Deliverable 3.2
Report on Genetic Annotations, ORPHAcodes alignments with other terminologies, and newly produced or updated texts

March 2026

Co-funded by the European Union

Co-Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or HADEA. Neither the European Union nor the granting authority can be held responsible for them

Table of contents

Introduction

The main scope of OD4RD2 Work Package 3 (WP3) is to improve the Orphanet scientific information collected in the Orphanet Knowledge base around the Nomenclature of rare diseases (RD). This is achieved though continuous survey of the literature and of reference resources and internal curation activity as well as though structured collaborations with European Reference Networks (ERNs), to accurately reflect the evolution of knowledge on RD. These activities contribute to enhancing the quality, consistency, and interoperability of RD data across health information systems and research infrastructures.

This summary report provides an overview of the main activities carried out over the entire duration of the OD4RD2 project (April 2023 - March 2026), focusing on key outputs related to the enrichment of the gene-disease knowledge base, the alignment of ORPHAcodes with international terminologies, and the production and validation of expert-reviewed textual information. The report is structured into three main sections:

1) **New discoveries around genes related to RD** and their representation in the Orphanet knowledge base, with references to data resources enabling reuse;

2) **RD coverage in other terminologies and classifications** including the alignment of ORPHAcodes with ICD-10, ICD-11, SNOMED CT and the Online Mendelian Inheritance in Man (OMIM) database, with links to the corresponding datasets;

3) **Production and update of rare disease textual information**, including definitions and summary descriptions, with an indication of contributions resulting from collaborations with ERNs.

1. New discoveries around genes related to RD

Methodology

Updates of the gene-diseases relationships in the Orphanet knowledge base are regularly performed by the Orphanet curation team according to a published methodology (link). These updates may also lead to a revision of the Orphanet nomenclature (addition or modification of an ORPHAcode), which follows the established methodology for updating the Orphanet nomenclature (link). When these updates impact nomenclature revision conducted in collaboration with ERNs, appropriate coordination is ensured with the activities of WP2 to avoid parallel revisions.

The complete Orphanet dataset of genes associated with rare diseases and their relationships is generated from the Orphanet knowledge base every 6 months (July and November) and is available for download and exploitation on the Orphadata Science website (Orphanet Scientific Knowledge Base section, a recognized Global Core Biodata Resource) (link).

Data report

The following indicators were calculated for the period going from April the 1st 2023 to March the 31st 2026.

Table 1 presents the updates performed on gene-diseases relationships during the reporting period. Independently of their disease-causing status, a total of 245 new genes were added to the Orphanet knowledge base (including 187 genes with a confirmed disease-causing relationship).

Overall, a total of 1530 curation actions (additions, modifications, and deletions of gene-diseases relationships) were performed during the project following systematic literature surveillance, external requests, internal quality control procedures and collaboration with ERNs. Of these, 818 were newly added relationships and 1 was a modification of an existing gene-disease relationship. In parallel, 711 gene-disease relationships were removed. A substantial proportion of these removed genes were associated with non-rare conditions.

Table 1: Gene-disease curation indicators and associated values for the reporting period

Number of new genes added to the Orphanet knowledge base	**245**
Newly added gene-disease relationships	**818**

Modification of existing gene-disease relationships	**1**
Suppressed gene-disease relationships	**711**

At the end of the project, a total of 4707 genes is found to be linked to at least one disease with a disease-causing relationship in the Orphanet knowledge base (Disease-causing germline mutation or Disease-causing somatic mutation). Of these, a total of 187 genes were newly linked during the project.

Table 2 reports the number of gene-disease relationships by relationship type at the beginning and at the end of the reporting period. Overall, an increase in the number of annotated relationships is observed, reflecting a net enrichment of the Orphanet knowledge base. This increase is primarily driven by disease-causing relationships, notably disease-causing germline mutations (from 5,170 to 5,339), disease-causing germline mutations loss-of-function (from 1,208 to 1,226) and disease-causing germline mutations gain-of-function (from 209 to 216). In line with the results provided in table 1, a substantial number of relationships were also removed because the corresponding genes were associated with non-rare conditions, impacting especially categories such as major susceptibility factor (516 to 461) and candidate gene tested (366 to 310), as expected.

