The Orphanet Nomenclature and Classification of rare diseases: a standard terminology for improved patient recognition and data interoperability

Caterina Lucano[1,*], David Lagorce[1], Annie Olry[1], Houda Ali[1], Valérie Lanneau[1], Mickael De Carvalho[1], Aysegul Dilsizoglu Senol[1], Marta Fructuoso[1], Emilie Gaillard[1,2], Marie-Cécile Gaillard[1], Seed Mihic[1], Mariana Tannoury[1], Florence Sauvage[1], Charlotte Rodwell[1], Sylvie Maiella[1], Marc Hanauer[1], Ana Rath[1,*]

[1]Inserm, US14-Orphanet, Paris, France

[2] Banque Nationale de Données Maladies Rares (BNDMR), Assistance publique - Hôpitaux de Paris (AP-HP), Paris, France

* Corresponding authors

E-mail: caterina.lucano@inserm.fr

E-mail: ana.rath@inserm.fr

Abstract

Background: Although individually uncommon, rare diseases (RD) collectively affect an estimated 329-624 million people worldwide There are over 6,500 known RD, 85% of which affect fewer than 1 person per million. As a result, the critical amount of data necessary to improve knowledge, care, and treatment can only be achieved through cumulative data collection across different countries in a standardized manner. However, RD are under-represented in medical terminologies and classification systems, hindering data sharing, interoperability, and public health monitoring.

Objective: The aim of this paper is to present the Orphanet Nomenclature and Classification of RD. We detail its content as well as its production and update methodology, and an overview of its mappings to other semantic resources. In addition, this work provides a clear and up-to-date count of RD based on the consensus operational definition of RD, and it details the distribution of RD by medical domain.

Methods: The Orphanet Nomenclature of RD is a multilingual standardized system composed of clinical entities, each defined by a unique and time-stable ORPHAcode, a preferred term, synonyms, a classification level, and a textual definition. This nomenclature is structured into three classification levels organized within a multi-hierarchical and multi-parental classification system by medical domain. Its production, updates, and mappings to major biomedical resources rely on standardized and published procedures, continuous literature review, manual curation and expert validation, reflecting advancements in RD knowledge and clinical practice. Presented data metrics were computed using the Orphanet July 2025 release to quantitatively characterize the content, structure, classification, and semantic alignments of the Orphanet Nomenclature and Classification system.

Results: As of July 2025, the Orphanet Nomenclature of RD includes a total of 9,784 active clinical entities, including 6,527 disorders (corresponding to the RD definition), 1,084 subtypes of disorders, and 2,173 groups of disorders. Disorders are multiclassified into 29 classification hierarchies, each corresponding to a distinct medical domain, accurately representing the complex multisystemic nature of RD. Extensive qualified mappings ensure semantic interoperability: 97.4% of disorders are mapped to at least one ICD-10 code (6.4% with an exact proximity relationship), 71.8% are mapped to at least one ICD-11 MMS code (14.7% with an exact relationship) and 94.8% are mapped to SNOMED CT (all with an exact relationship). Genetic disorders represent 72.2% of all RD, and 63.4% are mapped to at least one phenotypic OMIM number.

Conclusions: The Orphanet Nomenclature and Classification of RD is the only RD-specific interoperable medical terminology meeting the needs of healthcare, research, and public health systems. By addressing the underrepresentation of RD in medical terminologies, it enables accurate RD identification, coding, and monitoring, supporting cross-border data interoperability, and contributing to improved knowledge, policy-making, and ultimately better care for people living with a RD.

Keywords

Rare diseases, nomenclature, classification, ORPHAcodes, codification, interoperability, medical terminology

Introduction

Although rare diseases (RD) are individually uncommon, with over 6,500 different known RD, they collectively affect an estimated 329-624 million people worldwide [1,2]. Despite this significant global burden, available data indicate that 85% of RD affect less than 1 person per million [1]. As a result, the critical amount of data necessary to improve knowledge, care, and treatment can only be achieved through cumulative data collection across different countries in a standardized manner [3,4]. Cross-border data collection, sharing, and exploitation to support evidence-based policy making, healthcare, and research are therefore essential in the field of RD, as underscored by numerous international and European initiatives [5].

Over the past decades, numerous definitions of RD have been proposed, reflecting local legislations and public health priorities [4,6–11], most of which use point prevalence as the common epidemiological indicator. However, the threshold for rarity varies geographically, ranging from less than 1 per 2,500 in Japan [12], to less than 50 per 100,000 in Europe [13], or no more than 200,000 persons in the United States [14]. Furthermore, definitions can differ clinically, sometimes including restrictions such as “life-threatening” or “chronically debilitating” [6,13,15]. To address these inconsistencies, Rare Diseases International (RDI), in collaboration with a global panel of experts and the World Health Organization (WHO), has developed an Operational Description of Rare Diseases, defining a RD as a “medical condition with a specific pattern of clinical signs, symptoms, and findings that affect fewer than or equal to 1 in 2,000 persons living in any World Health Organization-defined region of the world”. [16]. Despite these efforts, the lack of harmonization among countries continues to impact RD research, diagnosis, and healthcare planning [3,5,8].

To take public health action on RD, it is fundamental to measure how many people are affected by these diseases in a defined population (prevalence), how these conditions impact those affected, and how to monitor their medical and societal consequences. [1,4,17]. In this regard, the United Nations Resolution on People Living with a Rare Disease (PLWRD) "encourages Member States and relevant United Nations agencies to collect, analyze, and disseminate disaggregated data on persons living with a rare disease... to assess progress towards the improvement of the status of persons living with a rare disease". This further highlights that adequate coding of RD patient data within health information systems is a prerequisite for an effective response to this global public health priority [18].

However, RD are under-represented in medical terminologies and classification systems used within health systems, with only a small fraction of RD possessing specific and unambiguous codes [19–22]. This limitation makes it difficult to trace RD patients in health information systems, preventing a robust understanding of the actual number of PLWRD. Only a small proportion of RD have a specific code in the 10th edition of the International Classification of Diseases (ICD-10) [20–22], the most widely used codification system worldwide; the representation of RD has increased in the 11th revision of the International Classification of Diseases (ICD-11) [21,23] and the Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) [24], though it still remains below the number of known RD. Moreover, RD cannot be easily identified in these reference systems, as they do not explicitly include any indication of the rarity of a disease. Finally, different regions use diverse semantic resources in their coding systems (e.g. ICD-10 or SNOMED CT), or diverse versions of the same resource (ICD local extensions, different ICD versions), hampering data sharing and interoperability. Given the already scarce nature of RD data, this barrier to data collection requires dedicated efforts to make RD recognizable in health information systems [25–27].

To tackle the challenges of RD codification and interoperability, Orphanet has developed and maintained a nomenclature of RD since 1997 and a classification of RD since 2006 [20,28]. Over time, adoption of the Orphanet Nomenclature of RD for RD codification within national health systems has been supported and endorsed by European and international policy initiatives aimed at improving RD identification and monitoring and RD cross-border data sharing. The Orphanet Nomenclature and Classification of RD is now recognized as the reference resource to support standardized RD codification across healthcare and research settings.

The aim of this paper is to present the Orphanet Nomenclature and Classification of RD, detailing its content as well as its production and update methodology, and an overview of its mappings to other medical terminologies and classification systems. Additionally, it provides a clear and up-to-date count of RD based on the above-mentioned consensus definition, and their distribution by medical domain. Finally, it discusses the importance of maintaining and adopting a standardized and reliable RD nomenclature and classification system to support clinical coding, research, healthcare interoperability, and policy-making.

Methods

Structural description of the Orphanet Nomenclature and Classification of RD

The core elements of the Orphanet Nomenclature and Classification of RD

The Orphanet Nomenclature of RD is a unique and multilingual standardized system aimed at providing a specific reference standard for RD, that is organized in a classification system, a multi-hierarchical and polyparental structure organized by medical domain

137 according to diagnostic and therapeutic relevance. The Orphanet Nomenclature of RD is
138 composed of clinical entities (a generic technical term used to describe the clinical items
139 included in the nomenclature), each of which includes several elements to ensure precise and
140 unambiguous representation of RD concepts (Figure 1) [29]:

141 - ORPHAcode: a unique and time-stable numerical identifier automatically assigned in
142 ascending order to every entry registered in the Orphanet Nomenclature;
143 - Preferred term (or main name): the most widely accepted name for the diagnosis in
144 the literature and/or in the medical community;
145 - Synonym(s): additional terms, including acronyms, fully equivalent to the preferred
146 term, as many as necessary according to the information found in the literature;
147 - Classification level: indicates the level of precision of the clinical entity, among three
148 possible granularity levels*:* Group of disorders, Disorder, or Subtype of disorder;
149 - Definition: a brief description of the major clinical characteristics that define the
150 diagnosis and differentiate it from other related clinical entities.

151

152

ORPHAcode	Preferred term	Synonyms	Classification level
ORPHA:2134	Atypical hemolytic uremic syndrome	Atypical HUS aHUS	Disorder
Definition	A rare, genetic thrombotic microangiopathy due to dysregulation of the alternative complement pathway and characterized by the triad of hemolytic anemia, thrombocytopenia, and acute renal dysfunction.		

