Work Package 5

Deliverable 5.2

Standard procedure and guide for the coding with Orphacodes

Issued on the 30st of May 2017

by the WP5 members of the RD-ACTION European Joint Action

Co-funded by
the Health Programme
of the European Union

This report is part of the project / joint action ‘677024 / RD-ACTION’ which has received funding from the European Union’s Health Program (2014-2020).

It has been produced by the co-leaders of the Work Package 5 and is part of the Task 5.1: To define and set the necessary strategy and tools to implement the Orphacodes in the European countries (Task Leaders: Remy Choquet and Ferdinand Dhombres [BNDMR, APHP, France] - Contributors: all WP5 contributors).

It has been reviewed by WP5 participating countries within an open for comments phase.

The RD-ACTION Joint Action was launched in June 2015 for a 36 months period.

More information on the activities of the RD-ACTION Joint Action can be found at www.rd-action.eu.

Disclaimer:

Summary

Underrepresentation of rare diseases (RD) in coding systems determines a general difficulty in tracing RD patients' pathways within healthcare systems. This issue is crucial as it affects the possibility of estimating the global number of persons living with RD and their access to healthcare services. This data paucity is perceived as a relevant issue at multiple levels: by patients, by researchers and clinicians, as well as by national/regional health authorities, responsible for health planning activities and the allocation of resources, human, technical and economic.

In order to tackle this issue Orphanet started to classify rare diseases since 2007, adopting a poly hierarchical approach. In this effort, still ongoing, each clinical entity in the Orphanet nomenclature is being assigned a unique and stable identifier, the Orphanumber. The subset of Orphanumbers that is applicable for coding patients is called the Orphacodes. The Orphanet inventory of rare diseases is continuously updated as new RD clinical entities are discovered. This work has been the basis of the World Health Organization (WHO) decision to establish a specific Topic Advisory Group for RD, coordinated by Orphanet, in order to achieve a better representation of RD in the International Statistical Classification of Diseases and Related Health Problems, 11th revision (ICD-11).

The WP5 of the RD-Action aims to support the implementation of a standardized and consistent way of coding RD using Orphacodes across Europe. Therefore, coding procedures, guidelines and tools will be developed, that are taking into account existing approaches in coding systems and guidelines in Member States (MS).

A review about the existing technical implementations for RD coding in MS, which was published in May 2016, showed, that most MS use ICD classification for coding morbidity and mortality. Only few MS started to use Orphacodes to code RD as a national and routine coding resource. Based on the results of this review, the WP5 started to develop strategies and tools in order to promote and support a consistent coding of RD across MS.

In this document, general rules for routine coding with Orphacodes are presented and guidelines are given in order to achieve internationally standardized data collection. Given the different data collection settings and purposes, additional rules and guidelines might be needed.

The guidelines are specified throughout the document along with the explanation for why each guideline is given. At the end of this document a summary of the guidelines is given for easy reference and implementation.

1. GENERAL CONSIDERATIONS ABOUT CODING FOR RARE DISEASES

The aim of the WP5 activities is to define rules and tools that can support the use of Orphacodes at Member States (MS) level in order to increase rare diseases patients' visibility across health information systems and to estimate the impact of rare diseases on social and healthcare systems.

Every policy regarding the implementation of Orphacodes depends not only on the characteristics and on the structure of the monitoring systems to be developed, but also on a variety of background elements that differ across MS. Among these factors, we can list the following:

- the different health care organizations and modalities to access the services
- the different organization of health statistics including the sources, the information pathway, the levels of aggregation, the prevalent uses, the availability of registration systems tracing patients' episodes of care, the existence of a unique patient identifier, different legislations in use dealing with privacy issues and data protection.

These background factors existing in each country affect the feasibility, the cost and the level of complexity of the actions required to implement monitoring systems using Orphacodes. On the other side, they help to drive the choice of monitoring system and influence its design and development.

The monitoring systems can be organized according to different designs. The difference can rely on the following domains:

1. Who (or what) is the object of the registration? This issue deals with the choice of what should be the statistical unit object of the monitoring activity. It can be an event (i.e. the hospital admission, the ambulatory access, a medical intervention, etc.) or a case, for example a patient when a rare disease diagnosis is formulated.
2. Who is the healthcare provider (or the professional) in charge of recording the information referred to the statistical unit (event or person)? It can be the clinician when a diagnosis is performed, the coder of the hospital discharge record, the geneticist working in a laboratory when the result of a diagnostic test is reported.
3. Which are the sources of the collected information and of the reporting activity? They can be the health care providers acting as Centers of expertise for groups of RD, or other hospitals, outpatient clinics, etc.
4. How is the information collected, using which classification and coding system and in which way it is used?

All these elements and the related choices constitute issues tackled by the RD-Action, and in particular by WP5, with the aim of defining rules and propose tools taking into account the different background situations existing in MS. Nevertheless, it has to be underlined that there is no ideal tool and/or rule applicable to all the situations of different MS.

Although situations may be heterogeneous, it is necessary to use a common vocabulary among MS to describe the rules defined through the RD-Action activities. From this perspective, we have proposed definitions for some basic concepts (classification, coding system) constituting the pillars for the adoption of Orphacodes within health information systems.

A classification can be defined as a logic tool that systematically groups a number of entities in categories or groups based on predefined logic criteria. A classification can be hierarchical or non-hierarchical, depending on its ability to be used to create broader or narrower categories, depending on the chosen level of granularity. The upper level includes all concepts present in the lower categories. Only hierarchical classifications allow exploiting different levels of granularity.

A code is a sequence of symbols, which can be qualitative attributes (i.e. colors), numeric or alphanumeric sequences. They univocally identify a defined entity or a category, independently from the level of aggregation or granularity considered. Thus, the code is univocal for each entity or category.

All the subjects or statistical units described by the same code can be grouped together because they belong to the same category described by that code. Codes as well can be hierarchical or not. Codes can be defined hierarchical when the level of granularity or aggregation of the entities included in the same category is directly described by the number or the sequence of symbols used.

Considering the structure of the classification and the codes, both can be hierarchical, as in the ICD, or the classification can be hierarchical, but not the codes, as is the case of the Orphanet classification. In other cases, both the classification structure and the codes can be non-hierarchical, as is the case of the OMIM (Online Mendelian Inheritance in Men) terminology[1].

The rationale for the use of Orphacodes within dedicated monitoring systems (or within more general health information systems) derives from the necessity to make rare disease patients more visible within data collections describing the health status of a population and/or across sources of data referred to the use of healthcare services. This need is increasingly perceived as urgent as currently used classification systems in European countries and worldwide, namely ICD, do not efficiently serve this purpose at the moment. Due to the under-representation of RD described by specific codes as well as possible bias observed in its current use for

[1] https://www.omim.org

DRG systems which is headed towards cost efficiency rather than epidemiological purpose (the use of a diagnosis code might represent the reason of admission and expense of health services rather than the rare disease – e.g. Renal failure rather than Cystinosis or Pulmonary distress rather than Cystic fibrosis). This limit might be reduced with ICD-11, but the effect of this effort will only be visible in the very long period of time and appropriate coding guidelines should be produced regardless of the availability of codes for rare diseases nonetheless. The need to develop specific classifications is common in other domains, for example specific classification systems have been developed for group of disorders or events, i.e. DSM for psychiatric disorders, ICD-O for oncological diseases or ICECI for injuries, etc.

At European level, the lack of data on RDs in healthcare systems represents a major issue to be addressed. Accordingly, recommendations have been issued stating the importance of using Orphacodes, together with ICD, for recording RD patient data in health information systems. More specifically, the recommendation on "*Ways to Improve Codification for Rare Diseases in health information systems*" concluded: "*MSs should consider adding Orphacodes to their country's health information system and explore the feasibility and resources needed to do so*". As health information systems in Europe mostly rely on ICD, in its different versions, the first question is how to logically integrate these two classifications.

Disease classifications can be used together with disease nomenclature for various use cases. The main identified use cases are the following for Orphanet products:

1. Coding assistance: navigate through diseases classifications, depending on physician specialty (as in LORD[2] which uses Orphanet classifications per medical specialty), differential diagnosis process,
2. General/unprecise coding of patients: whilst searching for patient precise rare disease diagnosis, upper-level categories could be used for temporary patient coding (e.g. a *Malformation syndrome with skin/mucosae involvement*),
3. Precise description of phenotypic forms of the disease (e.g. a *Cystinosis* case as a *Metabolic disease with corneal opacity* defect),
4. Statistical tool: to regroup number of affected patients by group of diseases.