Table 2: Number of gene-disease relationships at the start (April 2023) and end (March 2026) of the project, by type of gene-disease relationships

Type of gene-disease relationships	Number of gene-disease relationships at the start of the project	Number of gene-disease relationships at the end of the project
Disease-causing germline mutation(s)	5170	5339
Disease-causing germline mutation(s) (loss of function)	1208	1226
Disease-causing germline mutation(s) (gain of function)	209	216
Disease-causing somatic mutation(s)	226	236
Role in the phenotype	255	282
Part of a fusion gene	237	260
Modifying germline mutation	64	58
Biomarker tested	50	54
Candidate gene tested	366	310
Major susceptibility factor	516	461

Total of gene-disease relationship	**8301**	**8442**

In perspective, further efforts will focus on expanding gene coverage through the integration of a newly revised survey of the Online Mendelian Inheritance in Man (OMIM) database. This work will support the continuous updating of gene-disease relationships for newly described clinical conditions (rare disorders or subtypes of disorders) as well as conditions already included in the Orphanet knowledge base, as well as the OMIM mapping coverage (see Section 2, RD coverage in target terminologies).

In addition, gene-disease evaluation and validation in collaboration with ERNs will be further developed in the following project, to support both the update of existing relationships and the creation of new ORPHAcodes where appropriate. This approach is illustrated by the recent collaboration with ERN EURO-NMD on distal myopathies in relation with WP2 of this project, which led to the creation of 10 new ORPHAcodes with associated gene-disease relationships.

Finally, improved tracking of newly established gene-disease relationships generating newly added ORPHAcodes will be implemented to support monitoring within future project indicators.

2. RD coverage in target terminologies

Methodology

Alignment of ORPHAcodes with other terminologies in use in health systems and in registries is regularly performed by the Orphanet alignments project manager under the supervision of the Orphanet Scientific Director. The objectives of the OD4RD2 project were axed on four different terminologies, with different advancement statuses:

- ✓ Complete and maintain the alignment between ORPHAcodes and ICD-10 codes according to the rules already established and published here (link). This process implies the identification of the best suitable ICD-10 code and the definition of a semantic relationship between both codes.

- ✓ Continue and deliver for the first time the alignments between ORPHAcodes and ICD-11, in the frame of the participation of Inserm, US14-Orphanet to the French WHO collaborating centre. Two complementary approaches are followed to complete the alignments and to identify the gaps in ICD-11 in terms of rare disease coverage: a) First a syntactic approach based on language recognition algorithms developed by the Agence Numérique en Santé (ANS, lead of the WHO collaborating centre), was used to identify RDs already present in ICD-11 (Exact and Index matches). The results of the application of these algorithms were manually curated by Orphanet; b) a semantic approach, derived from the same rules used for ICD-10 alignments and needed to assess the proximity (exact or inexact) of both Orphanet and ICD-11 concepts. This approach is used in particular for RDs that are not represented in ICD-11, and must therefore be mapped to the closest clinically relevant ICD-11 MMS code (Attributed matching codes).

- ✓ Continue the mapping exercise between ORPHAcodes and SNOMED-CT in the frame of the collaboration agreement in place between Orphanet and SNOMED International, in order to have

a common release in October each year and containing new terms. This collaborative process focuses on aligning entities at the disorder level. The collaboration implies monthly follow-up meetings and exchange of information, in order to reach agreement on the concepts to be included in SNOMED-CT and mapped to ORPHAcodes (representing the exact same clinical entity). Orphanet releases a human-readable excel file with the updated mappings every year, and for the duration of the collaboration agreement.

✓ Continue the alignment between ORPHAcodes and OMIM concepts after capturing OMIM updates by running a script on a monthly basis, then manually curating the concepts in order to produce semantic relationships between concepts in both terminologies or to update existing ones. This activity can also result in initiating a demand for updating the Orphanet nomenclature, and follows the above-mentioned methodology. Should these updates impact the revision done in collaboration with ERNs, the appropriate connections are done to avoid parallel revision. In addition, a comprehensive review of OMIM codes not previously represented in Orphanet was carried out in 2025. This process enabled the manual assessment of each OMIM entry to determine the relevance and appropriateness of mapping with Orphanet.