153

154 **Figure 1. Technical elements composing the Orphanet Nomenclature of RD.**
155 Example of a clinical entity in the Orphanet Nomenclature of RD, defined by an ORPHAcode,
156 a preferred term, as many synonyms as needed, a classification level, and a textual definition.

157

158

159 ## Naming rules and multilingual translation of the Orphanet Nomenclature of RD

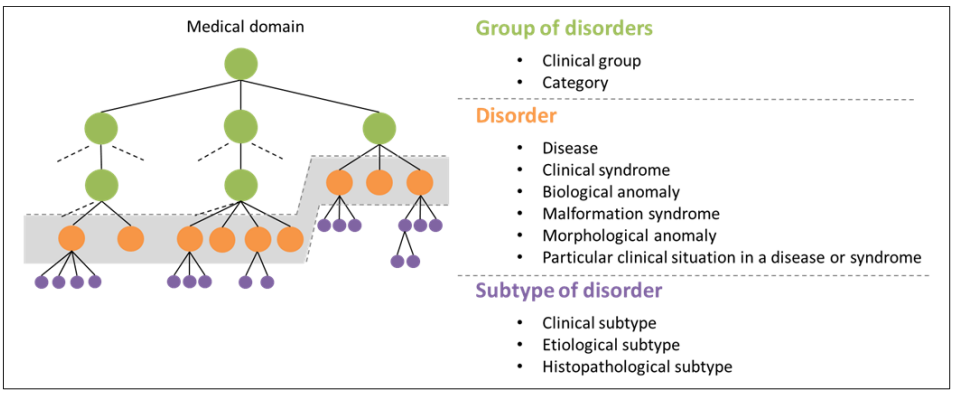
160 In the absence of an international consensus on how to name diseases, preferred
161 terms and synonyms are assigned according to standard naming rules developed by Orphanet
162 to create a comprehensive RD nomenclature that aligns with current knowledge and clinical
163 practice [30]. These rules ensure consistency across different disorder groups and medical
164 domains, facilitating better integration and use of RD-related clinical information in databases
165 and healthcare systems. The Orphanet Nomenclature of RD is produced in English (EN) and
166 entirely translated into eight languages, French (FR), German (DE), Spanish (ES), Italian (IT),
167 Dutch (NL), Portuguese (PT), Polish (PL) and Czech (CS), by translation teams located in
168 member countries of the Orphanet Network. All translations are validated by RD expert
169 medical doctors in each country. The English nomenclature serves as the basis for the
170 translation of concepts; however, the translation process considers the particularities and
171 local practices specific to every language, in particular with regards to the number of
172 synonyms [31].

173

174 ## Classification levels and typology

175 The Orphanet Classification system is composed of the three above-mentioned
176 classification levels, namely Group of disorders, Disorder, and Subtype of disorder (Figure 2).
177 The classification level serves as the basis for parent-child relationships between related
178 clinical entities. These relationships structure the Orphanet Classification according to
179 successive criteria that are relevant in the clinical setting [29]. Additionally, each clinical entity
180 is attributed a typology, a characteristic that, within the classification level, conveys the

181 nosological definition of the clinical entity. For Group of disorders, typologies include Clinical
182 group and Category; for Disorders, they include Disease, Clinical syndrome, Biological
183 anomaly, Malformation syndrome, Morphological anomaly, and Particular clinical situation in
184 a disease or syndrome; and for Subtype of disorder, they include Clinical subtype, Etiological
185 subtype, and Histopathological subtype (Figure 2, definitions in Additional file 4) [29]. Of note,
186 while generally used interchangeably to accommodate the widespread use of the term "rare
187 diseases", the terms "disease" and "disorder" designate different concepts within the
188 Orphanet Nomenclature of RD. The term "disorder" is used to refer to the classification level
189 that includes the six different typologies mentioned above, while the term "disease"
190 represents a specific typology and a narrower clinical concept within the "disorder"
191 classification level (Figure 2, Additional file 4) [29]. In the Orphanet Nomenclature and
192 Classification system, the "disorder" level corresponds to the WHO-RDI consensus definition
193 of "rare disease" [16].
194
195



196

197 **Figure 2. Classification levels and typologies of the Orphanet Nomenclature and**
198 **Classification of RD.**

199 Every clinical entity in the Orphanet Nomenclature and Classification of RD is associated with
200 a classification level conveying the precision of the diagnosis (Group of disorder, Disorder, or
201 Subtype of disorder), and a unique typology within the classification level.
202
203

204 Production and update process of the Orphanet Nomenclature and Classification of 205 RD

206 Curation and update procedure of the Orphanet Nomenclature and Classification of RD

207 The Orphanet Nomenclature and Classification of RD is produced and updated
208 according to a standardized and published procedure, in order to maintain an optimal quality
209 of information, by a scientific team of information scientists at the Institut national de la santé
210 et de la recherche médicale (Inserm) US14 service unit, Paris, France. The information
211 provided in the Orphanet Nomenclature is sourced from up-to-date scientific and medical
212 literature (more than 2,000 papers are consulted yearly), manually curated, reviewed and
213 validated by Orphanet information scientists and RD medical experts, [29,32].

214 The maintenance of the Orphanet Nomenclature of RD is ensured through three main
215 types of actions:

216 - Inclusion of a new clinical entity: to be included in the Orphanet Nomenclature, a
217 disorder level clinical entity must correspond to a well-delineated clinical presentation,
218 have a prevalence of fewer than or equal to 1 in 2,000 persons, and be described in at
219 least two unrelated individuals in the literature, or two individuals from the same
220 family when there is a clear genetic cause identified. Groups of disorders and subtypes
221 of disorders are included according to established clinical practice and evolving

222 medical knowledge. Before inclusion, the clinical entity is compared with existing
223 entities to avoid overlaps and ensure classification consistency;
224 - Modification of an existing entity: a clinical entity in the Orphanet Nomenclature of RD
225 can be modified to improve or update its information, including the preferred term,
226 synonyms, classification level, typology or the definition. Clinical entities described
227 before the genetic era, and lacking publications since then, are labeled as "Historic"
228 and are kept in the Orphanet Nomenclature because their initial descriptions are still
229 available and carry a potential for further characterization;
230 - Inactivation of a clinical entity: a clinical entity in Orphanet can be inactivated due to
231 deprecation (it is now included in another concept), obsolescence (it has no valid
232 reason to exist, such as being an imprecise or duplicate concept), or for no longer
233 meeting the rare disorder prevalence criteria.
234

235 ## Management of inactivation, reactivation and replacement of ORPHAcodes

236 An ORPHAcode is permanently linked to the same clinical concept and is never deleted nor
237 reused for a different entity, even if the original entity becomes inactive. In rare instances, an
238 inactive clinical entity may be reactivated after reexamination with updated knowledge and
239 reintroduced into the Orphanet Nomenclature and Classification with all relevant scientific
240 information. To ensure continuity of information, especially in patient records and coding
241 systems, Orphanet provides a replacement active clinical entity for deprecated and obsolete
242 entities. This replacement code, or target code, must be used instead of the corresponding
243 deprecated clinical entity as it aligns with the evolution of knowledge in the literature. For
244 obsolete entities, the target often refers to a group of disorders, guiding users to explore the
245 relevant classification to identify the most appropriate replacement code based on the

246 patient's diagnosis. Entities inactivated as non-rare in Europe are not given a replacement
247 code, as they no longer fall under the RD scope.
248

249 Classification of RD by medical domain

250 The Orphanet Classification hierarchies

251 The Orphanet Classification is structured in 34 major classification groups, individually
252 referred to as classification hierarchy. Of these, 29 represent a distinct medical domain and
253 are listed in Additional file 5. The remaining 5 hierarchies are ad-hoc classifications designed
254 primarily for organizational purposes:

255 - Rare genetic diseases: this classification does not represent a distinct medical domain
256 but is rather a derived genetic version that mirrors all the other clinical classification
257 hierarchies.
258 - Rare systemic and rheumatological diseases of childhood: this classification is fully
259 included in the Orphanet classification of rare systemic or rheumatologic disorders;
260 - Rare transplant-related diseases: this classification does not represent a medical
261 domain but is used to classify disorders according to an intervention.
262 - Rare teratologic disorders: this classification is fully included in the Orphanet
263 classification of rare developmental anomalies during embryogenesis.
264 - Rare disorder without a determined diagnosis after full investigation: this classification
265 does not represent a medical domain and is used to classify one single disorder
266 ORPHA:6616874, Rare disorder without a determined diagnosis after full investigation
267 [33]. This is a code that does not conceptually fall under any of the other Orphanet
268 classification hierarchies, as it allows for the identification in healthcare information
269 systems of patients who remain undiagnosed after a full investigation.