Although the use of Orphanet classifications is optional to the routine coding, it should be carefully done when used for statistics. Both ICD and Orphanet classifications, while being both hierarchical, differ in one important structural aspect: multi-axiality. The ICD is unidimensional, whilst the Orphanet classification is multidimensional. In the ICD, an entity has a single position in the hierarchy, independently from possible additional available information. The Orphanet classification is multi-hierarchical; each entry is classified in one or more hierarchies, usually following the organization of medical specialties and in one or more sections

[2] http://enlord.bndmr.fr

of a single hierarchy. This situation, describing the relationship between entries, has been defined “multiple parentage”.
As a consequence, a disease is described in the ICD by a code belonging to a macro-category, besides the phenotypic presentation in a particular patient. In addition, the use of a decimal coding system in ICD does not allow having many separate codes for each disease included in a group. This loss of detail may not interfere with other ICD purposes and uses, but is limiting when considering RD coding. This affects also the possibility to capture the wide phenotypic variability in terms of clinical presentation considering different patients with the same diagnosis or one patient over time.

The Orphanet classification, due to its multidimensional approach, allows an entry to belong to different macro-categories, i.e. different branches of the classification. This depends on the phenotypic variability in the clinical presentation of a disease. Rare disorders are often multi-systemic; thus the same RD entity can belong to different macro-categories, generally following an organization into body systems.

As a consequence of this difference in the structure of the two classifications, codes unlike classifications are hierarchical in ICD but not in Orphanet. ICD is built to enable grouping of codes for statistics within its code structure whereas using Orphanet classifications for statistics requires extra information such as linearization. Orphanet provides a default linearization of the classification, but each user could define his own linearization depending on his needs. Also, when the relation between the Orphacode and the ICD code is set, projection is possible through ICD logical structure for statistics. Taking into account all these characteristics, integration between Orphacodes and ICD could be achieved through two different approaches. The first is to follow the unidimensional logic and the hierarchy of ICD, integrating its categories with a series of included rare entities, not explicitly reported in the classification as terms at the moment, starting from the “flat” list (the nomenclature) of rare diseases entities produced and maintained by Orphanet (without considering the levels of intermediate branches).

A second possible approach is to follow a combined or multidimensional logic, that allows integrating the categories of the ICD with categories or entities that can be aggregated in different branches according to the prevalent considered phenotypic manifestation.

The product resulting from the first approach is a table, represented in an electronic sheet. To achieve the product resulting from the second approach, a mapping exercise establishing relations many-to-many between ICD entities and Orphanet ones has to be carried out. This work has been already carried out by Orphanet, according to defined rules[3]. This mapping activity has considered ICD-10, but we should take into account that in some European countries the ICD is not used in its

[3] http://www.orpha.net/orphacom/cahiers/docs/GB/Orphanet_ICD10_coding_rules.pdf

10th revision, but in other previous revisions or modifications, namely ICD-9 and ICD-9-CM, especially for morbidity recording.

To achieve integration between the ICD and the Orphanet classification the following approaches can be followed in the concrete implementation in health information or monitoring systems covering many MS:

- A one-to-one relation: an Orphanet code maps directly to one ICD code;
- A one-to-many relation: an Orphanet code maps directly to many ICD codes;
- A many-to-one relation: many Orphanet codes map directly to an ICD code;
- A many-to-many relation: an Orphanet code can be mapped to many ICD codes and vice-versa.

Besides these critical aspects deriving from the background analysis of each MS situation and from the analysis of the classification and coding tool, the concrete implementation in a real-world setting of monitoring systems using Orphacodes requires necessarily the consideration of the following issues:

1. How to guarantee, across different settings and countries, the quality and a homogeneous approach in the process of assigning a specific diagnosis to a patient and in the process of abstraction, on which the coding activity is based?
2. Which is the correct correspondence between nomenclatures (ie. the naming of pathological entities) and between classifications (ie the hierarchical organization of pathological entities) used in different countries?
3. Which is the setting in which the data are generated: the diagnostic/clinical setting or the administrative one (i.e. for reimbursement purposes)?
4. Which is the preferred moment for the collection of the information? At the moment in which the diagnosis is performed or even later, during the disease course, tracing the individual datum, even if not nominative?
5. Which is the organization of the coding system? Independent (double coding) or interconnected (joint coding)?
6. How to reduce the variability between coders and increase the level of completeness and quality of the data?

Finally, the real implementation of every monitoring system depends on a preliminary accurate analysis - performed MS by MS - of the additional costs required to support the implementation of Orphacodes, in terms of resources (technical and professional) and of the expected results. These results can deal, depending on the cases, with one or more of the following aspects:

- the increase in the quality and accuracy of health statistics serving the policy and the health planning process;
- the improvement of the reimbursement systems;
- the evaluation of the performance of the healthcare services and of the patients' clinical pathways;

- the improvement of epidemiologic and clinical knowledge on RD, and of the patients' needs;
- the availability of a potential recruitment system of patients, eligible for participation in clinical trials.

Depending on the level of granularity adopted in the process of data analysis, on the accuracy, completeness and quality of the data, it could be possible to achieve different goals in a variety of settings, through the use of a unique monitoring system and a unique classification and coding system.

Additional existing parameters should also be addressed, such as the presence of Centers of expertise for rare diseases, having RD coding regulated, the availability of a patient national identifier, etc. The following chapters should help MS implementation bodies to address the RD coding issue as a whole.

2. DESCRIPTION OF GENERAL USE CASES FOR RD CODIFICATION

As presented in the first section, the use of a coding system strongly depends on the use case the data is collected for. In order to meet the needs of all use cases while still generating data in a standardized way, coding settings should follow a minimum level of regulations that allow data comparability. This should enable as well to reduce the time necessary for coding by coding only once and being able to use the data multiple times for the different use cases.

In this chapter, we describe the main objectives in terms of statistical data exploitation that can be set separately or concomitantly to the implementation of such a strategy at MS level.

2.1. Coding for health care planning

National reporting and statistics for rare diseases right now are scarce and not sufficient in many countries. Some do rely on ICD-data; some are based on sampling methods. In order to have full knowledge of the number of patients with rare diseases and of the distribution of specific rare diseases the implementation of a more detailed coding system than ICD is necessary. Using a specific coding system (a new instrument) enables better characterization of data within usual ICD data. This is in line with the recommendation[4] of the Commission Expert Group on Rare Diseases (CEGRD). As stated by CEGRD, Orphacodes are the best candidate for this goal.

Historically, studies for health planning are set using a specific coding of diagnosis and medical acts in order to describe and eventually address health planning problems for specific conditions and populations. Specific surveillance information systems can also be set by authorities in order to react to health problems within populations. These systems are not specific to rare diseases and are generally adapted to the general population. Some rare diseases can be properly traced within those information systems when the disease is recognized as a specific condition for specific reimbursement but most rare diseases cannot be traced.

Of course, these systems suffer from biases which could be numerous especially for weak signals. Many European countries do have a specific system to record morbidities but these are generally linked to a DRG system which can also create biases. Existing system are generally using ICD-10 which is modified for the country specific needs.

[4] http://ec.europa.eu/health//sites/health/files/rare_diseases/docs/recommendation_coding_cegrd_en.pdf

In order to improve the existing morbidity recording system for rare diseases, the integration of a specific extension to ICD can be made. To maximize its usefulness within RD, such an approach should also fulfil with the following:

- The presence of a unique patient identifier at national level or a mechanism to avoid counting twice RD patients as they may be seen in many centers through their lifetime
- The capacity to capture all patient hospital encounters (in-patient and out-patient clinics)
- A clear national instruction (and/or regulation) to code a RD patient diagnosis at least once so it can be followed through its life since proper coding of the RD for each episode of care is unlikely to happen since RD patients have often multiple phenotypic expressions for a rare disease
- Patients should be coded when the rare disease diagnosis is confirmed (or a mechanism to include suspected or undetermined diagnosis assertion should be included, clearly separating situations)

Please note that national reporting and statistics use case should go hand in hand with the international use case but might be more detailed in some specific fields of interest within a country.

