



**Specification and  
implementation manual of the  
Master file for statistical  
reporting with ORPHAcodes**

**RD-CODE 1<sup>st</sup> update**

Based on the document produced by Work Package 5 of RD-ACTION



Co-funded by  
the Health Programme  
of the European Union

Updated on September 30<sup>th</sup> 2020

This update is part of the project / joint action '826607 / RD-CODE' which has received funding from the European Union's Health Program (2014-2020).

It has been produced by the members of the Work Package 5 and is part of the Task 5.4: Refined versions of the "Standard procedure and guide for the coding with Orphacodes" and the "Specification and implementation manual of the Master file".

It is based on the document produced by the members of Work Package 5 of the project / joint action '677024 / RD-ACTION' which has received funding from the European Union's Health Programme (2014-2020).

More information on the activities of RD-CODE can be found at <http://www.rd-code.eu>.

More information on the activities of the RD-ACTION Joint Action can be found at <http://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.

**Table of Content**

- 1. Introduction..... 1
- 2. Benefits of using the Master file for statistical reporting with ORPHAcodes ..... 1
- 3. Concept of the Master file for statistical reporting with ORPHAcodes ..... 2
- 4. Content definition of the Master file for statistical reporting with ORPHAcodes ..... 2
- 5. Description of the Master file for statistical reporting with ORPHAcodes..... 3
- 6. Using the Master file for statistical reporting with ORPHAcodes ..... 5
- 7. Guidelines..... 6
- 8. Release schedule ..... 6
- 9. Further information on the topic ..... 7

## 1. Introduction

The Specification and implementation manual of the Master file for statistical reporting with ORPHACodes (MF) has been developed in the frame of the Rare Disease Joint Action Workpackage 5 (WP5) “Steering, maintaining and promoting the adoption of ORPHACodes across Member States” (MS). It shall help to understand and work with the MF. With the Specification and implementation manual, it will be possible for ORPHACode implementing countries to test the MF, get support on the alignment of the pre-existing national coding system with the Orphanet classifications and give a feedback to the RD-CODE WP5-partners (contact details at the end of the document). According to the countries feedback the MF and the related documents will be further developed and adapted to the needs of MS. This feedback is given at RD-CODE workshops and WP5 conference calls. Furthermore a helpdesk was implemented on the RD-CODE website to receive additional feedback related to the Orphanet nomenclature content and the implementation of ORPHACodes in Health Information Systems (<http://www.rd-code.eu/helpdesk/>).

On a long-term view, the MF enables coding of rare diseases (RD) with a standardized code-set for national and international reporting for statistics as well as promoting an enhanced semantic interoperability between country specific IT systems and registries. It can also be used to support a standardized coding of RD patients on national and international level.

The benefits of using this manual as well as the concept of the MF and how to work with it are depicted. Additional to that, the guidelines, the release schedule and further information on the topic are described in the following chapters.

## 2. Benefits of using the Master file for statistical reporting with ORPHACodes

As already mentioned RD-ACTION WP5 developed a file in order to support RD data sharing across MS. Through the use of such file, namely the MF, the process of collection of coded RD-data should be simplified and structured. The MF should facilitate the standardised use of the Orphanet Nomenclature by providing only the data that is relevant for data sharing at EU-level and thus enable international statistical aggregation. Offering an alignment of disease names and their ORPHACodes and a terminal ICD-10 code in a dataset, should minimize the burden of using different classifications and support standardization. Furthermore, giving the MS the possibility to align such datasets to other classifications used locally for coding could give a higher added value to the file. Nevertheless, the MF allows an implementation at different levels so that all MS could use it independent of the available resources. (Please see [chapter 4](#) of this document and Deliverable 5.2 of the RD-ACTION ([http://www.rd-action.eu/wp-content/uploads/2017/05/D5.2\\_Standard-procedure-and-guide\\_final.pdf](http://www.rd-action.eu/wp-content/uploads/2017/05/D5.2_Standard-procedure-and-guide_final.pdf)) for further explanation)

Integrating the International Classification of Diseases (ICD) and the Orphanet nomenclature into one file simplifies the process of assigning an ORPHACode to a patient with a RD diagnosis. The coder does not have to select the codes from two separate coding systems but can select them simultaneously by selecting the relevant disease name as the MF combines both.

