



Work Package 5

Milestone 24

Specifications for an integrated coding application with Orphacodes

Issued on the 16.06.2016

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



Co-funded by
the Health Programme
of the European Union

This document is part of the project / joint action '677024 / RD-ACTION' which has received funding from the European Union's Health Programme (2014-2020).

It has been produced by the RD Coordinating Centre -Veneto Region, Italy and is part of the Task 5.3: Promoting the Orphacodes across MS by sharing coding tools and testing the master resource (Task Leader: Paola Facchin- Veneto Region [VR-IIBRD, Italy). It constitutes the Milestone24: Specifications for an integrated coding application with Orphacodes.

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:

The content of this document represents the views of the author only and is his/her sole responsibility; it cannot be considered to reflect the views of the European Commission and/or the Consumers, Health, Agriculture and Food Executive Agency or any other body of the European Union. The European Commission and the Agency do not accept any responsibility for use that may be made of the information it contains.

I. Glossary

A definition for some terms used in the document is reported in this section.

Disease: Deterioration or break of the physiological or morphological state initially considered as normal, balanced or harmonious.

Clinical terminology¹: a terminology required directly or indirectly to describe health conditions and healthcare activities (ISO 17115)

Reference terminology¹: a terminology containing only concept names as determined by an authorized organization. Reference terminologies must allow unambiguous communication of meaning across settings and among professionals and consumers.

Classification¹: an exhaustive set of mutually exclusive categories to aggregate data at a pre-prescribed level of specialization for a specific purpose.

Coding scheme¹: collection of rules that maps the elements in one set, the "coded set" onto the elements in a second set "the code set"

Coding system¹: a combination of a set of concepts [coded concepts], a set of code values, and at least one coding scheme mapping code values to coded concepts. Coded concepts are typically represented by terms, but can have other representation. Code values are typically numeric or alphanumeric.

Mapping¹: assigning an element in one set to an element in another set through semantic correspondence

Semantic correpondence¹: measure of similarity between two concepts

Orphacode: the unique and permanent identifier called Orpha number assigned by Orphanet to a given entity (a disease, a group of related diseases, subtypes of disease). When it identifies a disease entity it can be used to record it within an information system. The code can be assigned to a patient and used within an information system contributing to define the group of patients who are affected by the same disease.

¹ ISO Health informatics- Vocabulary for terminological systems- [ISO 17115-2007]

Introduction

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 ones. In order to tackle this issue, since 2007 Orphanet, besides collecting and making accessible available information on rare diseases, started to classify them, adopting a poly-hierarchy approach. In this effort, still ongoing, each clinical entity present in the Orphanet nomenclature is being assigned a number always preceded by "ORPHA" [1]. The Orphanet inventory of rare diseases is continuously up-dated as new RD clinical entities are discovered. The inventory relies on the European definition of rarity, (i.e. a prevalence lower than 1 in 2,000 in the general population) and on the definition of clinical entity. This work has been at 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) [2].

In many countries, ICD in its versions (10th-9th and 9-Clinical Modification (CM)) is used to record patient information in health information systems. As specific codes for many RD are lacking in ICD-10, and even more when considering ICD-9 and ICD-9-CM, morbidity and mortality statistics referred to single or to related groups of RD are not easily available. At present, most of the information on RD derives from specific data collection (i.e. disease registries), whose number is constantly increasing. A snapshot of which are the terminologies and classification systems used to identify RD in these registries is not available yet. A pitfall of this approach is a general data fragmentation that hampers the availability of data referred to RD considered as a whole and, furthermore, it represents one major problem of contents interoperability.

Information on RD is evolving rapidly as genomics has revolutionized not only the study of rare diseases, but also the way in which data are collected. Therefore, whilst a lot of information related to RD is available in gene, protein and human mutations variations databases, RD are still under-represented as distinct recognizable clinical entities with specific codes in the international classification of diseases. The narrowing of this gap is of the utmost importance as it can enable comparative and integrative analyses between different already available data sources, that otherwise will continue to stand alone.