As well it is important to have mechanisms in place to give the diagnosed patient a way to transfer the knowledge on the diagnosis along his pathway of care.

Such as the integration of any medical terminology implementation of Orphacodes within a country needs to be part of a comprehensive coding strategy[5].

Sometimes national use cases need to be catered and of course this will be a driver for the way of implementation of coding. For ICD-coding a key driver in recent years was to implement a coding system for reimbursement. This resulted in different approaches and versions used throughout Europe. Considering this status quo setting, implementation of Orphacodes should follow international needs but consider national requirements as well. To ensure comparability of morbidity coding across European countries despite differences between ICD versions and local guidelines for coding, Eurostat provides guidelines to MS to comply with.

A way to achieve this goal is to integrate ICD with Orphacodes by mapping the two coding systems in one file and to make sure that the implementation is not putting extra burden on the coder while producing standardized data. The result is then integrated with the local release of ICD for routine coding.

In the master file issued together with these coding guidelines such an approach is presented: Each rare disease name from the Orphanet nomenclature is presented with one Orphacode and the additional ICD-10 code mapping, according to the rules

[5] http://assess-ct.eu/fileadmin/assess_ct/final_brochure/assessct_final_brochure.pdf

defined by Orphanet (see chapter 1). In order to make the master file usable within a specific country setting it might be necessary to substitute the ICD-10 codes in the European master file with national morbidity codes.

2.2. Coding for RD expert centers

Some countries have created specific health structures within their general healthcare system to ensure equality and high level of expertise for rare disease patients. These healthcare providers may have specific needs in terms of healthcare resources, reporting or research. Besides, new diagnostic techniques (Whole Genome Sequencing, Exome, …), new therapies (drugs, others), the low number of cases sharing the same phenotype or genotype encourage the use of precise large scale identification system of patient diagnostic, phenotype and/or genotype for wider uses than producing health statistics.

Several approaches and tools can be set in order to achieve this; registries can be used as well as systems that can re-use data from electronic health record (the i2b2 example). But to be fully efficient, those systems should share the same data set and vocabularies: the same meaning in order to have comparable (and interoperable) data.

ICD classification is in most cases not suitable for precise identification of cases within the RD field. And since RD-specific health center may also have to produce statistics for health care planning, a link between this data collection can be made towards health care planning data collection.

To achieve this goal, RD-specific healthcare centers should include in their documentation systems at least the Orphacodes from the master file in order to be able to retrieve the data for health care planning needs. Some countries have followed the path of promoting minimum data sets, integrating different coding system alongside Orphacodes. Additional detail can of course be collected and it might be an option to include the whole Orphanet Nomenclature in a documentation system of a health care center. But emphasis should be placed on the possibility to automatically retrieve data according to the master file content and according to the basic coding rules and guidelines given in this document at any time. Tools and methods might vary according to center infrastructure, national regulations for health records or other regional settings as long as they are compliant to follow the basic rules of this document.

2.3. Coding for RD centers for research

In research settings, it might be necessary to have a more detailed coding and to express additional information with the coding. This should be defined according to the relevant setting requirements. Still, the use case of the international aggregation

and exchange needs to be considered and implementation of coding should allow for extraction of the data for the international use case as all other settings should do.

In an ideal situation, data capture of patient diagnostics, phenotype or genotype could be precisely done when captured for research needs (although cohorts may not be exhaustive). From this data, the codes for health statistics could be derived provided individual data can be exported and that a common patient identifier is used at cohort/registry level and at national level. If this is feasible from a regulatory standpoint, the transcoding of the data has to be founded and maintained.

As in the setting for routine patient care, additional detail can be derived by including the whole Orphanet Nomenclature in the data capture system. Even additional coding or ontology systems like Human Phenotype Ontology[6] (HPO), OMIM or other could be added. Still, as this relies heavily on the research intention of the center no international rules or guideline beyond the ones for routine coding can be defined in this document. It is still recommended, that Centers for research discuss and determine additional coding regulations if additional data should be shared internationally.

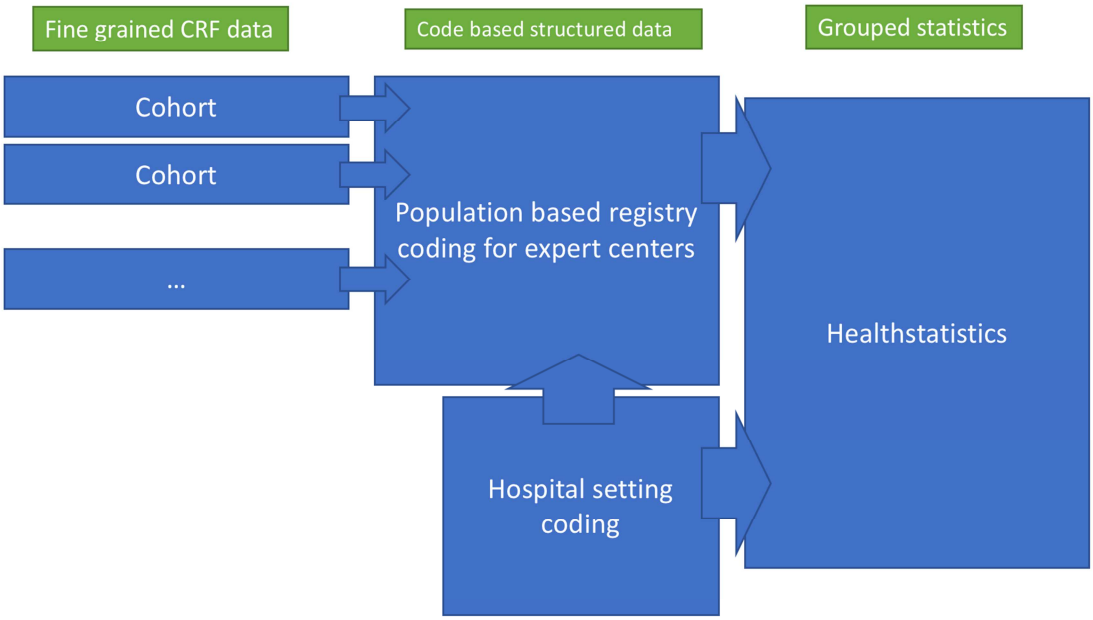

Figure 1 - Possible link between datasets to produce health statistics

The use of a single code system might not always be sufficient. Additional Orphacodes (genetic subtypes, nosological categories) or complementary codes (ICD-10, HPO, Genetic codes) could be required in some settings. It might even be possible that a region or country decides to capture such additional information for statistical data collection.

---

[6] Robinson PN, Mundlos S. The human phenotype ontology. Clin Genet. 2010 Jun;77(6):525–34.

Anyhow, for international comparability the main Orphacode should be identifiable in the data collection at any time.

2.4. Coding for international statistical aggregation of data

The main objective of the WP5 is to promote the implementation of Orphacodes according to the recommendation of the CEGRD mentioned above. On EU-level or beyond data collected in a standardized way can be compared to analyze the number of rare disease patients per country and the distribution of the different diseases. This will enable the EU and its countries to estimate the burden of disease resulting from rare diseases and to plan for better care for this group of patients. For WP5 purpose this is the main objective and the work of WP5 focusses on enabling this use case.

With data available now on rare diseases only few comparisons can be done. Many countries do use ICD-10 but in different ways and according mostly to national regulations. This data is therefore only usable for general planning but not for estimations on rare diseases which in ICD-10 are mostly coded with unspecific categories.

International statistical aggregation is only possible if the coding system used for generating the data is used in the same way in all countries. Using the Orphanet Nomenclature during the coding process allows multiple ways to reach a code and can result in different coding practices if not regulated. This can be avoided by defining a subset of codes from the Orphanet Nomenclature that should be used for coding for the use case of international statistical aggregation and by giving some basic rules on how to code with it.

Of course, a mechanism of collection of data needs to be implemented at the same time that enables aggregation. In order to make sure this is in line with country mechanisms, this use case needs to be aligned with the respective national use case and vice versa.