270 Every classification consists of a tree structure starting with a unique ultimate parent group at
271 the top representing the broadest clinical concept/medical domain (head of classification),
272 then branching out into as many intermediate groups as necessary with increasing precision
273 of clinical concepts, until reaching the first level of the confirmed diagnosis, the disorder. For
274 any given disorder, subtypes can be added when this level of precision is required in clinical
275 practice for appropriate management and follow-up.
276

277 ## Multi-hierarchical classification and linearization

278 Every clinical entity in the Orphanet Nomenclature of RD is introduced in as many
279 classification hierarchies as necessary according to major medical specialties involved in the
280 management of the disorder. This results in a multi-hierarchical and multi-parental
281 classification structure that reflects the multi-dimensionality and complexity of RD (Figure 3).
282 However, only one ultimate parent group is chosen as the entity's preferential classification
283 (preferential parent), according to the most severely affected body system, the most
284 determining involvement for the prognosis, and/or the medical specialty that will most likely
285 be relied on for the management of the disorder. This process, named linearization, enables
286 coherent statistical analysis. For a substantial number of RD, no firm scientific basis exists to
287 support assignment to a single medical domain, making preferential parent selection
288 inherently complex. Linearization in these situations is a conventional framework guided by a
289 set of published rules aimed at limiting arbitrariness and ensuring consistent application [34].
290
291

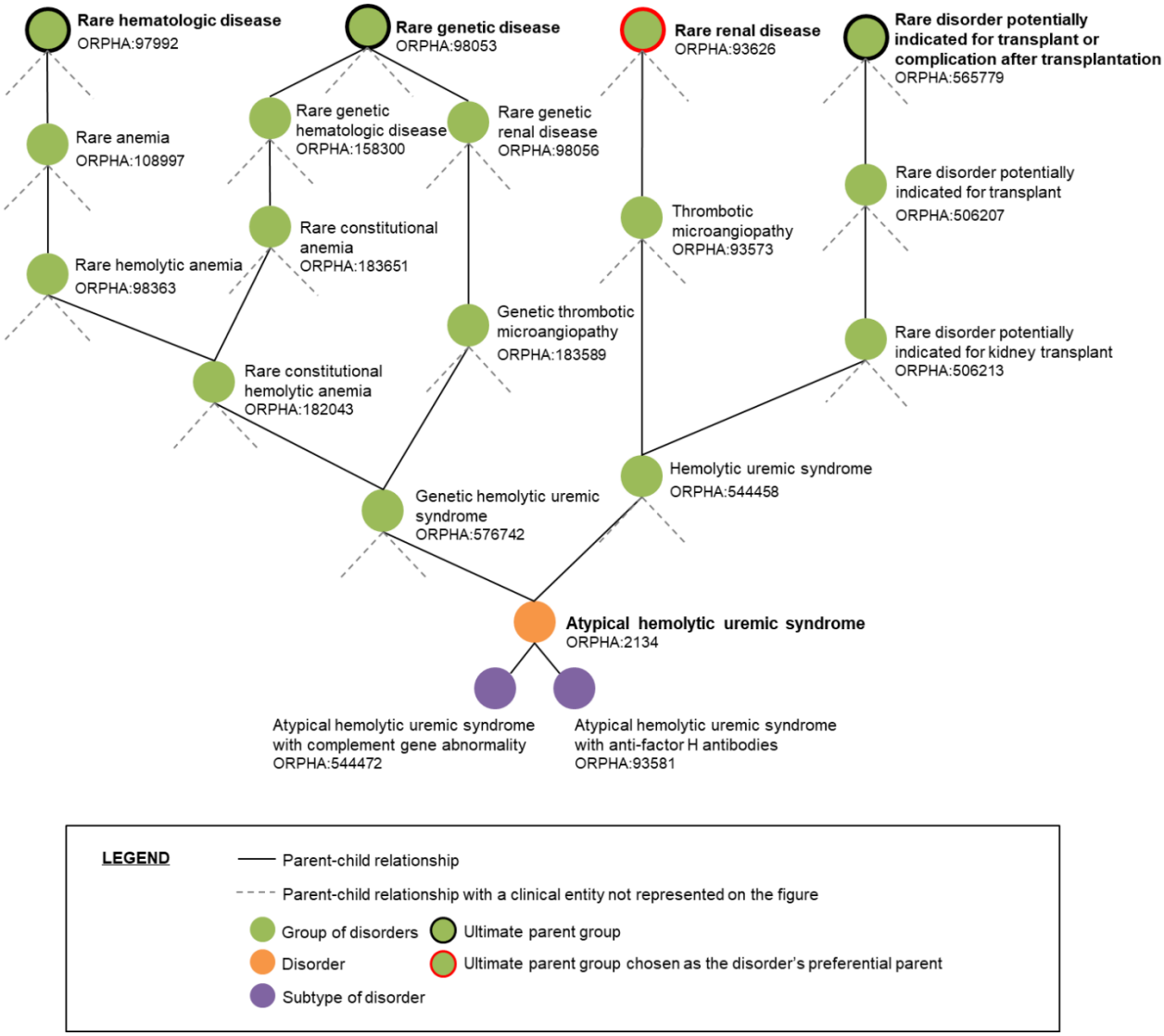


Figure 3. Representation of a disorder in the Orphanet Classification system.

The disorder Atypical Hemolytic Syndrome (ORPHA:2134) is classified into four classification hierarchies. It has two direct parent groups, each with broader parent groups, culminating at the top of the hierarchy, where each ultimate parent group represents a medical domain. The Rare Renal Disorder group (ORPHA:93626) was designated as the preferential parent of the disorder Atypical Hemolytic Syndrome (ORPHA:2134). Note that all groups in the figure have additional child entities not shown here.

302 ## Mappings to other semantic resources

303 The Orphanet Nomenclature of RD is mapped to other international terminologies,
304 classification systems, controlled vocabularies, knowledge bases, and ontologies according to
305 a standardized and published procedure, to enable interoperability between different
306 information systems [35]. Target standards include the WHO classification systems ICD-10 [36]
307 and ICD-11, both Mortality and Morbidity Statistics (MMS) and Foundation [37], the clinical
308 terminology SNOMED CT [38,39], the Online Mendelian Inheritance in Man (OMIM)
309 knowledge base [40], the metathesaurus Unified Medical Language System (UMLS) [41], the
310 controlled vocabularies Medical Subject Headings (MeSH) [42] and Medical Dictionary for
311 Regulatory Activities (MedDRA) [43], the NIH Genetic and Rare Diseases Information Center
312 (GARD) knowledge base [44], and the Mondo Disease Ontology (Mondo) [45].

313 The mapping cardinality always goes from the ORPHAcode to a code of the target resource.
314 All mappings are qualified by a proximity relationship that defines the nature of the
315 correspondence between the two concepts. The adopted proximity relationships are:

316 - E (Exact): the Orphanet entity designated by an ORPHAcode and the target code have
317 the same range of application; they describe the same pathological entity;
318 - BTNT (Broader Term to Narrow Term): the Orphanet entity designated by an
319 ORPHAcode has a broader range than the target code;
320 - NTBT (Narrow Term to Broader Term): the Orphanet entity designated by an
321 ORPHAcode has a narrower range than the target code.

322 Mappings are generally validated at the Orphanet disorder and subtype levels; however, when
323 an exact corresponding concept exists in the target resource, mappings may also be retained
324 at the group level. For ICD-10, ICD-11 and OMIM, mappings are maintained with all three

325 possible proximity relationships, whereas for SNOMED CT, UMLS, MeSH, MedDRA, GARD, and
326 Mondo, only exact matches are retained.
327 The mappings with the WHO ICD-10 and ICD-11 are further characterized by a specificity
328 relationship, describing how the equivalent clinical concept is represented in ICD [35]. The
329 relationships are defined as follows:

330 - Specific code: the clinical concept represented by the ORPHAcode corresponds to a
331 code in the ICD-10 tabular list/ICD-11 MMS exclusively denoting the concept;
332 - Index term: the clinical concept represented by the ORPHAcode is listed as an entry
333 term in the ICD-10 alphabetical index or in the ICD-11 Foundation, without being
334 assigned a dedicated tabular list/MMS code;
335 - Inclusion term (ICD-10 only): the clinical concept represented by the ORPHAcode
336 appears as an inclusion term under an ICD-10 term in the tabular list without a distinct
337 code;
338 - Attributed code: in the absence of an explicit corresponding term in ICD-10 or ICD-11
339 for the clinical concept represented by the ORPHAcode, an ICD code is attributed by
340 Orphanet as the closest available clinical concept.