Of note, outside the OD4RD2 project, a continuous effort is being made to align ORPHAcodes to other terminologies such as MeSH, UMLS, MedDRA, GARD and MONDO.

The complete Orphanet dataset of RD coverage in target terminologies and the relationships between ORPHAcodes to each of the target terminologies is generated every 6 months (July and November) except for SNOMED-CT, that is released annually in October. The Nomenclature Pack for coding, which is released annually (in July) contains the alignments to ICD-10, ICD-11 and OMIM. These different files available for download and exploitation at:

- the OrphadataScience website (Orphanet Scientific Knowledge Base section) (link) for ICD-10, ICD-11, OMIM, UMLS, MONDO, MeSH, MedDRA and GARD.
- the ORPHAcodes website for the SNOMED-CT mappings (link)
- the ORPHAcodes website for the Orphanet Nomenclature Pack (link)

Data report

The following indicators were calculated for the period going from April the 1st 2023 to March the 31th 2026.

RDs remain under-represented in medical terminologies and classification systems used within health systems, with only a small fraction of RD possessing specific and unambiguous codes. In particular, only a limited number of rare diseases are represented in ICD-10, while the representation of rare diseases has improved in more recent terminologies such as ICD-11 or SNOMED CT. However, coverage still remains below the number of currently known RDs.

To enable interoperability between different information systems, ORPHAcodes are mapped to international terminologies and classification systems. The alignment activity carried out during the project focused primarily on the disorder level, which represents the aggregation level for statistical purposes and data sharing across Europe. Results of the mapping activity are presented in Table 3, that contains the numbers and percentages per classification level of ORPHAcodes aligned to target terminologies at the start and end of the project.

ICD-10 alignments

At the beginning of the project, alignments between ORPHAcodes and ICD-10 codes already showed a high level of coverage (7288 of total clinical entities, namely groups of disorders, disorders or subtypes of disorders, mapped), which by the end of the project increased to 8484 mapped clinical entities.

Given that ICD-10 is widely implemented in health information systems and routinely used for coding, part of the work focused on performing quality control of existing mappings. Particular attention was given to multisystemic RDs, which involve several organ systems and therefore require careful assessment to determine the most appropriate mapping.

While the overall number of ICD-10 mappings increased, the proportion of aligned ORPHAcodes at the Disorder and Subtype level slightly decreased from 98,4% to 97,5% at the disorder level, and from 98,3% to 88,7% for the subtype level. This decrease reflects the evolution of the Orphanet nomenclature and classification, in particular the increase of the creation of new ORPHAcodes. The extensive ERN collaborations performed in WP2 contributed to the increase the number of ORPHAcodes to be mapped and therefore temporarily reduced the overall proportion of mapped entities. As expected, the proportion of aligned groups remains low, as groups are only mapped when an exact match exists in ICD-10.

ICD-11 alignments

Alignments coverage with ICD-11 increased substantially during the project. At the start of the project, alignments were performed by combining automated identification of potential syntactic matches (RDs explicitly represented in ICD-11), followed by manual validation.

For the RDs not represented in ICD-11, a semantic approach was applied to identify and attribute the closest clinically relevant ICD-11 code. This resulted in a significant increase in mappings, from 1711 of all ORPHAcodes aligned at the beginning of the project, to 6733 aligned ORPHAcodes at the end of the project. This increase corresponds more specifically to a rise in coverage from 20,5% to 77,5% at the disorder level, and from 10,5% to 78,9% at the subtype level, reflecting the majority of the mapping efforts with ICD-11 performed during the project. Similarly to ICD-10, the number of aligned groups remains low, as mappings are only established in the presence of exact matches.

Despite the substantial progress, some gaps remain, in particular for RDs that are not yet represented in ICD-11, which require the identification of codes that are the closest clinically relevant matches.

SNOMED-CT alignments

Alignments between ORPHAcodes and SNOMED CT codes are carried out as an ongoing collaboration with SNOMED International, in place since over 10 years ago. In the framework of this collaboration, mappings are focused on the disorder level and are restricted to exact matches, ensuring semantic equivalence between ORPHAcodes and SNOMED CT concepts.