There is an increasing need for updated and harmonized health care classifications and coding systems to allow comparisons of data coming from different sources. Data integration becomes particularly relevant when the same group of diseases is commonly described, classified and coded according to different available tools and/or different settings. An example are psychiatric diseases, usually classified and coded according to ICD as well as to Diagnostic and Statistical Manual of Mental Disorders (DSM), in their different versions. The same consideration can be applied to the classification of, as an example, injuries and tumors.

The scenario is complex as terminologies and classifications have been developed independently and because, even when considering the same classification, i.e. ICD, adaptations have been produced and are currently in use. The integration of the ICD and the Orphanet classification is driven by the need of increasing granularity of clinical content in order to make RD more visible in health information systems. The development of a resource integrating the two classification systems, one developed for all diseases, and one specifically dedicated to a group of RD, although numerous, requires a preliminary analysis of the main characteristics of the two systems.

All Member States use the ICD classification for the recording of mortality data and most of them of morbidity data [3] [4]. As already mentioned, whilst ICD is a classification in which both rare and non-rare diseases are represented, Orphanet classification has been specifically developed starting from the Orphanet nomenclature to allow the classification of RD, as those defined by the European prevalence cut-off.

The Orphanet effort to classify RD has been deeply described [1]. A number, always preceded by the "ORPHA" in capital letters, identifies each rare disease entity. To each RD entity can correspond a fully specified name, one or more synonyms, one or more eponyms or a combination of those. These descriptors, derived from international published literature can be modified over time, but the "ORPHA" plus the number assigned to each entity remains stable over time. The Orphanet classification adopts a clinical approach. It has a great potential, as each entity is included in a complex relational database (DB), so that per each entity a great amount of further information can be retrieved (i.e. genes involved, clinical manifestations and their age of onset, etc.).

The ICD as well adopts a clinical approach. The classification is based on a predefined number of granularity levels. A disease name can appear in the classification as a main term, an inclusion term or an index term. In the ICD each disease must be assigned a single code, whether the disease is specifically mentioned in the classification or not. Whilst several distinct diseases can share the same ICD code, in Orphanet classification each disease entity is assigned with a specific orpha-code.

Interoperability issues

Electronic health record systems able to 'interoperate' represent the basis for data sharing and exchange of information, a vital need to improve the quality and safety of patients' care and boost research. Interoperability of systems collecting data on RD assumes an added value, as in this domain information is particularly scattered and usually collected for different purposes, thus using different tools. The European Interoperability Framework classifies interoperability according to four levels: legal, organisational, technical and semantic.

The following aspects, among others, should be clarified when analysing information systems candidate to be interoperable: which is the statistical unit object of the data collection; which are the sources of the collected data; how data are recorded.

In fact, data can exist as free-text or as already structured data. The process of data structuration is achieved thanks to the use of terminologies, coding and classification systems.

Thus, any attempt towards integration has to tackle not only the issue of technical interoperability, but, above all, the issue of "semantic interoperability". The European Commission Recommendation on cross-border interoperability of electronic health record systems defines '*Semantic interoperability*' as the process that means ensuring that the precise meaning of exchanged information is understandable by any other system or application not initially developed for this purpose [5].

Furthermore, semantic interoperability deals with at least four sub-issues [6]:

1. temporal
2. meaning
3. granularity
4. structure

For all these issues a specific relation to the RD domain is provided.

1. The temporal issue

It is particularly relevant when dealing with rare diseases terminologies/classifications. It encompasses two aspects. First, the meaning of data can change over time, due to changes in clinical definitions, a common situation when dealing with uncommon diseases, for which progress in knowledge can introduce new diagnostic criteria and new classifications entering in use. Second, it deals with terminological changes due to the emerging of new concepts i.e. names of diseases shift from a general descriptive name to a more informative name referred to the etiological mechanism, previously unknown.