2.5. Coding for international exchange of data for RD patient care

Patients with rare diseases might require treatment within another country than the one they live in. This can be due to the fact that there is no center for their specific disease close to where they live but as well it might just be due to the increased mobility of people. This can trigger the need for data exchange on a specific person across country and language borders.

A similar use case for coding was addressed in the epSOS-project for emergency care treatment. This project aimed at having a central resource for translation of health data for each patient that receives emergency care in the EU. The data

collected in the home-country would be translatable to the countries language a patient receives the care in and vice versa. With this approach as well agreement had to be taken on which coding system would be used for the relevant data. In case a different coding system was in place in the respective project country a mapping to the chosen coding system needed to be provided. In the epSOS-project for diseases the ICD-10 code was selected. Agreement was to reduce it to a 3-character level for the purpose of the project and to have within one file translations to all languages of the project-partners.

First results in WP5 have shown that such an approach is too burdensome for many countries within the realm of this JA given the detail of the master file. Rather the existing translations from the Orphanet Nomenclature should be used and a mechanism to include them into the master file should be sought. As well the maintenance of many additional national coding systems in a centralized way did not get the JA-partners approval and maintenance of such information on country level was encouraged. Therefore no Basic coding guideline for this use case will be given and in the master file only Orphacodes and the current valid ICD-10 codes will be included.

As the consecutive projects of the epSOS-project have further evolved the idea of a central repository for translation of the patient data that seeks care in another country, the structures have become clearer. Once a central repository is in place and it is agreed to add information on rare disease patient to the patient record for emergency medical care, the master file structure can be adapted to the structure of the other classification and terminology resources and can become part of the central European repository. The structure and content of the master file is designed to be easily transferrable into the other data format.

3. GENERAL CONSIDERATIONS FOR THE PROCESS OF CODING FOR RD

The implementation methods for rare diseases coding should take into account the considerations presented in this section in order to fulfil the coding objectives. Given the very nature of rare diseases, reaching the diagnosis usually requires expert knowledge, techniques and time. Therefore, it can be a very long process, with various level of precision depending on the current knowledge or techniques available to confirm the diagnosis (ex: genetic test, etc.). Additionally, diagnoses might evolve over time, when new disease entities are defined or when new techniques are available. In some cases, tracking the "undiagnosed" patients might be relevant to drive public health policies at a national or international level.

Therefore, several coding implementation strategies can be adopted. For example, using the Orphacodes together with ICD for diagnosed patients could be supplemented with the usage of "other group of disease codes" or equivalent mechanism. Consequently, classifying undiagnosed patients or patients undergoing investigation can receive temporary assigned codes until the diagnosis can be confirmed clinically or genetically. In the specific dataset collection strategy for implementation, a specific diagnosis assertion data item can be captured. For some cases (undiagnosed cases) a set of specific extra terminologies could be implemented in addition to Orphacodes. These extra terminologies are for example the Human Phenotype Ontology for phenotypes and the HUGO HGNC nomenclature for genes. Using these terminologies will ensure a level of interoperability with RD-Connect[7] EU project or Match Maker Exchange[8] IRDIRC supporter project for phenotype/genotype correlation studies.

Guideline 1 - Several tools and strategies could be set at MS level to produce data or statistics for RD, nevertheless each country should set this strategy accordingly to a standard principle of maximizing exhaustiveness as well as possible re-use of existing data collections

[7] http://rd-connect.eu/platform/registries/ontologies/

[8] http://www.matchmakerexchange.org

3.1. Granularity of classification for coding

The Orphanet Nomenclature aims at providing a comprehensive reference of all rare diseases. Rare diseases are numerous and sometimes quite complex with various types of subtypes. Consequently, and from the perspective of routine RD coding, this large resource raises usability issues: it might be difficult for unexperienced coders to identify the appropriate code in a complex classification with so much detail. Further on it is also hard for the user of a given classification to identify the correct code in a short period of time, if a coding system is too complex. This might lead to frustration of the user or to incorrect coding due to the lack of time.

In order to avoid these problems in routine coding, it is recommended to present the classification to the user in a way that assists him in the selection of the right code by reducing the complexity of the coding system he has to choose from. This is true for Orphacodes as well as SNOMED CT implementations for example.

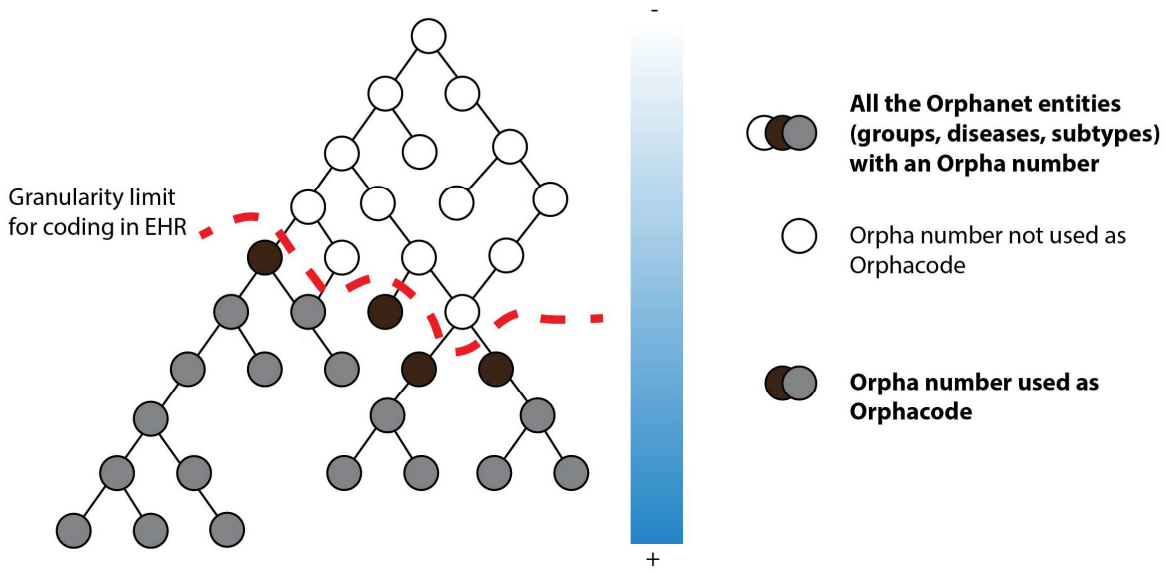

Figure 2 – France choice for the use of Orphanet classifications: separating coding entities for confirmed or suspected diagnosis of categorical entities for organization of diseases following a nosological viewpoint (France uses complementary nomenclatures to Orphanet's)

In the WP5 this will be tested by specifying a reduced list of codes that are necessary for international comparison and should be implemented in coding settings. The list (included in the "master file") can of course be part of a more complex setting but specifies the minimum agreeable level for international reporting and comparison. This way a coder will not be burdened by too much detail if not absolutely necessary.

Still, as the list is an extract of the Orphanet Classification, the same coding result can be compiled by implementing the Orphanet Classification in coding settings with enhanced requirements on complexity and detail like specialized research setting. When generating data sets for international comparability the more complex detail can then be aggregated to the level of detail of the internationally agreed coding list, as illustrated in figure 2.

Guideline 2 - Code the data in a way that the reporting can compile to the granularity of the international recommended list of Orphacodes ("master file"-granularity). If no further national needs for reporting are necessary, use the codes from the "master file" directly.

3.2. Diagnosis assertion option

In order to identify RD patients through their diagnostic journey, the Orphacode can be associated with a marker for the diagnostic certainty a specific qualifier can be given with the code. Most relevant cases are the confirmed diagnosis, the suspected diagnosis and maybe even the excluded diagnosis after thorough checking. For the use case of the international aggregation of data one can then choose if only confirmed cases will be registered or if suspected cases should be integrated into the compilation.

The following table describes the French chosen assertion mechanism to code RD patients using the Orphanet classification and complementary optional nomenclatures or value sets to describe investigations made to confirm the RD diagnosis.