During this project, the total number of aligned ORPHAcodes increased from 6397 to 6774, reflecting the continuous maintenance of mappings and the progressive improvement of RD representation in SNOMED CT. Following each modification of ORPHAcode, the mappings were reviewed, to determine

whether they remained an Exact match, and as a result, 32 mappings were removed. Furthermore, mappings were removed following each ORPHAcode inactivation.

OMIM alignments

OMIM is a key reference resource for genetic diseases. As Orphanet is clinically oriented, alignment with OMIM enables the representation of genetic information at the level of clinical entities rather than solely at the gene level.

As of the end of the project, alignment with OMIM covers 90.7% of genetic RD (at the disorder classification level) in Orphanet. A total of 6928 clinical entities (including groups of disorders, disorders and subtypes of disorders) in Orphanet are identified as having a known genetic basis, and only these can be associated with an OMIM number. Among these, 5006 are mapped to an OMIM code (4276 at disorder level, 676 at the subtype level and 54 at the group level).

To ensure continuous updating of alignments, new OMIM mappings are assessed both upon request and through a monthly automated OMIM survey script. This script supports the identification of updates, additions, and removals of OMIM phenotypes, ensuring that changes are accurately reflected in the Orphanet-OMIM mappings.

This script detects OMIM phenotypes not yet present in the Orphanet database and supports their evaluation. When the gene associated with an OMIM phenotype is already present in Orphanet, the script identifies candidate disorders sharing the same gene. If the corresponding disease is already represented, the OMIM phenotype is mapped to the appropriate clinical entity. If the gene is not present, the OMIM entry is reviewed as a potential candidate for creation in Orphanet. In cases where the disease exists in Orphanet but neither the OMIM phenotype nor the gene is linked to the corresponding ORPHAcode, the gene curation team assesses the gene-disease relationship.

During the project, a revision of the monthly OMIM survey script was undertaken. Moreover, we decided to conduct a comprehensive manual review of all OMIM phenotypes annotated with a number sign (#), which identifies phenotypic entries with a known genetic basis. This review, also called “T0 OMIM baseline”, focused on the analysis of OMIM phenotypes not yet mapped to Orphanet, and resulted in the identification of 544 new OMIM alignments and 676 possible new clinical entities to be evaluated for creation in the Orphanet nomenclature. The following phase will evaluate these entities one by one to evaluate inclusion in the nomenclature after consulting the medical literature. Moreover, 474 OMIM phenotypes were considered excluded from integration because they represented non-rare conditions, lacked characterization as distinct diseases, corresponded to historical entities, or were based on single case reports with one patient that did not meet the Orphanet inclusion criteria.

Table 3: ORPHAcodes aligned to target terminologies at the start and end of the project. Percentages are calculated on the total of Orphanet active clinical entities, by classification level, for ICD-10, ICD-11 and SNOMED-CT, while they are calculated on the total of Orphanet active genetic clinical entities for OMIM.

		Start of the project		End of the project	

ORPHAcodes aligned to ICD-10	**Total**	7288		**8484**	
	Disorders	6171	(98,4% aligned)	6492	(97,8% aligned)
	Subtypes of disorders	1003	(98,3% aligned)	998	(88,7% aligned)
	Groups of disorders	114	(5,34% aligned) *	132	(5,9% aligned) *
ORPHAcodes aligned to ICD-11:	**Total:**	1711		**6733**	
	Disorders	1285	(20,5% aligned)	4251	(77,5% aligned)
	Subtypes of disorders	108	(10,5% aligned)	884	(78,9% aligned)
	Groups of disorders	318	(14,9% aligned) *	415	(18,3% aligned) *
ORPHAcodes aligned to SNOMED-CT	**Total**	6397		**6774**	
	Disorders	5868	(93,5% aligned)	6169	(93,1% aligned)
	Subtypes of disorders	301	(29,5% aligned)	326	(29,0% aligned)
	Groups of disorders	228	(10,7% aligned) *	238	(10,6% aligned) *
ORPHAcodes (genetic entities only) aligned to OMIM	**Total**	4,526		**5006**	
	Disorders	3,915	(86,3% of genetic disorders)	4276	(90.7% of genetic disorders)
	Subtypes of disorders	559	(72,2% of genetic sutypes)	676	(80.9% of genetic sutypes)
	Groups of disorders	52	(3,90% of genetic groups) *	54	(3.9% of genetic groups) *