If the ORPHAcodes are integrated in a national coding system through other means, it should be ascertained that statistical reporting of coded RD-data will comply with the granularity and content of the data collected through integration of the MF into a national coding system. As the MF is compiled by using the Orphanet classifications, the same mechanism can be used to aggregate data with more granularity to the level of ORPHAcodes contained in the MF.

The linkage of the ORPHAcodes and the ICD codes (or other national coding systems) increases as well the standardization of rare diseases coding and consequently increases 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 on international level given the reduced number of patients.

At the end, the MF and the guidelines (see [chapter 7](#)) should contribute to data sharing at EU-level and ensure consistency in codification with ORPHAcodes amongst Member States.

### **3. Concept of the Master file for statistical reporting with ORPHAcodes**

The MF is a subset of the Orphanet classifications of RD that contains only the ORPHAcodes that should be used for data sharing and statistical purposes at EU-level. Orphanet produces and publishes the Orpha Nomenclature pack on an annual basis. It contains two XML data files, the Orphanet nomenclature file and the Orpha\_ICD10-mapping file, from which the MF can be produced (for details on how it is produced, see [chapter 5](#)).

The list of ORPHAcodes included in the MF can of course be part of a more complex setting but defines the minimum agreeable level for European reporting and comparison. Besides the implementation of the MF, users are free to use the whole Orphanet nomenclature to collect data in a more detailed way, if desired. This means that the MF could be implemented in the diverse data collection systems that use the Orphanet nomenclature. Emphasis should be placed on the possibility to automatically retrieve data at any time according to the ORPHAcodes in the MF and according to the basic coding rules and guidelines given by RD ACTION WP5 in Deliverable 5.2. Depending on the way the Orpha codification has been locally implemented, Orphanet classifications may be needed to reach the common level of granularity. Other tools towards this aim are available on the RD-CODE website (<http://www.rd-code.eu/tools/>).

### **4. Content definition of the Master file for statistical reporting with ORPHAcodes**

The Orphanet Nomenclature is a multidimensional, multihierarchical classification.

The classification consists of four different disjointed levels of ORPHAcodes: Category, Clinical groups, Disorders and Subtypes. Each level is further represented by typologies (i.e. “disorders” are represented by 6 different types, including “disease” and “malformation

syndrome” amongst others)<sup>1</sup> and have a decreasing complexity. This means that a disorder can be included in a category or a clinical group. It might be subdivided in different subtypes. The different levels are depicted in Figure 1.

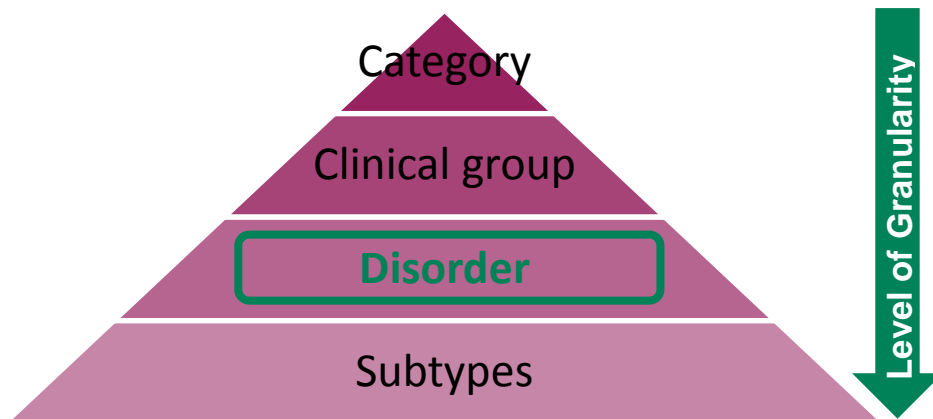


Figure 1: Different levels of the Orphanet classification

The MF was created by only selecting the disorder level. The categories, clinical groups and subtypes were discarded. As well, deprecated<sup>2</sup> and historical<sup>3</sup> entities were discarded.

For the creation of the MF it is also useful to know Orphanet’s “ICD-10 coding rules for rare diseases”. This document helps to understand the structure of the Orphanet nomenclature and why it enables interoperability, how the Orphanet nomenclature is produced and the set of rules according to which Orphanet attributes ICD-10 codes to disease entities and thus ORPHAcodes.