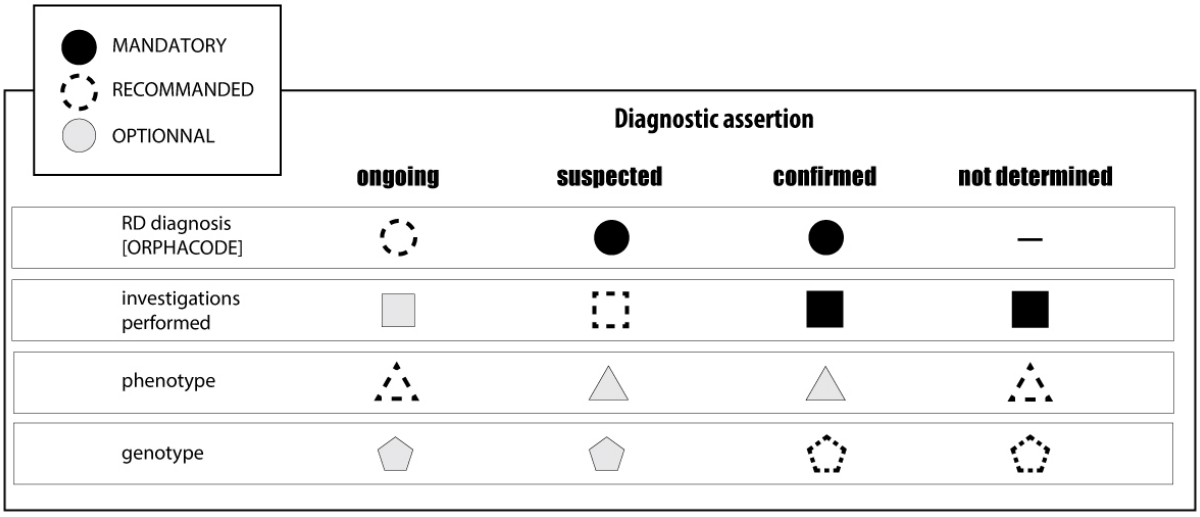

Figure 3 - RD patient description: Diagnosis, Investigations, Phenotype and Genotype

Important: Please note that in France, the data is captured by rare disease expert centers, assuming that the not determined assertion is used on purpose by experts and not applied within the general healthcare system.

A set definition of diagnosis assertion options is proposed. The use of such assertion mechanism should be carefully done. When used, we recommend it is accompanied with clear instructions directed towards rare disease experts.

3.2.1. Confirmed diagnosis

The diagnostic confirmation of cases should be left to physician judgment. However, the techniques used for the confirmation need to be associated with the diagnosis (this information is mandatory in the French system). In an implementation setting where patients are followed through all parts of the diagnosis within an expert center (which would act as a filter for rare conditions), the status of the diagnosis should be implemented alongside the RD code.

Moreover, if the data collection aims at making correlation hypothesis between genotypes and phenotypes, coding genotype information (as well as detailed phenotype information) might be of crucial relevance.

3.2.2. Suspected diagnosis

There are cases where a rare disorder is suspected because of its clinical presentation or the family history. In this situation, the diagnosis is not yet confirmed, but the diagnosis is the best hypothesis from the medical expert point of view.

For example, a neuromuscular disorder might be classified as suspected *Limb-girdle muscular dystrophy* until the gene is finally found years later and be then re-classified as confirmed *Autosomal dominant limb-girdle muscular dystrophy type 1A*.

3.2.3. Undetermined diagnosis: coding an RD disorder when the specific disease is still unknown

Another important information to capture is the patients with suspected rare disease of unknown kind. These patients are numerous[9],[10] and will have to get the most attention in health care settings as it is most important to guide them to the right diagnostic settings in order to provide them with the appropriate care provider. This will not only enhance the quality of care for these patients but as well it will lower the costs for the health systems.

Therefore, a way to code these patients is needed. The definition of the best coding approach is not the key part of WP5 and respective rules should be added once internationally agreed upon. Effectively, this assertion is not clearly defined in the literature. Current usage in France shows that clear inclusion criteria should be set for patient diagnosis to be asserted in such a way. These criteria should be set per RD network accordingly to RD experts.

---

[9] https://undiagnosed.hms.harvard.edu

[10] http://www.rarenewengland.org/Undiagnosed.html

In figure Figure 4, we present a proposal for coding such patients for the RD center specific dataset method. For the first implementation use case (ICD extension) the use of top-codes or other codes could be set.

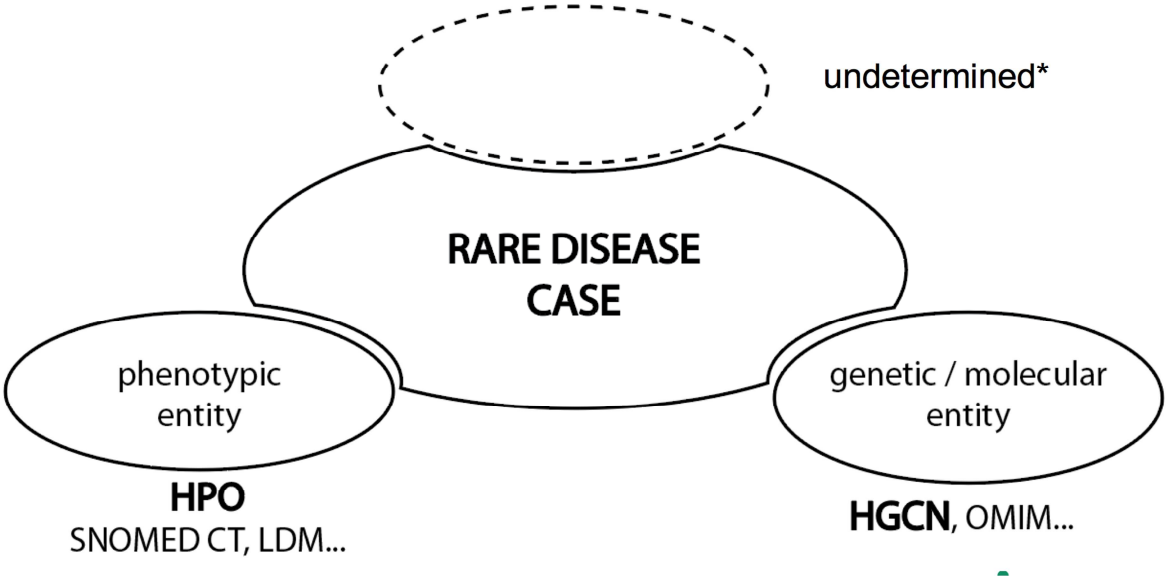

Figure 4 - Undetermined diagnosis. The RD case can be described using phenotype and/or genotype codes. [*when the diagnostic is undetermined, it is recommended to have at least 1 descriptor from other resources such as HPO]

Still, whenever possible the number of patients with undetermined diagnosis should be possible to collect in order to allow for health care planning. Until international agreement on the use of specific codes can be reached, at least a marker for undetermined diagnosis could be used.

3.2.4. Coding the "excluded diagnosis"

In the process of differential diagnosis, some diagnoses are tested and excluded during the investigations. Capturing this information might be helpful but is not mandatory on international level.

Guideline 3 - Whenever possible capture the information of the diagnostic assertion for all RD cases. Use the Options "Suspected rare disease", "Confirmed rare disease" and "Undetermined diagnosis". Additional options might be helpful.

---

4. REFERENCE METHODS FOR IMPLEMENTATION

In order to fulfil the recommendations of the Commission Expert Group on Rare Diseases (CEGRD) it is important to keep in mind the expert groups main objective for coding: "*Having codes for each rare disease would help European and national health authorities to obtain a better knowledge of healthcare path ways and of their impact on specialized health care services as well as on a country's budget planning for health and social service*." This is reflected in the following two use cases.

But one should not forget that national or more specific research requirements could be fulfilled as well. These have been addressed in this document in the section 2.

4.1. 1st level: extension of ICD system

4.1.1. Use of Orphacodes joined with other classifications in routine coding (e.g. for reimbursement systems)

Routine coding is performed in many countries for country-specific and international purposes. One setting of routine coding is of course the coding of mortality data for national and international statistics that has been in place in almost all European countries for a long time. In the last years coding for other purposes has become a more relevant part of national coding: most countries do code some morbidity data for morbidity statistics, but main focus shifts more and more to the reimbursement use case. In the European countries, the approach has developed in different ways from country to country, resulting in different classifications used in different levels of granularity.