*Please note that for group of disorders, Orphanet only aligns the exact concepts in all terminologies

3. Newly produced definitions and newly produced or updated RD summary information

Methodology

Production or update of textual information related to rare disorders either as a continuous internal updating activity or in collaboration with European Reference Networks. This activity follows a general methodology published here (link).
Collaborations can be established with ERNs in order to prioritize the common work to maintain up-to-date summary information on rare diseases in Orphanet, as well as the definitions that accompany every rare disorder and that are delivered together with the nomenclature to help assigning the right ORPHAcode to a patient.

The collaboration with ERNs follows these steps:

- ✓ Initiation of collaboration: there are three points of entry 1) following the finalization of a collaboration on the Orphanet nomenclature; 2) The ERN establishes contact for a large collaboration independently of a nomenclature collaboration; 3) an ERN is approached directly by the Orphanet Editorial team to update a selection of diseases;

- ✓ Identification of training required: for large collaborations an initial introduction to Orphanet along with Editorial training is useful. For smaller collaborations, experts are directly issued guidelines on the task and text content;

- ✓ Establishing a list of diseases and text type (definition quality control or production/update of a disease abstract) according to specialty and predefined set of priorities;

- ✓ Identification of suitable experts by the ERN coordinating team or working group leads. Diseases involving multiple specialties may require cross-ERN collaboration;

- ✓ Cycle of revision and curation of text(s) between expert(s) and Editorial team, resulting in the validation and publication of text(s)
- ✓ Maintenance of disease summary information (ideally every 3 years)

- ✓ Review and feedback of collaboration: given punctually throughout collaboration and at the end of a collaboration cycle.

This methodology has been established based on the previous ERN collaboration experience and is adaptable to the needs and resources of each network.

Data report

The following indicators were calculated for the period going from April the 1st 2023 to March the 31st 2026.

A total of 1059 texts were produced/updated during the reporting period (883 published and 176 pending validation), meeting and surpassing the total expected objective of 1015 texts (divided into 725 expected new/updated texts and 290 expected quality-controlled texts). Among the produced texts, 387 were abstracts reviewed by at least one expert (321 published and 66 pending validation) and 505 were new/updated definitions (454 published, 51 pending validation) and 167 were quality-controlled definitions (Table 4).

Among the 387 abstracts, 225 (188 published and 37 pending validation) were produced in collaboration with ERNs, whereas among the 505 definitions, 152 of them (103 published, 49 pending validation) were produced in collaborations with ERNs, which corresponds to 58% of the produced abstracts and 30% of the produced definitions. In total, 20 ERNs were involved in the production of the 377 textual information which corresponds to 36% of the total text production (Table 4 & Table 5).

Orphanet currently has a formal, full-step collaboration with ERN-ITHACA, Endo-ERN and ERN-EpiCARE, as described in the methodology section. A formal collaboration with ERN-BOND is also in progress. In addition, following the completion of classification and nomenclature revision of certain group of disorders (such as “Rare vascular anomaly”, “Neuronal ceroid lipofuscinosis” and “Neural tube closure defects”), definitions of the revised entities were validated by VASCERN, ERN-EpiCARE & MetabERN, ERN-ITHACA & ERN eUROGEN respectively. Similar collaborations with ERN eUROGEN, ERN-NMD, ERN-PaedCan and ERN-EURACAN are in progress. Overall, 12 ERNs among the 20 listed in Table 5 participated in punctual non-structured collaborations to update texts.

The complete list of newly produced or updated RD summary information and definitions (with a focus on those realized in collaboration with ERNs) is available in Annex 2. In the Orphanet website, when a text is the result of an ERN collaboration, the name of the network appears next to the expert(s)' name(s) on the disease page where the experts are credited. This information can also be retrieved from the Orphanet knowledge base.