341 All mappings are manually curated and validated by an Orphanet information scientist and a
342 medical doctor before publication, unless explicitly marked as *Not yet validated* to indicate
343 pending medical review. Mappings with SNOMED CT concepts result from a collaborative
344 process with SNOMED International, in which Orphanet disorder concepts are first evaluated
345 for inclusion in SNOMED CT by a SNOMED International terminologist, and the resulting exact
346 mappings are subsequently validated jointly by SNOMED International and Orphanet before
347 publication. Notably, only mappings to ICD-10, ICD-11 and OMIM are distributed together

348 with the official Orphanet Nomenclature files for coding, while the entirety of the mappings
349 are distributed through Orphadata Science.
350
351
352 ## Data extraction and analysis
353 ### Datasets description
354 Data metrics were computed using the Orphanet Scientific Knowledge Files delivered
355 on the Orphadata Science website [46]. These datasets, structured in an Extensible Markup
356 Language (XML) file format and in JSON format, are released in July every year at the same
357 time as the official Orphanet Nomenclature files for coding (Nomenclature Pack) [47]. For this
358 study, we analyzed the following datasets of the July 2025 release:
359 - the "rare diseases and alignments with terminologies and databases" dataset
360 (<iso>_product1.xml) in English, Czech, Dutch, French, German, Italian, Polish,
361 Portuguese and Spanish languages; version 1.3.42 / 4.1.8 [2025-03-03] released 24 Jun
362 2025;
363 - the "linearization of rare diseases" English dataset (en_product7.xml); version 1.3.42 /
364 4.1.8 [2025-03-03] released 24 Jun 2025;
365 - the "genes associated with rare diseases" English dataset (en_product6.xml); version
366 1.3.42 / 4.1.8 [2025-03-03] released 24 Jun 2025;
367 - The "classifications of rare diseases" English datasets (en_product3_<x>.xml); version
368 1.3.42 / 4.1.8 [2025-03-03] released 24 Jun 2025;
369 Moreover, the differential files of the Orphanet Nomenclature packs for coding
370 (ORPHAnomenclature_diff_en.xlsx) released from 2020 to 2025 were used to manually
371 retrieve the annual inclusion and inactivation of disorders, and the human readable Orphanet-

372 SNOMED-CT mapping file (ORPHA-SNOMEDCT_Mapping_File_production.xlsx), July 2025
373 version, released in October 2025, was used to report on the Orphanet-SNOMED CT mappings.
374

375 ## Metrics calculation

376 Metrics were computed using Python 3.8 scripts made available as dedicated Jupyter
377 Notebooks (.ipynb files) in a Binder/Github-ready repository
378 (https://github.com/Orphanet/orphanet_nomenclature_binder). This repository provides a
379 publicly accessible, reproducible, and interactive computational environment that enables on-
380 demand calculation of metrics using Orphadata files. A total of eight scripts were developed
381 to calculate the necessary metrics and were annotated to document the calculation
382 methodology. Metrics not computable via scripting were calculated manually, as detailed
383 below.

384 The distribution of active clinical entities by classification level and typology was
385 calculated using the script "1_number_of_Orphanet_clinical_entities". This script uses the
386 "rare diseases and alignments with terminologies and databases" dataset in English, filters
387 active and inactive clinical entities, and subsequently counts them by classification level and
388 typology. The same dataset was used to manually calculate the number of historical clinical
389 entities by filtering for the "Historical entity" annotation.

390 The distribution of synonyms by classification level in English, as well as the number of
391 adapted synonyms in Czech, Dutch, French, German, Italian, Polish, Portuguese and Spanish
392 was calculated with the script
393 "2_production_of_the_Orphanet_RD_terminology_in_english_and_other_language". This
394 script uses the "rare diseases and alignments with terminologies and databases" dataset in all

395 target languages and sequentially filters and counts annotations for active clinical entities, and
396 retrieves and counts their synonyms, grouping them by classification level.

397 The number of translated preferred terms in Czech, Dutch, French, German, Italian,
398 Polish, Portuguese and Spanish was calculated manually by comparing the preferred terms of
399 active clinical entities in each target language with the corresponding English preferred terms
400 in the "rare diseases and alignments with terminologies and databases". Terms differing from
401 the corresponding English term were considered translated; terms identical to the
402 corresponding English term were considered untranslated, with the exception of acronyms
403 and Latin terms, which were considered translated.

404 The coverage of definitions at the disorder level in each language was calculated using
405 the script "8_number_of_Orphanet_RD_with_definition_in_all_languages". This script uses
406 the "rare diseases and alignments with terminologies and databases" dataset in all target
407 languages, filters active clinical entities at the disorder classification level, and retrieves and
408 counts the corresponding definitions.

409 The annual disorder inclusions and inactivations for the period 2020–2025 were
410 manually retrieved from the differential files of the Orphanet Nomenclature packs for coding
411 released between 2020 and 2025.

412 The number of disorders classified in more than one classification hierarchy versus those
413 classified in a single classification hierarchy was calculated using the script
414 "4_number_of_Orphanet_RD_monoclassified_vs_multiclassified" which uses the
415 "classifications of rare diseases" datasets to select disorders and assess their inclusion in one
416 or multiple hierarchies. The same datasets are exploited by the script
417 "5_distribution_of_Orphanet_RD_in_Orphanet_classifications" to retrieve and count the
418 number of disorders included in each classification hierarchy. Of the 34 classification

419 hierarchies in the Orphanet classification system, only 29 were included in the analysis. The
420 previously described 5 hierarchies not corresponding to distinct medical domains or
421 representing organizational or non-clinical groupings, were excluded. The distribution of
422 disorders by preferential parent was calculated using the script
423 "6_distribution_of_Orphanet_RD_by_preferential_parent". This script selects all active
424 disorders from the "rare diseases and alignments with terminologies and databases" dataset
425 and uses the "linearization of rare diseases" dataset to retrieve their preferential parent,
426 subsequently counting the total number of disorders linearized for each classification
427 hierarchy.

428 The proportion of genetic disorders was calculated using the script
429 "3_proportion_of_genetic_versus_non_genetic_Orphanet_RD". This script retrieves all active
430 disorders from the "rare diseases and alignments with terminologies and databases" dataset
431 and assesses their inclusion in the "genes associated with rare diseases" dataset and the
432 "classifications of rare diseases - rare genetic diseases classification" dataset. Inclusion in
433 either dataset was considered indicative of a genetic disorder.

434 Mappings to other semantic standards were retrieved using the script
435 "7_alignment_between_Orphanet_RD_and_international_terminologies" for ICD-10, ICD-11,
436 and OMIM, and manually for UMLS, MeSH, MedDRA, GARD, and Mondo. In both cases, the
437 "rare diseases and alignments with terminologies and databases" dataset was filtered by
438 target standard to retrieve mappings at the Orphanet disorder level and their relationships.
439 Orphanet-SNOMED CT mappings were retrieved manually by filtering the human-readable
440 Orphanet-SNOMED-CT mapping file for mappings at the disorder level.

441

442

443 # Results

444 ## Quantitative characterization of the Orphanet Nomenclature and Classification of
445 RD

446 Counting and distribution of active clinical entities in the Orphanet Nomenclature and
447 Classification of RD by classification level and typology

448 As of July 2025, the Orphanet Nomenclature of RD contains a total of 9,784 active
449 clinical entities, including 6,527 disorders corresponding to the consensus definition of "rare
450 disease" [16] (among which is a specific disorder code for undiagnosed patients after full
451 investigation [33]), 1,084 subtypes of disorders, and 2,173 groups of disorders. Among all
452 active entities, 3.3% (325 disorders and 1 subtype) are considered historic, with no
453 publications over the last 25 years. The distribution of all active clinical entities by their
454 classification level and their typology is presented in Table 1 and in Additional file 1. Definitions
455 of each typology are provided in Additional file 4.

456

457

458 **Table 1. Distribution of active clinical entities in the Orphanet Nomenclature of RD by**
459 **classification level and typology.**

Classification levels		Number of active clinical entities	% on total clinical entities	% on each classification level
Disorders		6,527	66.7%	
	Diseases	4,172		63.9%
	Clinical syndromes	50		0.8%

	Malformation syndromes	1,788		27.4%
	Morphological anomalies	446		6.8%
	Biological anomalies	11		0.2%
	Particular clinical situations in a disease or syndrome	60		0.9%
Subtypes of disorders		1,084	11.1%	
	Clinical subtypes	839		77.4%
	Etiological subtypes	195		18.0%
	Histopathological subtypes	50		4.6%
Groups of disorders		2,173	22.2%	
	Clinical groups	450		20.7%
	Categories	1,723		79.3%

Counting and distribution of English synonyms across active clinical entities in the Orphanet Nomenclature of RD

Of the 9,784 active clinical entities in the Orphanet Nomenclature of RD, 6,311 (64.5%) carry at least one synonym in English, while the remaining 3,473 (35.5%) are not associated with any synonym. Of the 6,527 disorders, 4,805 (73.6%) carry at least one synonym in English, with a total number of synonyms at the disorder level amounting to 10,547. The minimum number of synonyms associated with a disorder is 1, and the maximum is 15, with a mean of

469 2.2 synonyms for disorders with at least one synonym. The complete distribution of synonyms
470 by classification level is reported in Additional file 2.
471
472 ## Counting and distribution of translated preferred terms and synonyms across translation
473 languages
474 All preferred terms associated with active clinical entities registered in the Orphanet
475 Nomenclature of RD are translated in the following languages: French, German, Spanish,
476 Italian, Dutch, Portuguese, Polish and Czech. Table 2 shows the rate of translation of preferred
477 terms for every language. The very minor delay in translation is due to the different timings
478 and frequencies at which translations are processed by every translation team. Non translated
479 preferred terms are nevertheless included in English, pending translation. Synonyms can be
480 adapted in all translated languages according to the medical practice in the target language;
481 the number of adapted synonyms for every translation language is shown in Table 2. The
482 complete distribution of synonyms by classification level in all translated languages is reported
483 in Additional file 2. Additional translations are also available in Chinese (9,158 translated
484 preferred terms, last update June 2020), Turkish (9,168 translated preferred terms, last
485 update September 2021) and Ukrainian (9,612 translated preferred terms, last update July
486 2024), and are provided in separate files in the "Rare Diseases and Alignment with
487 Terminologies and Databases" section of Orphadata Science [46].
488
489
490 **Table 2. Number of active clinical entities with translated preferred term and adapted**
491 **synonyms by language.**

	FR	DE	ES	IT	NL	PT	PL	CS
Number of translated preferred terms	9,784	9,778	9,784	9,569	9,777	9,576	9,605	9,628
% of translated preferred terms*	100%	99.9%	100%	97.8%	99.9%	97.9%	98.2%	98.4%
Number of adapted synonyms	13,000	12,854	12,912	11,234	14,546	5,858	12,804	14,298

*Percentages are calculated relative to the total number of active clinical entities with English preferred terms (n = 9,784).