While Orphacode implementation is intended to facilitate not only aggregation of data but as well the reuse of data for multiple other purposes, in a routine setting for coding it needs to be aligned to the pre-existing coding systems to achieve best possible outcome of standardized data with the least extra effort for the coders. This can be achieved by having the Orphacodes integrated within the national coding files. MS can choose to add additional detail to the coding (like codes for manifestations of diseases or stages) but the basic coding should use the code combinations from a stem file for all European Countries in all settings where such double coding will be performed. An approach that follows these premises has been implemented in Germany in a pilot project. In order to avoid extra burden on the coder as well as inconsistent code combination, Orphacodes have been added to the ICD-10-GM-Code-files already in use. Once a coder choses the diagnosis text from this file he will get the code from both coding systems and can add them to the patient record. This way the assigned diagnosis always results in the same basic

code combination. This approach goes in concordance with the approach proposed to the Member States in this document. It enables the integration of the Master file (described in section 5) to the national coding files, and harmonizes the international implementation guidelines with the national practices.

Once the data is available in routine settings multiple use cases can be achieved at the same time. Examples can be additional detail for reimbursement settings, medication contraindication checking, treatment advice, guidance to the relevant reference centre for treatment and links to more information on the disease within Orphanet for physicians outside specialist settings. These use cases of course require additional prerequisites and cannot be achieved by just implementing the Orphacodes. Certain additional measures have to be taken, like the implementation of a rare disease code field in the routine data set or another way to link the code to the patient data.

It is important as well to define the way the data will be collected and aggregated in order to allow its flow regardless of the way of implementation of the RD coding. For example, if the data is just collected locally in a specialized center for rare diseases and then reported in an aggregated way, it is not necessary to adapt the national data flow path for routine data. If data is collected for comparison of data between specialized centers for rare diseases or as part of the national reporting, then a respective data flow path has to be established together with the coding implementation to make sure that all collected data can be used respectively.

As national settings for data collection do differ due to different data protection/legal settings and different evolutions of national health systems, an EU-recommendation has to cater for the differences and only should regulate the common denominator. Still, the recommendation is to follow the same path as much as possible throughout Europe to allow maximum comparability of the collected data.

Another consideration is that since RD patients are often affected through their lifetime, a single coding of RD diagnosis could be sufficient to trace its episode of care within its lifetime if the country supports patient record linkage.

4.2. 2nd level: tools for RD centers

4.2.1. A 1st use case: the use of the full Orphanet nomenclature: the experience of the RD Registry of the Veneto Region

The traceability of rare diseases in health information systems has been recognized as an important issue in the Italian RD Plan 2013-2016. In particular, the Plan suggested the use of the Orphacodes in addition to the ICD-Code, in an experimental way, in selected health information systems, particularly at regional level.

The use of Orphacodes has been introduced in the Veneto Region RD Registry since 2006. The Registry is designed as an Information System (IS) to support Centers of expertise and the other actors of the RD care network in their comprehensive

management of rare diseases' patients. According to the national legislation, a patient with a clinical suspicion of a rare disease has to be referred to a specific Centre of expertise, in order to have a complete assessment. Only if performed in an officially labelled RD Center, this assessment is free of charge for the patient. Health data are collected and stored separately from other data in the IS. The possibility of linking the two components is guaranteed by the assignment of an ID, on the basis of a random number algorithm. This represents a non-meaningful identifier. This identifier is used when epidemiologic or clinical research has to be performed starting from registry data. Authorized users have access to personal identifiable data, only when the IS is used for care purposes, i.e. to provide patients with benefits and other services.

The specific identification of RD entities monitored by the RD Registry has been a necessity since the Italian RD list includes single RD entities as well as groups of rare diseases. Some of these groups can include a vast number of different entities, although they do not appear explicitly in the national list. The necessity to adopt a system able to identify each RD entity responds to two necessities in the context of the development of the IS. First, to produce reliable epidemiological data the identification of a RD entity must be as precise as possible. From a health-planning point of view, it is important not only to know the contribution of different nosological groups to the global epidemiological datum referred to rare diseases as a whole, but also to know which is the specific contribution of single rare disease entities. As an example, at regional level specific benefits have been recognized to RD patients, in addition to the so-called essential levels of care established at national level.

In the IS, the process of Orphacodes recording is part of the diagnosis module. A RD diagnosis is recorded in the IS only if performed by clinicians working in RD Centers, officially labelled as "of expertise" for a specific rare disease or for a group of rare diseases. The clinician cannot enter free-text, but uses a thesaurus of diseases names, which appear as a dropdown list. The list is constantly updated according to Orphanet nomenclature and international classifications. International codes are assigned to each disease name. Since 2002, this work is performed and maintained by the medical staff of the registry with the support, when needed, of expert clinicians working in the Centers of expertise. When using the information system to register a new patient, clinicians need only to select the name of the disease, without any code, i.e. "Alport syndrome". The system automatically shows the possible general branch in which Alport syndrome is represented, according to the Orphanet classification. The clinician then selects the branch that best represents the clinical representation of the patient. Automatically, the corresponding ICD code and Orphacode for Alport syndrome appear in the patient form that the clinician can visualize and use to fill hospital records. All the information collected and coded is registered in the system to generate a "coding pathway". This method allows capturing a single disease or disease subgroup, or a group of diseases, depending on the granularity of the Orphacode and informs on the prevalent health needs of the patient depending on the branch selected. The recording of the codes' pathway allows the multi-

dimensional aspect of this classification to be preserved and fully exploited. This approach is useful to describe the phenotypic variability of presentation of RD patients' in a real-world clinical setting. It can be used in research activities, in the clinical practice and for public health purposes. This automatic tool, embedded in an informative system for mandatory RD patients' registration, supports the diagnosis activities performed by all Centers of expertise in 8 Italian Regions with a covered population of nearly 25 million inhabitants.

Advantages of using this approach are:

- The broad exploitation of the Orphanet classification as it has more hierarchical levels compared to the ICD. This allows not losing interesting intermediate terms, potentially describing a specific patient. This can be used also to describe patients without a diagnosis, who can only be defined using a broader level of granularity at a certain point of the diagnostic process.
- A broader exploitation of the Orphanet classification derives also from the fact that its organization is more flexible, as disease entities can have a multiple parentage. This turns out to be useful when we want to reuse data in order to capture patients that fall into categories of interest i.e. "genetic glomerular disease" rather than focusing on a specific disease i.e. "Alport syndrome". Thus, the Orphanet classification can support secondary RD patient data analyses, even if the whole classification itself is not embedded in the coding system.

In the system, clinicians only need to enter the name of the disease, selecting the appropriate one from a list without having to enter a code or to select the corresponding ICD and Orphacode. Thanks to this procedure, it is easy to select all the cases with a given disease.

4.2.2. Extended description of RD patients

In Italy since 2002 and in France since 2007, RD expert centers were officially identified through a selection process. In France, within the second plan for rare diseases (2011-2016), a national database for rare diseases project was set. It upgraded and complemented the previous project, recording at minimum Orphacodes together with a minimum data set. Whilst the previous Orphacode integration, which is comparable to the Italian experience, has enabled the identification of approx. 370,000 RD patients (31st Dec 2016), a need to add additional descriptors for extended use cases has risen. From that perspective, an expert review of the Orphacodes as well as the Orphanet classifications was started together with RD health networks. As the result of the choice to prepare France to personalized medicine (plan *France Medecine Genomic* 2025) and to enforce interoperability between hospital electronic health care records and the national project, propositions were made to extend the Orphacodes with complementary resources for phenotype coding as well as genes that could be used alongside the

Orphacode in some situations. Extension is not mandatory for all centers but it is recommended in some situations.

RD case description (that could fulfil the description of the full patient phenotype and/or genotype) cannot be set using a single code system. Therefore, the precision and the completeness of the patient state could be captured using complementary resources as stated within Figure 5. *Please note that these complementary coding resources can also cover the same concepts in some cases.*

In this setting, which extends the actual Italian setting, healthcare experts are invited to code patients using extra descriptors when:

- a rare disease cannot be yet confirmed, professionals can use Orpha categories and ICD codes or HPO phenotypes to code suspected cases
- the diagnosis is confirmed, the use of the appropriate Orphacode is mandatory, additional descriptions are supported to describe unusual signs for example
- the diagnosis is Undetermined, physicians are encouraged, to be compatible with RD-CONNECT or other local studies to coordinate the use of some specific phenotypic descriptions using HPO and/or genetic descriptions (if available) for further patient grouping or correlation population based studies

In this extended setting, the Orphacodes are mandatory if the diagnosis assertion is suspected or confirmed.

