For privacy reasons, this data is to be communicated to the platform with the appropriate level of precision (e.g. mapped to the province, or to postal code). Moreover, it is important that geographical names are mapped to the INSPIRE identifiers. This will enable the link with platforms organized around environmental spatial information, such as environmental pollution databases. This may offer an additional opportunity to indicate the place with an appropriate granularity to comply with privacy needs.	attribution of the case to a geographic area; prevalence, incidence, mobility
		Patient country of residence	DEFINITION: "Name of country where the patient usually lives"; FORMAT: full name of country	attribution of the case to a geographic area; prevalence, incidence, mobility
		ID Treatment Centre	Treating Centre Full name/code; contact data are optional to improve identification	attribution of the case to the treating setting
		Treating Centre City-Town	FORMAT: full name of city It is important that geographical names are mapped to the INSPIRE identifiers.	attribution of the centre to a geographic area; patient mobility for treatment; planning research/clinical trials
		Current and past participation in clinical trials	Yes/No	planning research/clinical trials
		Patient willingness to be contacted to participate in a future clinical trial	Yes/No	planning research/clinical trials
		Patient willingness to be contacted about donating biological samples	Yes/No	planning research/clinical trials
		Patient consent	based on graduated consent forms	
		Patient contact	contact details; preferred means of contact (including via intermediary physician); language	

Domain 2) Determinants and services	Case characterization	**Education level**	Values from 0 to 8, based on the ISCED 2011 classification	Studies of socio-economic burden. Comparison and matching of patient populations from different data sources on the basis of socio-economic data. Applicable to individuals from early childhood.
		Occupation	Self-defined current economic status (PL031 EU-SILC Target Variable): 11 possible values. (http://epp.eurostat.ec.europa.eu/portal/page/portal/income_social_inclusion_living_conditions/documents/tab/Tab/Personal%20data%20-%20labour.pdf)	Studies of socio-economic burden. Comparison and matching of patient populations from different data sources on the basis of socio-economic data. Applicable to individuals more than 16 year old.
		healthy carrier	Yes/No	
		other cases in the family	Yes/No (If Yes: degree of kinship)	Contribution of consanguinity; socio-economic burden of disease
		healthy carriers in the family	Yes/No (If Yes: degree of kinship)	Contribution of consanguinity
		case parents are consanguineous	yes/no	Contribution of consanguinity
		Biomaterial donated	(Yes/no); If Yes: list to be defined (e.g. Tissue or body fluid or other specifications)	planning research/clinical trials
		ID Biobank where the biological sample is stored up	Biobank Full name/code; contact data are optional to improve identification	Link to Biobanks; planning research/clinical trials
		(if the biobank storing the sample is not known) ID Centre which sampled the biomaterial	Sampling Centre Full name/code; contact data are optional to improve identification of the centre	Link to Biobanks; planning research/clinical trials
	History of diagnosis	Genetic features patient	Gene-HGNC Gene Symbol	Link to genetic research platfomrs; patient cohort selection
			Chromosome number	
			Nucleotide sequence analyzed and reference sequence systems with accession and version number	
			Variant description in HGVS format	
			Variant description in other formats	
		Date of first symptoms onset	DEFINITION: "Date when patient first began experiencing symptoms or signs of the rare disease"; FORMAT: complete date (year, month, day) in any	age at onset; time to diagnosis

			format	
		Date of first contact of patient with the public Health Service	Date of the first time the patient	time to diagnosis
		ID Centre/physician referring the patient to the RD centre	Centre/Physician Full name/code; contact data are optional to improve identification	integration of RD centres in the general Health Service
		Date of current diagnosis	DEFINITION: "Date when the current rare disease diagnosis was made" FORMAT: complete date (year, month, day) in any format	time to diagnosis; life expectancy at diagnosis
		status of current diagnosis	suspected-confirmed	diagnostic patterns; time to diagnosis; life expectancy at diagnosis
		methods used for current diagnosis	list to be defined	diagnostic patterns
		ID Centre which made diagnosis	Centre Full name/code; contact data are optional to improve identification	
		Centre which made diagnosis City-Town	FORMAT: full name of city It is important that geographical names are mapped to the INSPIRE identifiers.	patient migration for diagnosis
		Patient referred after positive neonatal screening result	Yes/no	sensitivity and PPV of neonatal screening tests; effectivenes of neonatal screening program
	Treatments and services	**Current orphan drug treatment**	DEFINITION: "A list of all current orphan drugs that a patient is currently taking"; FORMAT: name of all active ingredients (ORPHANET list)	
		Current off-label drug treatment	DEFINITION: "A list of all current drugs (different from orphan drugs) that a patient is currently taking"; FORMAT: name of active ingredients	
		Current drug treatment	DEFINITION: "A list of all current drugs (different from orphan drugs) that a patient is currently taking"; FORMAT: name of active ingredients	
		Hospitalizations	DEFINITION: "Cumulative number of patient's admissions to the hospital due to the rare disease"; FORMAT: number	
		Transplantations	Yes/No (If yes: date of transplantation; tranplant material)	
		Surgeries	Yes/No (If yes: date of surgery; ID code of Surgery)	
		Current dietary regimens prescribed as treatment	Yes/No (If yes: type of regimen)	
		Current assistive devices	Yes/No (If Yes: Type of assistive devices used by patient; ID Code of type of device.	
		Other treatments	If Yes: Type/Code of treatment; indicate if current or date of administration	
Doma in 3)		**Patient vital status** (and date of death)	Live/Dead (If Dead: complete date of death (year, month, day) in any format Required Sources: National Death Registry or National Population Registry	