Table 4: Summary of produced texts including published or ongoing abstracts and definitions

Indicators (01.04.2023-31.03.2026)			Total
Total number of produced texts			**1059**
Abstracts (new/updated)	Published	321 (188 with ERNs)	**387**
	Ongoing	66 (37 with ERNs)	
Definitions (new/updated)	Published	454 (103 with ERNs)	**505**
	Ongoing	51 (49 with ERNs)	
Quality controlled definitions	Published	108	**167**
	Ongoing	59	

Table 5: Summary of the texts reviewed by an ERN-affiliated expert or part of an ongoing collaboration, divided by ERN

	Number of Published new & updated texts in collaboration with ERNs			Number of Ongoing new & updated texts in collaboration with ERNs		
ERN	**Total**	**Abstracts**	**Definitions**	**Total**	**Abstracts**	**Definitions**
Endo-ERN	18	14	4	20	11	9
ERN-EpiCARE	16	11	5	7	7	0
ERN-EpiCARE & MetabERN	12	0	12	0	0	0
ERKNet	1	1	0	0	0	0
ERN-CRANIO	16	7	9	0	0	0
ERN-GENTURIS	6	6	0	0	0	0
ERN RARE-LIVER	1	1	0	0	0	0
ERN-PaedCan	1	1	0	0	0	0
ERN-PaedCan & MetabERN	1	1	0	0	0	0
ERN-RITA	7	6	1	0	0	0
ERN-BOND	6	5	1	0	0	0
ERN-Eye	3	3	0	0	0	0
ERN-Skin	3	3	0	0	0	0
EuroBloodNet	22	19	3	2	2	0
ERN-eUROGEN	2	2	0	8	0	8
ERN-EURO-NMD	21	16	5	20	8	12
ERN-EURO-NMD & MetabERN	2	2	0	0	0	0
ERN-ITHACA	48	47	1	1	1	0
ERN-ITHACA & EndoERN	0	0	0	1	1	0
ERN-ITHACA & ERN-Skin	1	1	0	0	0	0
ERN-ITHACA & ERN eUROGEN	40	0	40	0	0	0
ITHACA & ERN-CRANIO	2	2	0	0	0	0
ERN-Lung	0	0	0	1	1	0
MetabERN	29	28	1	5	5	0
VASCERN	24	10	14	21	1	20
ERN-RND	2	2	0	0	0	0

ERN-EURACAN	6	0	6	0	0	0
EndoERN-ERN-BOND	1	0	1	0	0	0
TOTAL	**291**	**188**	**103**	**86**	**37**	**49**

Lessons learned and perspectives

The activities carried out during the OD4RD WP3 have highlighted several key areas for improvement and consolidation in future projects.

1) The close relationship between the activities conducted in WP2 and WP3, particularly in relation to the validation of definitions and disease summary information following nomenclature revisions, has demonstrated to be a point of strength for the project. Ameliorating the integration of workflows between WP2 and WP3 will be necessary to ensure better coordination between nomenclature updates and associated textual content, as well as to facilitate validation processes in collaboration with ERNs.
2) The mapping activities carried out during the project highlighted the importance of continuous quality control of mappings. Given the evolution of the Orphanet nomenclature, further efforts will be required to reinforce mapping quality control with collaborating partners and to define a more structured workflow for the validation of new and existing mappings in this framework.
3) The evolution of the Orphanet nomenclature, in particular through the creation of new ORPHAcodes in the context of ERN collaborations, has a direct impact on mapping coverage and indicators. This highlights the need to anticipate alignment updates in parallel with nomenclature changes, in order to maintain stable coverage levels across terminologies.
4) The activities related to the alignment with OMIM, including the revision of the T0 baseline and the development of the revised OMIM survey script, demonstrated the relevance of combining automated processes with manual curation. Future work will focus on consolidating this workflow, improving the follow-up of OMIM updates, and treat all possible mappings and creations of entries in the nomenclature resulting from the revision of the T0.
5) The collaboration with ERNs proved to be essential both for nomenclature evolution and for the production and validation of textual information. However, variability in the level and structure of engagement across ERNs indicates the need to further structure and harmonize collaboration processes in order to optimize expert contributions.