A detailed description of the “ICD-10 coding rules for rare diseases” ([https://www.orpha.net/orphacom/cahiers/docs/GB/Orphanet\\_ICD10\\_coding\\_rules.pdf](https://www.orpha.net/orphacom/cahiers/docs/GB/Orphanet_ICD10_coding_rules.pdf)) and of the “Procedures: Orphanet inventory of rare diseases” ([https://www.orpha.net/orphacom/cahiers/docs/GB/eproc\\_disease\\_inventory\\_R1\\_Nom\\_Dis\\_EP\\_04.pdf](https://www.orpha.net/orphacom/cahiers/docs/GB/eproc_disease_inventory_R1_Nom_Dis_EP_04.pdf)), can be found on the Orphanet website (see chapter 9).

## 5. Description of the Master file for statistical reporting with ORPHAcodes

The MF is an extract from the Orphadata and contains only the selected ORPHAcodes with the relevant information that should be used for data sharing at EU-level. For each

---

<sup>1</sup> Disorders include: disease, malformation syndrome, clinical syndrome, morphological anomaly, biological anomaly and particular clinical situation in a disease or syndrome.

Sub-type include: clinical sub-type, etiological sub-type and histopathological sub-type.

<sup>2</sup> “These are entities formerly considered as distinct diseases and now recognised as being part of another disorder present in the database.” (Source: <http://www.orphadata.org/cgi-bin/docs/userguide.pdf>)

<sup>3</sup> “Entries that were described long time ago, most of the time before the genetic era, and for which the princeps article is still available but no further literature exists that confirms their existence” (Source: <http://www.orphadata.org/cgi-bin/docs/userguide.pdf>)

ORPHAcode, the following information is available: ORPHAcode, Name, Synonym (s) and the ICD-10-WHO-code for most entities.

The MF can be extracted from the Orphanet nomenclature file (XML format) included in the Orphanet nomenclature pack, that can be found on the Orphadata website at “Orphanet nomenclature files for coding “ (<http://www.orphadata.org/cgi-bin/ORPHAnomenclature.html>). It is currently available in 9 different languages.

To get the content of the MF, follow these steps:

- 1) From the Orphanet nomenclature file, get only the Active ORPHAcodes at the Disorder level (FlagValue:1, 129 or 513 and GroupOfType id="36547") with their preferred term and synonym(s).
- 2) Then, from this data subset, query the Orphanet-ICD-10 cross referencing file to gather the cross referencing with the ICD-10 codes.

The description including examples can be found in the document “DESCRIPTION OF THE ORPHANET NOMENCLATURE FILES FOR CODING” by Orphanet. It is also provided on the Orphadata website at “Orphanet nomenclature files for coding”.

The resulting MF should contain the following columns:

**1<sup>st</sup> Column (ORPHAcode):**

This column contains the ORPHAcode, a unique and time-stable numerical identifier attributed randomly by the database upon creation of the entity. Their specificity should support data sharing at EU-level of RD for statistical purposes. Only active ORPHAcodes are available in the MF.

**2<sup>nd</sup> Column (Preferred term):**

This column contains the preferred term: the most generally accepted name according to the literature, and as adopted by the medical community. Every name is assigned to one ORPHAcode (1<sup>st</sup> column) by Orphanet. (To read more about how Orphanet attributes names to rare disorders, see “Rare disease nomenclature in English”; link in chapter 9 below)

**3<sup>rd</sup> Column (Synonym):**

This column contains terms that are perfectly equivalent to the preferred term (2<sup>nd</sup> column) they are attached to, and that are attached to the same ORPHAcode (1<sup>st</sup> column). As many synonyms as necessary are added to a preferred term. The number of synonyms can differ between the Nomenclature pack data sets of different languages (Orpha nomenclature pack version 2020 available in 9 different languages).

**4<sup>th</sup> Column (ICD-10 V2019):**

The column contains the ICD-10-WHO code in its most recent Version (Version 2019) that was mapped by Orphanet to a disorder, a group of disorders or subtypes,

according to the Orphanet ICD-10 coding rules (link: see chapter 9). The number of ICD-10 codes can differ from one to multiple codes.

In general, the number of lines for one ORPHAcode depends on the number of synonyms that are appendant to it (see table Table 1).

**Table 1: Variation of lines for different ORPHAcodes depends on the number of synonyms**

ORPHAcode	Name	Synonym	ICD-10 V2019
213	Cystinosis		E72.0
213	Cystinosis	Protein defect of cystin transport	
15	Achondroplasia		Q77.4
95	Friedreich ataxia		G11.1
95	Friedreich ataxia	FRDA	
95	Friedreich ataxia	FA	

For implementation into national coding systems it is not necessary to use the English disease names and ICD-10-WHO V2019 codes. If a country uses the ICD-10-WHO Version 2019 there is an existing translation in the Orphadata. If a country uses another disease coding system it might need to provide a respective mapping. The file used for integration into a national coding system can be structured just like the MF itself.