This new mechanism was incorporated within the national e-health data exchange framework regulated by the French agency for e-health (ASIP Santé). The electronic formats are shared with health application vendors to be integrated at hospital levels within EHRs or specific applications. Technical formats do use international standards for health data standardization (HL7, CDA, SNOMED, ORPHACODES, HPO, etc.). They are freely available[11].

[11] http://www.bndmr.fr/un-cadre-dinteroperabilite-pour-les-maladies-rares/

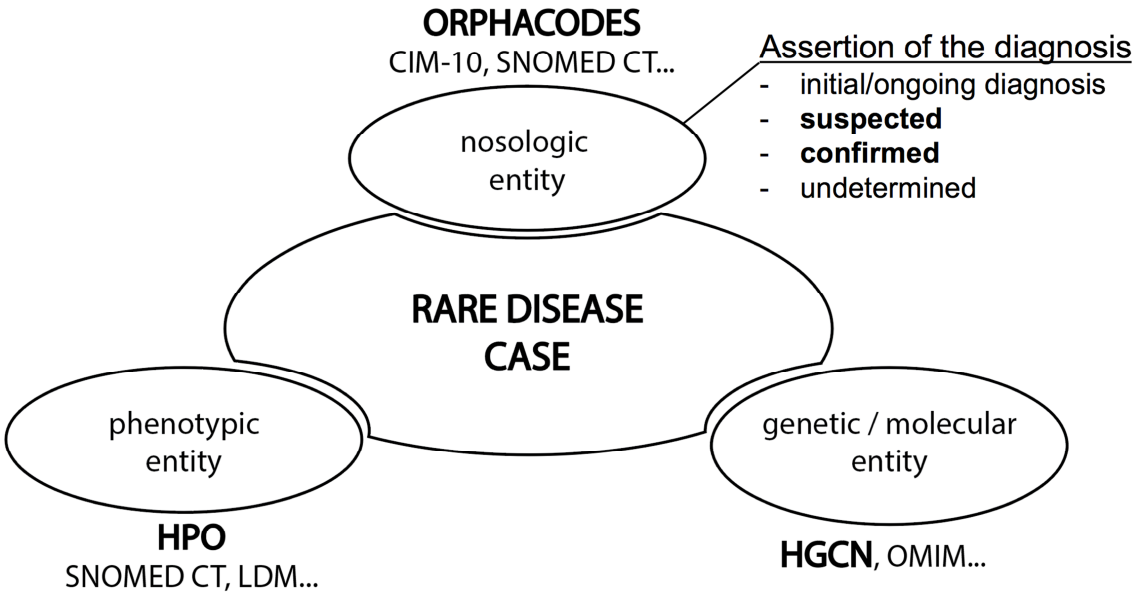

Figure 5 - The 3 dimensions of RD coding: nosology, phenotype, genotype. The level of certainty (assertion of the diagnosis) is a modifier of the nosological entity used for coding.

In France, the contribution to the national database project (BNDMR) to data collection is mandatory (regulation) and linked with a specific financing mechanism.

4.3. 3rd level: tools for registries/cohorts/etc.

Registries are key tools to generate knowledge about rare disease patients. Given the low prevalence of most rare diseases or conditions, it is much likely that data from a large number of hospitals will be required to generate sound knowledge depending on the purpose of the registry.

As stated in the EUCERD recommendations for rare disease registries and data collections[12], a *patient registry 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 serves one or more predetermined scientific, clinical, or policy purposes.*

It is usual to distinguish between population-based registries, which refer to a geographically defined population and aim to register all cases in that population, and non-population-based registries based on clinical centers or other criteria - such as a disease condition, *members of a patient organization, participants registered via an ERN or other disease-specific registry*, *etc. - where the population coverage may not be comprehensive. These types of registry have different* purposes *but both are useful provided they serve identified target aims. Both types of registry are the*

[12] http://www.eucerd.eu/wp-content/uploads/2013/06/EUCERD_Recommendations_RDRegistryDataCollection_adopted.pdf

targets of these EUCERD *recommendations. Multiple RD registries (>600) already exist in Europe*[13]*. The key principles proposed apply also to these existing datasets as they adapt to the changing environment for registries in a European and international context. The current recommendations for the basic principles underlying RD registration should take as a starting point generally accepted guidelines for registry development* which will be further developed by the EU JRC platform for rare disease registries.

A classification of registry use cases was proposed[14]:

- **Knowledge dissemination:** distribution of information to patients and their clinicians on new therapies, best practices, and safety issues
- **Patients' recruitment:** providing patient-population information for designing trial protocols that optimize size and length of trials
- **Clinical epidemiology:** population descriptive statistics, natural history of disorders, medical practice variation
- **Clinical effectiveness:** evaluation of the effects of preventive, diagnostic, and curative interventions delivered in real-world settings
- **Safety monitoring:** orphan drugs are generally not tested in large phase 3 studies, which makes the need for post marketing safety surveillance via registries even more important than with conventional drugs
- **Quality and outcomes improvement:** enhancing patients' outcomes by standardizing practice and reducing practice variation
- **Genotype/phenotype association studies:** the registry provides phenotypic data which can be linked to genetic and other exposure data
- **Linkage to bio-specimens and bio-repositories:** to detect phenotypic correlates of cell and tissue biology

Unfortunately, given the low structuration of EHR data and the heterogeneity of ICT solutions, data re-entry is the most common way of implementing such registries at large scale. The high number of rare diseases and the heterogeneity of knowledge that might be available make it difficult to build a generic approach or tool at EU level. Besides, given the relatively low therapies (care) available for those diseases, healthcare professionals have historically build local, national or European registries that might integrate generic care management and research functionalities together with other requirements as cited above.

---

[13] Disease Registries in Europe, Orphanet Report Series, Rare Diseases Collection, January 2013

[14] The case for a global rare-diseases registry, Forrest, Christopher B et al. The Lancet , Volume 377 , Issue 9771 , 1057 - 1059

Electronic health records systems can offer rich and sound longitudinal view upon patient disorder especially for long term conditions such as rare diseases. The need to enable collection of data for RD patients at national or European level is greatly growing. Whether this data collection is set within a population based registry or a cohort instrument, interoperability between health information systems and these tools should be encouraged. Rare disease patients have different follow up requirements within care. While many still don't have a genetic diagnosis, most have lifetime conditions that require specific expertise and follow up. Given the small number of patient per disease, building a coherent catalog of rare disease patients for research screening takes time as some rare disease patients may be seen only once by the RD expert. The data pooling may then be stable overtime and adapt itself to many local IT servicing situations. This catalog, such as the French or the Italian projects is although not enough for most scientific studies. Disease registries are often built and necessary. A further option is the creation of disease specific registries embedded into more general ones, serving different purposes, while sharing the same infrastructure.

4.3.1. Example: The European Cystinosis registry

A European registry for Cystinosis patients was created in 2011 in France in order to investigate the natural history of the disease. The registry was also set to better understand the treatment impacts on patient's quality of life, to evaluate the treatment effect on the diseases as well as its observance.

Renal transplantation and the availability of cystine-depleting medical therapy, cysteamine (EU/1/97/039/001, EU/1/97/039/003), have radically altered the natural history of cystinosis. Cystinosis is a good example of a "paediatric" disease where patients now survive into adolescence and adulthood. These individuals have complex, multisystem problems that require lifetime care and eventual data capture.

The registry was set to study the clinical outcomes of a cohort of French adults with nephropathic cystinosis and the impact of long-term oral cysteamine administration. The analysis of this cohort (86 adults) showed that early and prolonged treatment with cysteamine has a positive effect on life expectancy and on the onset of end stage renal disease and extra-renal complications.

Thirteen centers in France participated as well as three European centers from Italy, Belgium and the Netherlands. The registry data collected were:

- Demographics
- Kidney survival
- Effect of cysteamine treatment
- Growth together with the use of recombinant growth hormone
- Validation of techniques for cystine measurement
- Genotype/phenotype links

In the context of this registry, many clinical data are collected: patient status, specific symptoms recording, biological measures before and after renal transplant, treatment status, weight and height, recording of specific clinical signs at the gastro-intestinal, neurological and muscle levels, etc... Data is recorded at patient visit (or later) by the physician.