Coverage and distribution of disorder definitions across languages

Disorder definitions are produced in English for all disorders included in the Orphanet Nomenclature of RD and are subsequently translated into the eight translation languages: French, German, Spanish, Italian, Dutch, Portuguese, Polish and Czech. Table 3 presents the coverage of disorder definitions at the disorder level in each language. Currently, approximately 94% of rare disorders have a definition in English, while the rate of coverage in other languages depends on the priorities defined by each translating country.

Table 3. Number of disorders with available definitions per language.

	EN	FR	DE	ES	IT	NL	PT	PL	CS

Number of disorders with a definition	6,116	5,941	4,572	6,095	4,592	6,068	1,186	1,372	17
% on the total number of disorders (n=6,527)	93.7%	91.0%	70.1%	93.4%	70.4%	92.9%	18.2%	21.0%	0.26%

506

507

508 ## Annual inclusion and inactivation of disorders in the Orphanet Nomenclature of RD over

509 the last 6 years (2020-2025)

510 Since 2020, an average of 123 ± 72 (mean ± SD) new disorders per year have been

511 added to the Orphanet Nomenclature, while an average of 44 ± 18 disorders per year have

512 been inactivated. A detailed report of the number of newly included, obsoleted, deprecated

513 and non-rare disorders for each year from 2020 to 2025 is provided in Table 4.

514

515

516 **Table 4. Yearly overview of newly included and inactivated disorders in the Orphanet**

517 **Nomenclature of RD (2020-2025).**

	Newly included disorders	Obsoleted disorders	Deprecated disorders	Non-rare disorders
2020	74	35	24	5

2021	53	8	16	1
2022	92	29	5	4
2023	114	21	8	5
2024	253	38	22	9
2025	152	13	19	2

518

519

520 ## Distribution of rare disorders by medical domain in the Orphanet Classification of RD

521 The Orphanet Classification of RD has a multi-parental structure: 4,415 (67.7%)
522 disorders are classified in more than one classification hierarchy while 2,106 (32.3%) are
523 classified in just one classification, independently of the number of parents that they have in
524 this same hierarchy. Figure 4 illustrates this multi-parental structure by comparing the number
525 of disorders linearized to each medical domain and therefore counted once (gray bars), with
526 the number of disorders classified in one or more medical domains and therefore counted as
527 many times as they appear in each classification (white bars). As an example, while 1,096
528 disorders are primarily classified as neurological disorders, a total of 2,492 disorders present
529 a major neurological involvement. Based on the linearization rules, the majority of rare
530 disorders fall into the rare developmental defect during embryogenesis classification (32.3%,
531 2,109 disorders), closely followed by the rare neurologic disorders classification (16.8%, 1,096
532 disorders), amounting together to almost 50% of all rare disorders. The following most-
533 represented classifications are the rare neoplastic disorders at 7.6%, the rare inborn errors of
534 metabolism at 6.1%, the rare skin disorders at 6%, the rare bone disorders at 5.4%, the rare
535 ophthalmic disorders at 3.7%, the rare immune disorders at 3.2%, the rare endocrine disorders

536 at 3%, the rare hematologic disorders at 2.8%, the rare infectious disorders at 2.7%, and the
537 rare systemic or rheumatologic diseases at 2.5%. The remaining classifications each include
538 fewer than 100 linearized disorders each, collectively representing the residual 8% of rare
539 disorders (Figure 4 gray bars; Additional file 3).

540 A similar pattern is observed when considering multiclassified disorders, where the
541 largest numbers are again found in the rare developmental defect during embryogenesis
542 classification (2,793 disorders) and the rare neurologic disorders classification (2,492
543 disorders). The following classifications, however, contain considerably less multiclassified
544 disorders and are, in descending order: rare ophthalmic disorders (868 disorders); rare skin
545 disorders (835 disorders); rare bone disorders (793 disorders); rare neoplastic disorders (743
546 disorders); and rare inborn errors of metabolism (704 disorders). All remaining hierarchies
547 contain less than 500 multiclassified disorders each, with only 3 hierarchies containing less
548 than 100 (Figure 4 white bars; Additional file 3).

549

550

551

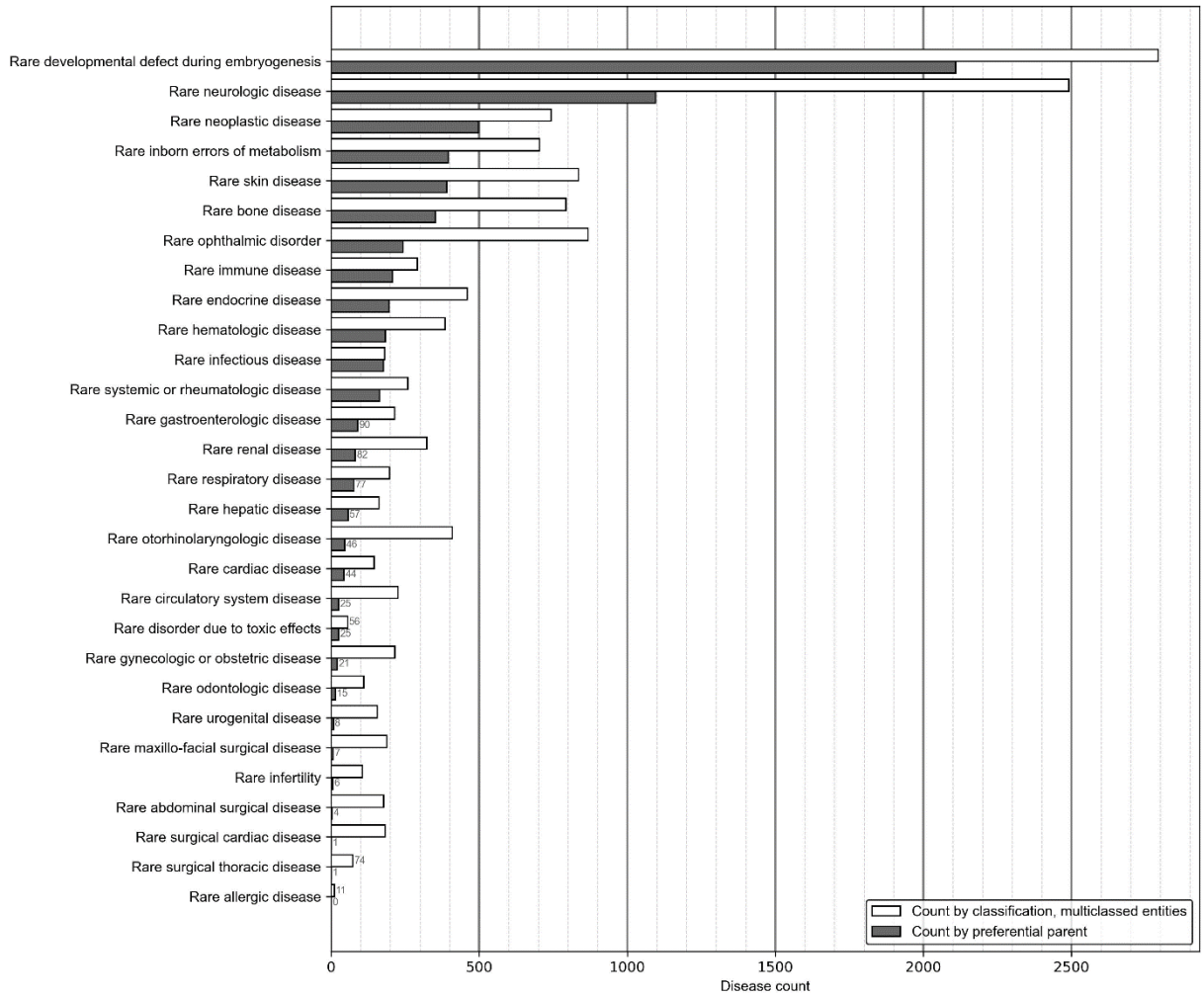


Figure 4. Distribution of rare disorders by medical domain according to their classification. Grey bars represent the number of disorders linearized according to their preferential parent classification, while white bars indicate the distribution of multi-classified disorders, counted as many times as they appear in the relevant classification hierarchies. Classifications containing fewer than 100 disorders are annotated with exact counts to enhance readability. The analysis includes 29 classification hierarchies corresponding to distinct medical domains; 5 hierarchies representing organizational or non-clinical groupings were excluded, as described in the Methods section. Six disorders were excluded from this analysis because they correspond to biological anomalies without clinical manifestations or were linearized under the classifications of rare genetic disorders or rare transplant-related disorders.