2. The meaning issue

For many RD a set of different terms (describing the same RD entity) are in use. Their frequency of use is dependent on the context, whether geographical or related to the care setting (e.g. centers of expertise vs primary care setting for example). Second, the use of a term can vary over time, with some becoming obsolete.

3. The granularity issue

Different systems can be designed in order to capture different levels of granularity. These levels of granularity can be predefined or not. In the RD field this issue is also related to the possibility of dividing the same RD entity into subtypes, the so-called “salami-slicing” effect. In some cases these subtypes can be defined as clinical entities, whilst in others they can be considered as artificial sub-sets of the same “parent rare disease”. The required granularity of the coding can vary according to the scope of the data collection (i.e. public health purposes vs research purposes) and the characteristics of the user.

4. The structure issue deals with how the information is categorized.

Classification implies the categorization of relevant concepts for the purposes of systematic recording or analysis. The categorization is based on one or more logical rules [7]. In a classification a certain number of concepts are grouped within a category. How these concepts are grouped depends on the purpose of the classification and its use. In particular, the process of diseases classification starts from the definition of patterns of features and establishing those that allow one disease to be distinguished from another. Classifying diseases is not a static process as it is driven by evolving diagnostic tools, better etiological understanding, and furthermore implies a scientific community consensus.

Another issue is that the categorization process may vary according to the scope of the classification. If the classification is used for statistical purposes, there could be a major interest to group uncommon concepts into broader categories. If the purpose is reimbursement the concepts can be the same, what changes is the logic according to which categories are built (i.e. case-mix systems).

When reference terminologies are developed for specific groups of diseases or interventions, the concepts are usually fine-grained. It is the case of terminologies dealing with rare concepts. ICD is a classification dealing with both common and uncommon diseases, thus, represented categories, into which concepts are grouped, are not per se fine-grained. Mapping from a fine-grained terminology (evolved into a complex poly-hierarchical classification) to a less fine-grained classification represents a challenging process.

Integrating the two classification resources

A preliminary requirement for any integration process is the analysis of the main characteristics of the two resources, ICD and Orphanet one, 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. 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.

Some of them regard the terminological domain and can be summarized as following:

- the use of terms that can become obsolete as new terminologies enter in use
- the use of broad categories, per definition not intended to identify a precise list of entities (i.e. “not elsewhere classified”/”other”)
- difficulty to serve a multi-hierarchy approach

The ICD-10 compared to ICD9-CM has tackled mainly the obsolescence issue, and in a very limited way the other cited aspects. These ICD limitations become even more evident when considering rare diseases. These limitations should be taken into account when we want to proceed through a mapping exercise between the ICD classification and the Orphanet one.

RD present some peculiarities that affect their representation in ICD, considering the above-mentioned limitations.

First, rare diseases are a heterogeneous group encompassing many clinical entities. The belonging of a disease to this group is defined by two concepts: the rarity one (that can vary across countries) and the “clinical entity” one.

For the purpose of this document, we will refer to the EU definition of rarity and to the definition of clinical entity adopted by Orphanet for setting up the list of RD present in the Orphanet DB.

The great numerosity of this group of diseases and the fact that they can involve different body systems and functions, often simultaneously, explains how challenging it is to set up and maintain a terminology of these diseases based on the published literature on one hand and, on the other hand, is one of the reason for their scarce explicit representation in the ICD classifications.

Regarding the content, an issue is represented by how RD are named in the two systems. Many RD are not present with their names in ICD neither as main terms, nor inclusion terms or index terms. Orphanet has instead a wide term coverage, including, but not limited to, the management of synonyms and eponyms, a relevant issue when dealing with uncommon conditions.

Another issue is concept permanence. This issue deals with the fact that once a concept is created, namely a disease entity, it remains stable.