Unfortunately, none of this data is built up on existing data standards. A recent French national survey showed[15] that most disease based data collections do not follow any data standards. Many different reasons could be invoked: The limit of the funding, the lack of specialist to build such tools, the a-priori absence of need to share data with others, or to combine data at international level, etc. The result is that registries are poorly interoperable at any level. This situation is seen as a key priority by the European Commission hence a team of specialist was set at the EU Joint Research Centre in Ispra (Italy). Main activities of this team for the years to come are to promote guidelines and tools to enrich interoperability of rare disease registries across Europe. ERNs will surely also promote European disease or population registries.

We believe in the perspective that registries are formidable tools for epidemiology although their running costs are rather high to produce sound data over years and space. Recommendations from EUCERD are important for newly built registries. We also believe registries should implement means so their data can be compatible with the guidelines carried by this document. We don't think though these types of registries alone can answer to RD population based analysis, they should be though linked to national level datasets for health statistics.

Whenever possible, existing standards should be used to help registry builders. Registries should be build implementing standardized data interfaces and the required security from design. The FAIR Data Initiative which was launched in January 2014 is quite interesting with regards to data sharing for registries.

Promoting interoperability from EHR to RD registries on the one hand, and interoperability between RD registries on the other hand, is a really complex task. It involves promoting the use of standards for the modelling of RD registries as well as ensuring the possible re-use of care generated data from EHRs within a registry context that is supposed to generate new knowledge and/or insights on the rare disease natural history or treatment course evaluation and therefore implies strong data quality procedures.

Through the future ERNs, recommendation of healthcare data modelling and terminology standards for ERN level registries should be a requirement in order to lift registry standardization and mechanically promote interoperability as a state of mind. Care generated data, following all works underway at national and EU level to enable

[15] http://www.bndmr.fr/wp-content/uploads/2016/06/Poster_enqueteBDD_ECRD2016.pdf

interoperability between HCPs, will become more and more standardized and therefore, interoperable although much care generated data is still plain text or semi-structured data or captured into another context for another objective (care vs research).

For the building and promotion of registries, it is also key to help registry makers to be aware of clinical data standards (openEHR, HL7, LOINC, etc.) as well as research standards (CDISC, OMIM, HGVS[16]) and that they are better known, promoted and used. Promotion of open standards could also be a key point here given the limitation of resources for research on low number of patients.

Guideline 4 – Although rare disease registries (disease, population or patient based) should promote the use of data standards to increase interoperability of their data, they should not be the only instruments upon which the EU strategy to produce health statistics for RD at population level relies.

---

[16] The Human Genome Variation Society published a Sequence Variation nomenclature

5. TECHNICAL REQUIREMENTS

Certain technical requirements arise from the intention to use Orphacodes as international standard for routine coding of rare diseases. This includes a regular (annual) provision of a standardized file (EU master file) on the Orphanet platform together with a history of changes to the file and the provision of all previous files in an archive. A cycle for posting these files has to be agreed upon.

WP 5 will give a recommendation at the end of the Joint Action in Deliverable 5.5. Still, in order to make best use of such a regular cycle some coding Guidelines are necessary to be set in this document already and are outlined in the following sections.

5.1. Updating of the coding file

As mentioned above, the WP5 started to explore the use of possible tools supporting the routine encoding of RD across MS. One of these is the “master file”, through which the routine coding process should be simplified and structured. The master file should facilitate the use of the Orphanet Nomenclature by providing only the data that is relevant for routine coding. Offering an alignment of diagnostic terms with the Orphacode and a terminal ICD-10 code in a dataset, could minimize the bureaucratic burden of using different classifications and support standardization. Furthermore, giving the MS the possibility to align such datasets to other classifications used locally for morbidity coding could give a higher added value to the file. Nevertheless, the master file should allow an implementation at different levels so that all MS could use it independent of the available resources.

Integrating the ICD classification and the Orphanet nomenclature into one file as described above, should increase the standardization of rare diseases coding and consequently increase interoperability on data sharing. The interoperability of data is a key aspect in processes where data sharing plays a role. In the rare diseases field data sharing becomes much more relevant than in common diseases. On one hand, there is a higher need of networking between health care providers given the reduced number of experts in each disease. And on the other hand, there is the need to aggregate data for epidemiological purposes and research given the reduced number of patients.

Besides the already mentioned aspects, the master file should also provide a frame for the routine coding process in order to make it practicable and stable. This could be achieved through versioning of the file and regulation of the update cycle. Regular updates of the coding files are necessary as the field of rare diseases is evolving fast. Still, an agreed update cycle that is followed by all countries is necessary to ensure, that all countries use the same version. From experiences with ICD it is recommended to have an annual cycle of updates. The WP5 team agrees that the master file should be published (updated) once a year. The publication of a file that should be used for routine coding in a more frequently rhythm could result in a

bureaucratic burden for the implementation and lead to inconsistences in the captured data.

The versioning information should be captured together with the data collected. Every year the list of codes (“master file”) should be provided for implementation in a standardized way. Ideally this would be aligned to ICD updates (File to be used from January first onwards for one year) as the two coding systems will be used together in the same settings. This presented file has to be implemented by the countries as well in order to cater to the frequent updates in the Orphanet Nomenclature.

Guideline 5 - Update your coding resource according to the internationally agreed cycle in order to have the most recent coding file and to ensure comparability.

5.2. Coding for international and national purposes at the same time

A preliminary requirement for any integration process is the analysis of the main characteristics of the two resources, ICD and Orphanet, focusing on the specific issue of RD representation and thus coding.

The WHO International Classification of Diseases, 9th edition, Clinical Modification (ICD-9-CM) and ICD-10 are widely used for reimbursement purposes and public health reporting (WHO recommends using ICD-10 in its most recent version including all updates). These classifications are at the basis of the production of morbidity and mortality statistics worldwide. Given the purposes of these classifications, they present some limitations when applied for other specific use cases.

The ICD limitations become evident when considering rare diseases. These limitations should be taken into account when we want to proceed thorough a mapping exercise between the ICD classification and the Orphanet nomenclature.

Rare diseases are a heterogeneous group encompassing many clinical entities and present some peculiarities that affect their representation in ICD.

Given these limitations it might still be necessary to use two coding systems at the same time. In most settings, this will be ICD-10 joined with Orphacodes.

In order to achieve standardized data and allowing the coder to code two coding systems at the same time, both systems should be linked by a standardized mapping according to the disease name. In a file, which holds all necessary diseases relevant for the use case but a minimum of the diseases specified in Coding rule 1, both system codes should be given with the disease name. This can be achieved through a code linkage in the master file or through some other local system.

Guideline 6 - If Orphacodes are used together with another national coding system for morbidity coding, the two systems should be linked in a standardized way to ensure that code combinations are standardized and the coding effort for the user is minimized.

6. International rules and guidelines for coding rare diseases

In order to achieve international comparable data collection, it is important to use a classification system in a unified way. Implementation of ICD-10 worldwide has shown that guidelines for coding provided together with the classification can enhance the comparability of the collected data. With mortality statistics that follow clear guidelines for coding and are synchronized to an annual update cycle of the classification, international comparability of data is high. For morbidity coding with ICD-10 many countries have implemented individual coding guidelines; sometimes even multiple different coding guidelines for the different settings within one country. Resulting morbidity data is difficult to compare and can be used for international statistics in a very limited way.

Still, the necessity to adapt the coding to national needs has triggered the different coding guidelines for ICD-10 in morbidity and show a need for national definition of coding with a classification system.

Learning from the ICD-10 experience, coding with Orphacodes should follow guidelines that are agreeable on an international level and do leave room for national adaptation without risking international incompatibility. Therefore, the following rules and guidelines should be tested and implemented together with the implementation of coding with Orphacodes to achieve the European goal of comparable data and – with that – of a better picture of rare disease patient numbers and distribution.

The Coding guidelines consider the use cases described above but do focus on the use cases for international aggregation and exchange of rare disease data.

In this section, the coding guidelines given in the document above are summarized for easy reference. For further explanations see the sections 1-5.