563 ORPHA:616874 Rare disorder without a determined diagnosis after full investigation was also
564 excluded.
565
566

567 Coverage and proximity of the mappings between the Orphanet Nomenclature of 568 RD and other semantic resources

569 Of the 6,527 active disorders registered in Orphanet, 6,355 (97.4%) are aligned to at
570 least one code ICD-10, with only 415 (6.4% of all disorders) having an exact proximity
571 relationship. The remaining aligned codes are either present in the ICD-10 alphabetical
572 index/are included in the tabulated list under the ICD-10 code (655 disorders) or they have
573 been attributed by Orphanet to the ICD-10 code that corresponds to the closest entity
574 according to Orphanet's rules (Table 5). The work on ICD-11 mapping has not been completed
575 yet, however, as of July 2025, 4,683 disorders (71.8%) were mapped to at least one code in
576 the ICD-11 MMS, with 958 (14.7% of all disorders) characterized by an exact proximity
577 relationship and 3,013 (46.2%) present in the ICD11 Foundation with an exact uri. The
578 remaining mappings have been attributed by Orphanet to the ICD-11 code corresponding to
579 the closest entity (Table 5).

580 According to the October 2025 release of the Orphanet-SNOMED CT mapping, 6,191
581 active Orphanet disorders (94.8%) are mapped to one SNOMED CT code, all with an exact
582 proximity relationship (Table 5).

583 Among all active disorders present in the Orphanet Nomenclature, 72.2% (n= 4,715)
584 are genetic disorders. A total of 4,141 Orphanet disorders are mapped to at least one
585 phenotypic OMIM number and 3,319 mapped disorders have an exact proximity relationship
586 with the corresponding phenotypic OMIM number (Table 5).

587 In addition, exact mappings to external semantic resources are distributed as follows:
588 6,501 (99.6%) for the UMLS Metathesaurus; 2,735 (41.9%) for MeSH; 1,455 (22.3%) for
589 MedDRA; 3,279 (50.2%) for GARD; and 6,194 (94.9%) for Mondo. Of note, only mappings to
590 ICD-10, ICD-11 and OMIM are distributed together with the official Orphanet Nomenclature
591 files for coding, while the entirety of the mappings are distributed through Orphadata Science.
592
593
594 **Table 5. Number of active disorders mapped to other semantic resources**

	Number of active disorders with at least one mapping	% on total disorders (n=6,527)	Number of active disorders with an exact proximity relationship	% on total disorders (n=6,527)
ICD-10	6,355*	97.4%	415	6.4%
ICD-11	4,683*	71.8%	958	14.7%
SNOMED CT	6,191	94.8%	6,191	94.8%
OMIM	4,141	63.4%	3,319	50.9%
UMLS	6,501	99.6%	6,501	99.6%
MeSH	2,735	41.9%	2,735	41.9%
MedDRA	1,455	22.3%	1,455	22.3%
GARD	3,279	50.2%	3,279	50.2%
Mondo	6,194	94.9%	6,194	94.9%

595 * Counts include mappings attributed by Orphanet when the rare disorder concept was not
596 present in the target resource.
597

598

599 # Discussion

600 ## Understanding the necessity of a RD-specific terminology

601 While collectively common, RD are numerous, clinically heterogeneous, and share
602 challenges directly related to their individual rarity: persons living with each RD are few and
603 dispersed; knowledge about each RD is limited and often inaccessible to non-specialist health
604 professionals; and expertise is scarce and dispersed as well. RD are hard to identify in medical
605 records since most are absent from medical terminologies in use in hospitals or not labeled as
606 rare, rendering them invisible in health systems. This lack of identification results in
607 inequalities in accessing a timely diagnosis and appropriate care, hinders data collection on
608 RD natural history and their individual and societal burdens, and complicates patient
609 recruitment for clinical research and therapy development [4,8,17,48].

610 The availability of adequate codes for each RD is a well-recognized need at the
611 international level [25]. Improved codification for RD has been cited as a priority in the 2009
612 EU Council Recommendation on an action in the field of rare diseases [27], and more recently
613 by the Rare2030 recommendations [49]. Having codes for each RD would help international,
614 European, national, and local health authorities obtain better knowledge of healthcare
615 pathways and of the impact of RD on both specialized (such as centers of expertise) and
616 general health care services, while also supporting budget planning for health and social
617 services [8,25,27]. Adequate coding is also necessary for identifying patients for clinical trials
618 at both national and transnational levels. Finally, a large proportion of RD patients remain
619 undiagnosed despite undergoing a full investigation, but this particular situation cannot be
620 captured by the codification systems in use.

621 RD codification requires a nomenclature and classification system that meets the
622 requirements of a standard medical terminology to ensure implementation in health
623 information systems and in systematic research collections, such as patient registries. A
624 trustworthy and reliable medical terminology must meet strict quality requirements: univocity
625 and unambiguity of coded concepts, unrestricted management of synonyms, consistency of
626 semantic relationships between concepts, multilingual support, accuracy of translations,
627 correctness of semantic alignments, and relevant coverage of identified use cases.
628 Additionally, terminology standards must adhere to rigorous and essential quality procedures,
629 ensure sustainability and scalability to support the rapid development of knowledge, and
630 guarantee long-term stability, preservation of the meaning of concepts from previous
631 versions, and traceability of changes [50].

632 A key prerequisite for a RD terminology is to adopt a clear and operational definition
633 of RD. Orphanet adopts a clinical definition that encompasses all RD, regardless of whether
634 the etiology is known or not. This core concept ensures a terminology that can be adapted to
635 different use cases, i.e. health systems with diverse diagnostic capabilities or different stages
636 of the patient's diagnostic journey. A RD in Orphanet is therefore defined as a clinical entity
637 characterized by a set of well-delineated phenotypic abnormalities, described in at least two
638 independent cases in the medical literature, and a consistent evolution, allowing for a
639 definitive clinical diagnosis. The rarity of a disease is defined according to the European
640 prevalence threshold (not more than 1 in 2,000 persons in the European population). This
641 choice was made for historical reasons, as Orphanet was developed in Europe and co-funded
642 by the European Commission (EC). However, this definition is consistent with those used in
643 other geographical areas, which adopt either a similar or more restrictive cutoff, with the
644 exception of the United States, where diseases with a higher prevalence are included [1,14].

645 Another limitation of this definition is that it includes infectious diseases which are rare in
646 Europe, even if they are common in other parts of the world. To address this, some countries
647 having adopted the Orphanet Nomenclature of RD outside the European space, like Argentina,
648 have adapted this chapter to exclude infectious diseases that are common in their countries
649 [51]. These limitations underline the need for consensus around a global operational
650 definition of RD, a goal the WHO has worked towards in collaboration with RDI, with the
651 participation of a multi-stakeholder, international group of experts including an Orphanet
652 representative (A.R.) [16]. The results of this work support Orphanet's adoption of a clinical
653 definition of RD and may lead in the future to adaptations of the Orphanet RD lists for the
654 different WHO regions. In the meantime, Orphanet collaborates with non-European countries
655 to help adapt the Orphanet Nomenclature of RD to local needs, improving global RD
656 interoperability.
657

658 ## Implications of the Orphanet terminology structure, mappings and updates

659 The Orphanet Nomenclature and Classification of RD is, to the best of our knowledge,
660 the only RD terminology clearly defining three levels of granularity: group of disorders,
661 disorder and subtype of disorders, with the disorder level complying with the internationally
662 agreed-upon definition of RD. This distinction has essential implications.

663 A first implication is that it allows to unambiguously count the number of existing RD
664 (i.e. the disorder classification level), preventing overcounting due to the aggregation of
665 heterogeneous granularity levels. As of today, many enumerations of RD circulate, some
666 derived from semi-automated integration of different databases and ontologies (most
667 frequently Orphanet and OMIM), sometimes resulting in estimates exceeding 10,000 RD
668 [8,52,53]. However, these methods aggregate different levels of granularity without clearly

669 stating which RD definition is adopted, rather reflecting the growing knowledge of disease
670 mechanisms or treatable genetic variants than an actual increase in the number of RD. While
671 this overall figure aligns closely with Orphanet's total active clinical entities (9,784 as per July
672 2025 release), it is essential to recognize that superimposing different levels of granularity
673 and/or different etiologies of the same RD can lead to counting the same disease multiple
674 times. Using this broader count in public health discussions without clarification can lead to
675 misinterpretations, potentially affecting policy decisions, budget allocation, and healthcare
676 planning. For epidemiological and public health purposes, the count of clinically defined RD
677 (6,527 according to the Orphanet July 2025 release) provides an appropriate common
678 denominator for comparing countries and healthcare systems with varying technological and
679 diagnostic capacities. It also permits to follow the evolution of therapy development and of
680 healthcare resource allocation, including assessment of how many RD benefit from these
681 measures over time and across geographies. Conversely, for precision medicine or molecular
682 diagnosis, a more granular approach including subtypes (1,084 in the July 2025 release) or
683 information on genetic variants collected in patient records may be needed.

684 A second implication is that defining the disorder granularity level allows the
685 prevalence of RD to be established accurately and facilitates comparability across countries
686 and continents. A clear RD definition therefore results in accuracy in counting RD and,
687 consequently, in counting patients, which is essential for assessing the magnitude of RD as a
688 public health issue. Furthermore, while allowing patient coding at the lowest granularity level
689 (i.e. Subtypes of disorder), the classification hierarchies also allow data aggregation at the
690 disorder level, thereby permitting to unambiguously establish the prevalence of RD. Notably,
691 the Orphanet Nomenclature of RD is the only system which provides a code for undiagnosed
692 RD patients who underwent all available diagnostic possibilities. This RD population has been

693 estimated to represent up to 50% of all RD patients, based on the experience of RD expert
694 centers [33]. Being able to code these patients at the disorder level makes it possible to
695 include them in research pathways and undiagnosed programs, as recommended by IRDiRC's
696 goal #1 [54–56], and to include them in future global RD point prevalence estimations.