RD names can evolve and also different names can be used more often in some countries and/or settings at a given time. The preferred name may change, but the entity exists as such and different names are conducive to the entity because it has a unique identifier, stable over time. Disease concepts may evolve as well, as knowledge progresses and new classifications enter in use. This is particularly the case of RD, as the use of new diagnostic techniques is changing our understanding of many underlying disease mechanisms. So, the issue of the characteristic of the disease entity identifier becomes very relevant.

In the ICD the identifier is the hierarchical code that can be associated to a main term, an inclusion or a index term. Many RD are lacking this direct explicit relationship as coders can attribute different possible codes to the same RD, even when some general rules are established [8]. Another issue is related to the permanence of the entity in the classification. It remains in Orphanet (where the names of a disease can change, but each entity has a stable unique identifier). It can change in different versions of the ICD. In this sense, each ICD version can be considered as a “stand alone” version, compared to the others.

Whilst in the ICD, the code assigned to a RD conveys a hierarchical information, in Orphanet the Orphacode does not convey this information. The Orphanet classification allows a N number of concepts at different levels of the classification, creating a tree architecture.

Some codes correspond to RD entities, whilst others, at upper level of the classification, represent branches that can be a group of related diseases with very different levels of granularities. The upper levels described by an Orphacode are wide groups of RD such as “rare endocrine diseases” [ORPHA97978].

The Orphanet classifications is not based on a predefined number of granularity levels. A clinical RD entity can be divided into a number of subtypes, that can be both genetic, or etiologic ones. At the upper level of the classification, a RD entity can have more than one parent. This situation has been defined as “multiple parentage”. These hierarchical structure is based on a logical design according to which more general terms are parents to more specific children. A rare disease entity can be part of more than one hierarchy, and even “parents” can do.

The Orphanet structure of the classification responds to the need for poly-hierarchy. According to this design, the same disease entity can appear in multiple places in the classification and in the hierarchy.

Of note, poly-hierarchy is very important when considering the different purposes of a RD classification and the different possible users.

Different possible situations can arise in the effort of integrating the Orphanet classification within the ICD. The starting point is how diseases are represented in the two systems, but the process

includes also the analysis of the concept associated to a certain term. A preliminary issue to achieve an integration is to establish which versions of the two classifications are the objects of this integration. Two aspects have to be taken into account:

- the availability of a consolidated version of the classification
- the availability of a definite temporal version of the classification

We consider the lack of precise and agreed information regarding which is the temporal version of the Orphanet classification as a major issue that has to be tackled before going through the mapping exercise. It represents an essential prerequisite.

In this document, some examples of the problems of integration will be provided considering the Orphanet version of the classification (accessed May, 2016) and the ICD-10 (9) and their possible solutions.

1. There is a direct correspondence (one-to-one relationship) between a main or an inclusion term or an index term associated to an ICD code and a disease entity name present in the Orphanet terminology described by an Orphacode. In this case, the ICD code corresponds to a unique Orphacode. Both identify the same rare disease entity. The match is exact. The match can also be partially exact, but no doubt exists that the disease entity is the same in the two systems. In the integration process there will be a direct trans-codification from one ICD code to one Orphacode.

Example: Behçet disease

“Behçet disease” in the ICD-10 corresponds to the code: “M35.2”

“Behçet disease” in the Orphanet classification corresponds to the Orphacode: “ORPHA117”

2. There is a direct correspondence (one-to-many relationship) between an ICD main term and a name present in the Orphanet terminology describing not a disease entity but a group of rare diseases (a branch of the Orphanet classification). In this case, the ICD code describes a set of diseases (all rare). The match can be exact or partially exact. In this case, it has to be established if the group of diseases is coincident.