		Patient disability profile	Patient disability generic and domain-specific questionnaires (modules) with separate recording of domain scores	patient disability profile and disease course
		Patient HRQoL index score	Patient health-related quality of life generic questionnaires with calculation of the utility score	assessment of burden of disease; QALYs; equitable decision-making
		Comorbidity	DEFINITION: "Other diseases observed in the patient"; FORMAT: ICD10 (ORPHA-codes in case that other RD are observed)	
		Remarkable or unusual symptoms	Remarkable or unusual symptoms, including adverse effects of treatments, and their severity (based on a 5-degree scale).	
		Metadata of disease-specific data and data sources	ID, metadata and contacts of registries, clinicians or other sources collecting disease-specific data of the patient and description of data collected	Facilitate tracing of existing disease-specific data on the patient

Appendix

Some tools available for patient reported outcomes

Disability and QoL indicators, based on generic questionnaires developed for Patient Reported Outcomes (PRO), address different health domains. Many tools are available, and it is necessary to review them and to agree on the questionnaire(s) which is (are) more suitable for the aims of the platform. The NIH GRDR has adopted the NIH-developed PROMIS[7] SF General Health module and four additional modules specific for physical functioning, pain, fatigue and depression. PROMIS SF tools are modular questionnaires with different lengths which have been extensively validated. However, their availability in different languages relevant to EU is varied. NIH has also been involved in a project with WHO for the development of the WHO Disability Assessment Schedule 2.0 (WHO-DAS 2.0)[8], a questionnaire recently developed in a number of versions and languages on the experience of a previous WHO instrument (WHO DAS II). It is grounded on the conceptual framework of the ICF dimensions and produces domain-specific scores (the disability profile) for six different functioning domains: cognition, mobility, self-care, getting along, life activities (household and work) and participation. Domain-specific scores and a single summary score can be calculated with both these tools. Some training is necessary to fill the questionnaires; however, both of them are produced in different versions for self administration, for administration by an interviewer or for administration to a patient's proxy. Therefore it might be possible that trained personnel in patients associations assist the patients and their relatives in filling these forms to improve consistency among responses.

As to the available questionnaires for HRQoL, a recent review[9] has evaluated some questionnaires used in children with reference to reliability, validity and sensitivity to change. The KIDSCREEN[10] and DISABKIDS[11] projects, funded by the EU within the FP5 and the Public Health Programme (2001-2004), have developed different generic questionnaires in different versions suitable for children and adolescents. Generic comprehensive HRQoL questionnaires to calculate summary scores (utilities), such as the EQ-5D[12] and the HUI[13] have also been developed and extensively validated for the calculation of QALYs in children and adults of different populations. Moreover the experience of a EU funded project (BURQOL-RD) can help identifying the best instrument to assess HRQoL in RD patients.

[7] http://www.nihpromis.org/
[8] http://www.who.int/classifications/icf/whodasii/en/index.html
[9] L. Rajmil, L. Perestelo-Pérez, and M. Herdman: Quality of Life and Rare Diseases. In: M. Posada de la Paz, S.C. Groft (eds.), Rare Diseases Epidemiology, Advances in Experimental Medicine and Biology 686 (2010)
[10] http://www.kidscreen.org/english/questionnaires/
[11] http://www.disabkids.org/questionnaire/disabkids-core-instruments/
[12] http://www.euroqol.org/about-eq-5d.html
[13] http://www.healthutilities.com/