697 The three-level classification structure of the Orphanet Nomenclature and
698 Classification of RD supports consistent semantic relationships between concepts, enabling
699 effective use across various use cases and settings. While the disorder level serves as the
700 primary reference level for data sharing among systems and countries, and for
701 epidemiological purposes, groups of disorders allow data aggregation for use cases analysis,
702 i.e. for public health decision-making such as healthcare organization (for example,
703 designation of centers of expertise by RD area). Meanwhile, subtypes of disorders allow for
704 fine-tuning of RD codification at the clinical, etiological and histopathological level, and can be
705 used for applications in genetics, precision medicine, and other research and innovation
706 purposes [57–63].

707 The Orphanet Classification system is also able to capture the multisystemic nature of
708 RD, as clinical entities are organized into 29 distinct medical domains and classified as many
709 times as needed according to their relevant domain of care. This is reflected in the distribution
710 of RD across classification hierarchies: the majority of RD fall within the rare developmental
711 defect during embryogenesis and rare neurologic disease classifications, together accounting
712 for nearly 50% of all RD. However, almost 70% of RD are classified in more than one
713 classification, reflecting their multisystem involvement. This approach ensures a more
714 accurate representation of the complex nature of RD and of their linked resources, allowing
715 to better orient patients and medical doctors towards appropriate information and specialized
716 care [57–59]. At the same time, the distribution of entities across preferential parent

717 classifications is uneven, with some domains encompassing a large proportion of RD and
718 others including comparatively fewer entities. This distribution has implications for statistical
719 analysis and data representation, as domains with fewer entities may appear less represented
720 when preferential parent classifications are used for data aggregation. However, the multi-
721 hierarchical structure allows flexible aggregation strategies that can be tailored to specific
722 epidemiological or public health questions.

723 The field of RD is a rapidly evolving one: with the continuous advancement in genetic
724 knowledge and the growing interest in personalized medicine, new RD are described each
725 month. Additionally, the scope of previously identified RD can be refined, with medical
726 concepts being either merged or divided based on new discoveries or the examination of
727 larger numbers of patients. To ensure that the Orphanet Nomenclature and Classification of
728 RD remain up to date with current scientific knowledge, Orphanet conducts a monthly
729 systematized review of the literature and continuously collaborates with medical experts,
730 particularly from the European Reference Networks (ERNs) and international RD learned
731 societies [32,64]. Decisions on modifications to ORPHAcodes are made after an extensive
732 review of the literature, and are validated by the Orphanet medical and scientific committee,
733 following published standardized procedures [29]. Between 2020 and 2025, a total of 738 new
734 RD were included in the Orphanet Nomenclature of RD, averaging about 120 RD per year. This
735 number reflects not only newly described RD, but also the outcomes of numerous time-
736 effective collaborative projects with experts, aimed at restructuring and updating the
737 classification to incorporate revised medical concepts. At the same time, Orphanet also
738 applies a strict inactivation policy to preserve the stability and univocity of ORPHAcodes.
739 Inactivated codes are never reused for other entities, and a substitution code is always
740 suggested to maintain traceability. While unusual, it is possible to reintroduce an inactivated

741 ORPHAcode when new evidence emerges that justifies a reconsideration [29]. This has
742 occurred in cases such as Systemic lupus erythematosus ORPHA:536 and Immunoglobulin A
743 nephropathy ORPHA:34145, which did not initially meet RD prevalence criteria and are now
744 considered RD. Taken together, the integration of new RD and the number of inactivated ones
745 contribute to a small and relatively stable annual increase in the total number of RD, averaging
746 fewer than 80 RD per year. A small proportion of active clinical entities are identified as
747 historic, corresponding to RD for which no new patients have been reported in the medical
748 literature for 25 years. These entities are retained in the nomenclature to preserve diagnoses
749 that could potentially re-emerge and to support retrospective analyses, while their specific
750 status limits the risk of confusion.

751 To further ensure the univocity and accessibility of concepts, Orphanet implements an
752 unrestricted management of synonyms. Synonyms must cover the exact scope of the concept
753 and are added as needed, without limitations in number [29]. This approach maximizes
754 semantic coverage across different usages and data sources. Additionally, Orphanet provides
755 a standardized, medically validated definition for each RD included in the Orphanet
756 Nomenclature. These definitions consist of short texts describing the clinical features to
757 unambiguously delineate the entity and distinguish it from other RD, and are intended to
758 support healthcare professionals in assigning the correct ORPHAcode to a patient [29,65]. To
759 further ensure international interoperability, the Orphanet Nomenclature and Classification
760 of RD is also a multilingual resource. Translations are performed according to established
761 procedures that include a final manual medical validation [31]. The different translations also
762 include an unlimited number of synonyms that are tailored according to the relevance for each
763 language. Taken together, these elements ensure semantic consistency and facilitate
764 ORPHAcode integration into health and research information systems.

765 RD is a field in which the boundaries between diagnosis, healthcare, and research are
766 often blurred, therefore maintaining semantic interoperability between health and research
767 sectors is essential. The multiplicity of reference semantic standards used in healthcare,
768 research and regulatory science reflects the fact that each serves a distinct purpose, adopting
769 diverse levels of granularity, and encoding clinical concepts according to its own organizing
770 principles. For example, ICD classifications are optimized for morbidity and mortality statistics,
771 billing and reimbursement, and administrative reporting, which often results in broad disease
772 groupings. In contrast, research-oriented resources such as OMIM organize information
773 around gene-phenotype relationships, frequently splitting a single clinically defined disease
774 into multiple molecular entries or, on the contrary, merging distinct clinical disorders related
775 to the same gene. Consequently, semantic alignments between terminologies entail inherent
776 complexity and need domain expertise, and must not be represented as simple one-to-one
777 correspondences, but rather as structured relationships, as described in studies investigating
778 RD terminology harmonization and mapping processes [66]. Mapping strategies must
779 therefore explicitly document mapping cardinality and capture varying degrees of conceptual
780 proximity and specificity, in order to preserve meaning across contexts and avoid
781 misinterpretation. In addition, while automated approaches may assist in identifying
782 candidate correspondences, the validation of RD mappings requires expert medical review to
783 ensure that assigned relationships accurately reflect clinical meaning. This is reflected in the
784 way ORPHAcodes are mapped to different resources, with defined relationships and medical
785 validation. As of July 2025, Orphanet has aligned 97.4% of its active rare disorders to at least
786 one ICD-10 code, with only 6.4% of all RD being exactly represented. The work on ICD-11
787 mapping is ongoing, but by July 2025, 71.8% of Orphanet's RD were mapped to at least one
788 ICD-11 MMS code, with 14.7% all RD being exactly represented. Additionally, the October

789 2025 release of the Orphanet-SNOMED CT mapping shows that 94.8% of Orphanet's active RD
790 are mapped to SNOMED CT, with all exhibiting an exact proximity relationship. Regarding
791 genetic diseases, Orphanet classifies 4,715 RD (72.2% of active disorders) as genetic (defined
792 as either caused by a known gene or having a clear familial inheritance pattern), with 4,141
793 RD mapped to one or more phenotypic OMIM number. It is interesting to note that many
794 genetic RD are mapped to multiple OMIM numbers, reflecting the fact that Orphanet adopts
795 a clinical definition of disease, whereas OMIM organizes entries around individual gene-
796 phenotype relationships.

797 Orphanet is currently the only primary producer of a scientifically validated
798 nomenclature and classification of RD. Other terminologies are increasingly integrating
799 Orphanet concepts as a source of RD information, acknowledging the critical scientific
800 contribution Orphanet makes to the RD community. Outstanding collaborations between
801 Orphanet, SNOMED International and WHO ensure that the RD content of SNOMED CT and
802 the ICD-11 continue to be improved by the inclusion of RD concepts as defined by Orphanet
803 [23,24].

804 In parallel with mappings to major reference standards, the alignment of ORPHAcodes
805 with other semantic resources, including UMLS, MeSH, MedDRA, GARD, and Mondo, supports
806 a range of use cases such as literature indexing, pharmacovigilance, regulatory reporting, and
807 knowledge integration. Overall, this extensive and accurate mapping system ensures
808 consistency of transcoding across multiple platforms, further enhancing the utility of the
809 Orphanet Nomenclature of RD in clinical, research, and regulatory contexts [60–63,67,68].
810

Recognition and adoption of the Orphanet Nomenclature of RD

Since 1997, Orphanet has maintained through its publicly available website an inventory of RD to provide information on RD and their dedicated services across Europe [28]. This has evolved over time into a nomenclature and classification system of RD, that is today the only RD-specific standard meeting the high-quality requirements of a medical terminology [20].