Example:

“Disorders of urea cycle metabolism” in the ICD-10 corresponds to the code: “E72.2”

“Disorder of urea cycle metabolism and ammonia detoxification” in the Orphanet classification corresponds to the Orphacode: “ORPHA79167”

In the ICD-10 inclusion terms of “Disorders of urea cycle metabolism” are the following:

- Argininaemia
- Argininosuccinic aciduria
- Citrullinaemia

-Hyperammonaemia

In the Orphanet classification more diseases entities than the ones explicitly reported in the ICD as inclusion terms are children of the parent “Disorders of urea cycle metabolism and ammonia detoxification”. The ones reported in bold below are not explicitly present in the ICD, but logically included in the group “Disorders of urea cycle metabolism and ammonia detoxification”:

Argininemia ORPHA90

Argininosuccinic aciduria ORPHA23

Carbamoyl-phosphate synthetase 1 deficiency ORPHA147

Citrullinemia ORPHA187

Hyperammonemia due to N-acetylglutamate synthase deficiency ORPHA927

Hyperammonemic encephalopathy due to carbonic anhydrase VA deficiency ORPHA401948

Hyperinsulinism-hyperammonemia syndrome ORPHA35878

Hyperornithinemia-hyperammonemia-homocitrullinuria syndrome ORPHA415

Ornithine transcarbamylase deficiency ORPHA664

A possible solution for integration is to report for each RD entity described as an inclusion term in the ICD the corresponding Orphacode. And for the remaining, to report them as additional included entities (with their name and Orphacode).

Of note, “Hyperammonemia”, reported as an inclusion term in the ICD-10, is a generic term describing a clinical condition rather than a disease entity.

When an ICD code corresponds to an Orphacode representing not a RD entity but a branch, a direct trans-codification can still be performed. Nevertheless, it has to be considered that, according to the purposes of the coding activity, the user can further be able to choose an additional level of detail if a specific RD entity has to be reported.

3. In another possible existing situation, the ICD appears to be more specific than the Orphanet classification in the description of a RD entity (many-to-one relationship). In this case a RD entity matches well with an ICD code (usually with three digits), that is further sub-divided into related 4-digits codes.

Example:

D57 Sickle-cell disorders

Excl.: other haemoglobinopathies (D58.-)

D57.0 Sickle-cell anaemia with crisis

Incl.: Hb-SS disease with crisis

D57.1 Sickle-cell anaemia without crisis

Incl.: Sickle-cell:

- Anaemia NOS

- Disease NOS

- Disorder NOS

In these cases, the relationship is one Orphacode corresponding to a set of ICD codes. The further selection of which ICD depends on the user according to different mutually exclusive specific manifestations of the same disease. This situation deals generally with more frequent RD.

This is also the case when a rare disease entity can be double coded in the ICD-10 with the dagger and asterisk system.

Example: "Whipple disease" corresponds to "ORPHA3452"

and can be double coded in ICD-10 by the association of:

- (K90.8†) Other intestinal malabsorption - Whipple disease

- M14.8*Arthropathies in other specified diseases classified elsewhere - Whipple disease

4. In the majority of cases RD are not mentioned with their names, as reported in the Orphanet terminology, in the ICD-10. The integration process (see fig.1) passes through the choice of which is/are the most appropriate codes for the considered disease entity (many-to-many relationship). This process involves a careful examination of the context, starting from a disease knowledge in terms of its place in the Orphanet classification and the relationships with other RD entities described by a defined ICD code, of its main features, etiological mechanisms, etc. Of note, a specific disease can have a multiple parentage in Orphanet and thus can have a multiple parentage of different ICD codes. In these cases, the mapping is to be carefully discussed. Furthermore, it has to be underlined that this depends, among others, on the following contextual variables: coder expertise, setting, purpose of the coding activity.