The MF data set can be extracted from the Nomenclature pack from the Orphadata website in XML-format (<http://www.orphadata.org/cgi-bin/ORPHAnomenclature.html>). Due to the fact that the MF is extracted from the Orphanet Nomenclature it only contains information provided by Orphanet. Only the translated disease names (if they cannot be extracted from the Orphadata) and national morbidity codes need to be added to integrate the file into routine coding systems. The codes that are included in the MF represent the minimal number of ORPHAcodes that should be used for data sharing of RD in MS in order to support the retrieve of statistical data at this level.

Countries should be encouraged to integrate the file within their national coding files and to make sure that if a national coding system is to be used together with the ORPHAcodes, both systems are aligned as much as possible to maximise standardisation of results and to help the coders to find the right code from both systems in one coding step. The addition of the different morbidity coding systems from all countries into one MF has proven not to be feasible as discussed and agreed during the WP5 Session of the Annual Meeting of the RD-Action in 2016. Therefore, an approach of centrally maintaining all MFs has not been followed further. It can be considered again, if necessary, in the future, as all prerequisites for this approach are in place.



## 6. Guidelines

In the framework of Deliverable 5.2 (Standard procedure and guide for the coding with ORPHAcodes), the RD-ACTION WP5 has published guidelines (link: see chapter 9), which are highly recommended to be used for the routine coding of RD. This procedure will create consistency and a standardized level of quality of coding as well as minimize the coding effort for users and the effort of statistical evaluations.

The guidelines extracted from D 5.2 are the following:

**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.

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

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

**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.

**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.

**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.

## 7. Release schedule

The MF should provide a frame for the data sharing at EU-level. This could be achieved through versioning of the file and regulation of the update cycle. With this, also the routine coding process could become more practicable and stable. 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 RD-ACTION WP5 team agreed that the MF should be published (updated) once a year. Actually, Orphanet releases the codification dataset (ORPHA nomenclature pack) annually in July/August. The publication of a file that should be used for data sharing between routine coding systems at EU-level in a more frequently rhythm could result in a bureaucratic burden for the implementation and lead to inconsistencies in the captured data.

The versioning information must be captured together with the data collected. Every year the list of codes (MF) 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 in many EU-countries. This presented file has to be implemented by the countries as well in order to cater to the frequent updates in the Orphanet Nomenclature.

Every year with the release of a new file the old file should be marked as old version but should be kept available online for historical reasons and as reference. Ideally a list of changes should be provided with the new file in order to ease implementation for the users (e.g. adaptation of the national morbidity codes).

## 8. Further information on the topic

To learn more about the work of the RD-CODE WP5 see <http://www.rd-code.eu/> and <http://www.rd-code.eu/workpackage-5-enhancements/>

To learn more about the work of the RD-ACTION please visit the RD-ACTION website ([www.rd-action.eu](http://www.rd-action.eu)).

To learn more on how the Orphanet nomenclature is produced, see “Procedure: Orphanet inventory of rare diseases” ([http://www.orpha.net/orphacom/cahiers/docs/GB/eproc\\_disease\\_inventory\\_PR\\_R1\\_Nom\\_04.pdf](http://www.orpha.net/orphacom/cahiers/docs/GB/eproc_disease_inventory_PR_R1_Nom_04.pdf)).

To read more about how Orphanet attributes names to rare disorders, see “Rare disease nomenclature in English” ([http://www.orpha.net/orphacom/cahiers/docs/GB/eproc\\_Disease\\_naming\\_rules\\_in\\_English\\_PR\\_R1\\_Nom\\_01.pdf](http://www.orpha.net/orphacom/cahiers/docs/GB/eproc_Disease_naming_rules_in_English_PR_R1_Nom_01.pdf)).

To learn more about the Orphanet ICD-10 coding rules for rare diseases please visit the Orphanet website (<http://www.orpha.net/consor/cgi-bin/Education.php?lng=EN>).

To learn more about the Orphadata and its content, see <http://www.orphadata.org/cgi-bin/docs/userguide.pdf>.

To learn more about the work of the RD-ACTION WP5, especially the Deliverables (D) and Milestones (MS, see <http://www.rd-action.eu/> and <http://www.rd-action.eu/leaflet-and-documents/>