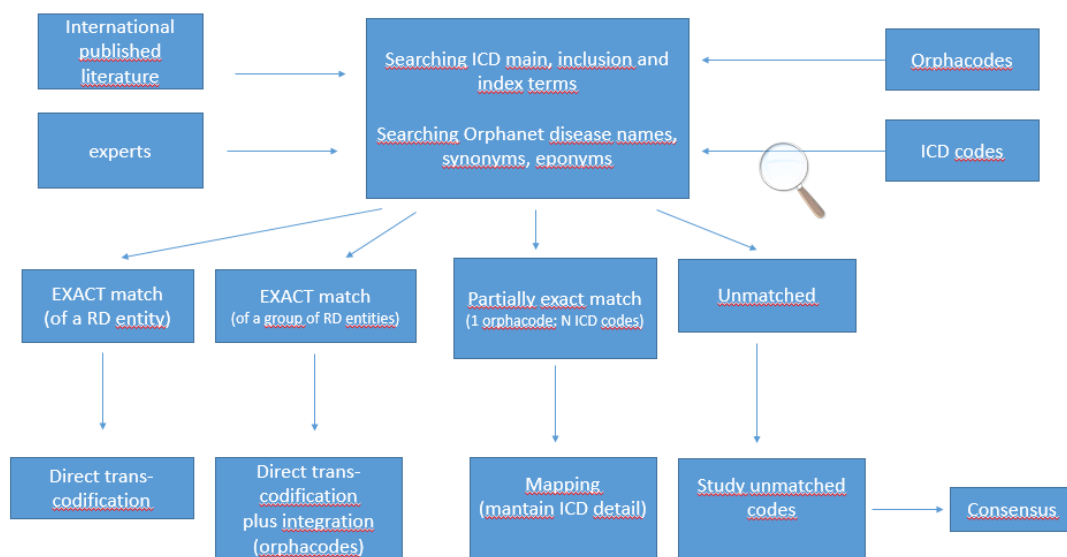


Fig. 1 Proposed flowchart for the integration process

Conclusions

1. A disease coding system's purpose is recording disease's diagnoses and relating them to a population, making them comparable in space and time. Other aims, concerning patients' health care, might not be attributed to a diseases coding system alone;
2. A disease coding system deals with a disease, diagnosed or suspected, with different granularity (a disease, a group of related diseases, subtypes of disease);
3. The disease coding system has to be specific for rare diseases, but linked in a non-casual way to the coding systems utilized in the current informative flows;
4. To perform the link to current informative flows, four directives can be used: one-to-one, one-to-many, many-to-many, many-to-one;
5. These directives are not alternative, but might be present altogether or be used in a different way, depending on the context and the aims of coding.

References

- ¹ Rath A., Olry A, Dhombres F, Brandt MM, Urbero B, Ayme S. Representation of rare diseases in health information systems: the Orphanet approach to serve a wide range of end users. *Hum Mutat.* 2012;33(5):803-8.
- ² Aymé S, Bellet B, Rath A. Rare diseases in ICD11: making rare diseases visible in health information systems through appropriate coding. *Orphanet J Rare Dis.* 2015; 10:35.
- ³ Commission Regulation (EU) No 328/2011 of 5 April 2011 implementing Regulation (EC) No 1338/2008 of the European Parliament and of the Council on Community statistics on public health and health and safety at work, as regards statistics on causes of death Text with EEA relevance.
- ⁴ Eurostat: http://ec.europa.eu/eurostat/statistics-explained/index.php/Health_statistics_at_regional_level.
- ⁵ EU Commission recommendation on cross-border interoperability of electronic health record systems http://ec.europa.eu/information_society/newsroom/cf/itemlongdetail.cfm?item_id=4214- [accessed June, 2016].
- ⁶ Liyanage H, Krause P, De Lusignan S. Using ontologies to improve semantic interoperability in health data. *J Innov Health Inform.* 2015; 22(2):309-15.
- ⁷ Madden R., Sykes C., Ustun B., World Health Organization Family of International Classifications: definition, scope and purpose. <http://www.who.int/classifications/en/FamilyDocument2007.pdf?ua=1> [accessed June, 2016].
- ⁸ Orphanet. ICD-10 coding rules for rare diseases. Procedural document. http://www.orpha.net/orphacom/cahiers/docs/GB/Orphanet_ICD10_coding_rules.pdf
- ⁹ WHO- International Statistical Classification of Diseases and Related Health Problems 10th Revision <http://apps.who.int/classifications/icd10/browse/2016/en> [accessed May, 2016]