In 2014, the EC Expert Group on Rare Diseases adopted a recommendation on improving RD codification [25]. In this document, EU Member States are encouraged to assess the feasibility of implementing ORPHAcodes at a national level and to include RD codification into their national RD plans/strategies. In 2017, the EC Steering Group on Promotion and Prevention (SGPP) recognized ORPHAcodes as a best practice [69]. Since then, ORPHAcodes have been included in the set of common data elements for Rare Diseases Registration, released by the EU RD platform on RD registration to enhance interoperability between registries [70]. The Orphanet Nomenclature of RD has also played a key role in implementing the EU 2011 Cross-Border Healthcare Directive in regards to RD. As a result, ORPHAcodes have been recommended for tracking RD diagnoses in the European Common Semantic Strategy (2019) [71] and the Guidelines on Patient Summary (Release 3.2, March 2022 and Release 3.4, November 2024) [72], and have been integrated into the Code Systems used in the MVC - My Health @ EU - eHealth Digital Service Infrastructure (eHDSI) [73,74]. The functional specifications developed by the X-eHealth project were instrumental for including ORPHAcodes in the minimum dataset required for the electronic Patient Summary [72,75]. Finally, the World Health Assembly (WHA) resolution “Rare diseases: a global health priority for equity and inclusion,” adopted on the 24th of May 2025, urges Member States to commit to “considering the implementation of ICD-11, and where appropriate, interoperable

835 codification systems for RD such as the Orphanet Nomenclature of rare diseases, at their
836 earliest possibility, and in accordance with their available resources, in order to enable the
837 recording, reporting and monitoring of rare diseases at the national and international levels"
838 [76].

839 To support the effective and sustainable implementation in health and research
840 information systems, the Orphanet Nomenclature of RD (ORPHAcodes) is distributed under
841 an open license (CC-BY 4.0) on different platforms depending on use: Orphadata Science for
842 structured datasets intended for large scale analyses [46], and the Orphacode.org platform
843 for coding resources for health information systems [47]. Each annual release (in July of each
844 year) is versioned and accompanied by differential files to ensure traceability of changes and
845 continuity of coding, as per the recommendation of the RD-ACTION working group for routine
846 maintenance of codification [77]. In addition, a suite of technical tools is provided to facilitate
847 implementation into diverse systems, promote accurate patient codification, and ensure
848 continuous support by the community. Finally, the implementation of ORPHAcodes within
849 information systems should be guided by established and evolving guidelines and best-
850 practice recommendations, including those developed within the RD-CODE and the
851 OD4RD/OD4RD2 project [33,78–80]. This supporting infrastructure ecosystem is essential to
852 ensure evolution over time and respond to emerging use cases in a rapidly evolving field.

853 Establishing a seamless data ecosystem for RD based on the Orphanet Nomenclature
854 and Classification of RD is a key step towards increasing the knowledge in a field that has long
855 remained largely invisible within health systems, and ultimately improving the lives of PLWRD
856 [49]. Indeed, accurate codification of RD patients enables the production of data for both
857 primary and secondary use. Primary use includes, among others: recognition of the RD
858 condition by non-expert health professionals; ensuring best care practices at all stages of the

859 patient's care pathway; and recognition of undiagnosed RD patients who might benefit from
860 new discoveries, potentially leading to establishing a diagnosis. Secondary use of RD data
861 includes accurately measuring the prevalence of the entire RD population and monitoring
862 public health interventions to inform evidence-based decision-making; producing real-world
863 evidence for research and therapeutic development; and facilitating recruitment for clinical
864 research.
865

866 ## Conclusions

867 The Orphanet Nomenclature and Classification of RD is the only RD-specific, high-
868 quality standard, interoperable medical terminology meeting the needs of healthcare,
869 research, and public health systems. By providing a clinically based and scientifically validated
870 framework, it helps close the RD visibility gap in health and research information systems,
871 enabling consistent RD identification and monitoring. Its wide adoption will support cross-
872 border and cross-sectoral data interoperability, contributing to improved knowledge, policy-
873 making, and ultimately better care for people living with a rare disease.
874
875

876 # Acknowledgements

877 We would like to thank the Orphanet translating members for their essential
878 contribution to the translation of the Orphanet Nomenclature of RD, as well as the European
879 Reference Networks (ERNs) and the Filières de santé maladies rares (FSMR) for their continued
880 involvement in nomenclature and classification revisions, and in the production of textual
881 content.

Funding

The Orphanet Nomenclature and Classification system has been developed and maintained thanks to Inserm support since 2007 as well as European support since 2009 (*RD-PORTAL S12305098-2006119; RD-PORTAL2 S12324970-20091215; ORPHANET EUROPE JOINT ACTION 20102206 ; EUCERD Joint Action 2011 22 01; ORPHANET OPERATING GRANT 20133305; RD-ACTION JOINT ACTION 677024; ORPHANETWORK DIRECT GRANT 831390; RD-CODE 826607; OD4RD 101070531; OD4RD2 101110100*). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Ethics statement

This study did not involve human participants, patient-level data, or identifiable personal information. The analyses were conducted exclusively using data derived from published literature and publicly available sources curated within Orphanet. Consequently, approval from an Institutional Review Board (IRB) or equivalent ethics committee was not required, and informed consent was not applicable.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Data availability

The datasets supporting the conclusions of this article are publicly available. Orphanet provides open access to rare disease nomenclature, classification, genetic data, and mappings to other terminologies (https://www.orpha.net). Structured datasets for large-scale analysis are distributed via the Orphadata Science platform (https://sciences.orphadata.com/) and all previous versions are archived in a dedicated GitHub repository (https://github.com/Orphanet/Orphadata_aggregated). Additional resources include the Orphanet Nomenclature files for coding (Nomenclature Pack) distributed on the ORPHAcodes platform (https://www.orphacode.org/) and the Orphanet Rare Disease Ontology (ORDO, https://www.orphadata.com/ordo/). All formats are maintained and updated annually or bi-annually to ensure accuracy and accessibility.

Authors' contributions

AR, CL, AO, and HA conceived the study and contributed to the overall research goals and aims. CL, MDC, ADS, MF, EG, MCG, SM, MT, FS, VL, and DL curated and maintained the data. CL and DL carried out the formal analyses. AR, MH, and SM secured the funding. Methodology was developed by AR, CL, AO, HA, and VL. AR and CL coordinated and

924 administered the project. DL developed the software. AR and MH provided supervision.
925 Validation was performed by CL, DL, FL, and VL. CL and DL prepared the visualizations. CL, AR,
926 AO, and HA drafted the initial manuscript. CL, AR, SM, CR, VL, and MH critically revised and
927 edited the manuscript. All authors read and approved the final version of the manuscript.
928
929

930 ## Abbreviations:

931 BTNT: Broader Term to Narrow Term
932 CS: Czech
933 DE: German
934 EC: European Commission
935 EN: English
936 ERN: European Reference Network
937 ES: Spanish
938 EU: European Union
939 FR: French
940 FSMR: Filière de santé maladies rares
941 GARD: Genetic and Rare Diseases Information Center
942 ICD: International Classification of Diseases
943 IT: Italian
944 MeSH: Medical Subject Headings
945 MedDRA: Medical Dictionary for Regulatory Activities
946 MMS: Mortality and Morbidity Statistics

Mondo: Mondo Disease Ontology

NL: Dutch

NTBT: Narrow Term to Broader Term

OMIM: Online Mendelian Inheritance in Man

ORDO: Orphanet Rare Disease Ontology

PL: Polish

PLWRD: People Living with a Rare Disease

PT: Portuguese

RD: Rare Diseases

RDI: Rare Diseases International

SGPP: Steering Group on Promotion and Prevention

SNOMED CT: Systematized Nomenclature of Medicine Clinical Terms

UMLS: Unified Medical Language System

WHO: World Health Organization

Multimedia Appendix

Additional file 1 (Excel file .xlsx)

Distribution of Orphanet clinical entities by classification level and typology. This file is the output of Notebook 1_number_of_Orphanet_clinical_entities.

Additional file 2 (Excel file .xlsx)

969 **Production of the Orphanet RD terminology in English and translation languages.** This file is
970 the output of Notebook
971 2_production_of_the_Orphanet_RD_terminology_in_english_and_other_languages.
972
973 **Additional file 3 (Excel file .xlsx)**
974 **Distribution of disorders in Orphanet Classifications according to their preferential parent**
975 **and multiclassified disorders.** This file combines the outputs of Notebook
976 5_distribution_of_Orphanet_RD_in_Orphanet_classifications and Notebook
977 6_distribution_of_Orphanet_RD_by_preferential_parent.
978
979 **Additional file 4 (PDF file)**
980 **Glossary of key concepts and definitions used in the Orphanet Nomenclature and**
981 **Classification of Rare Diseases.**
982
983 **Additional file 5 (Word file .docx)**
984 **Overview of the 34 Orphanet classification hierarchies and their status as distinct medical**
985 **domains or organizational classifications**
986
987

988 ## References

989 1. Nguengang Wakap S, Lambert DM, Olry A, Rodwell C, Gueydan C, Lanneau V, Murphy
990 D, Le Cam Y, Rath A. Estimating cumulative point prevalence of rare diseases: analysis
991 of the Orphanet database. Eur J Hum Genet 2020 Feb;28(2):165–173. PMID:31527858

2. Orphanet, Inserm. Rare diseases in numbers. 2026 Feb. Report No.: 1. Available from: https://www.orpha.net/pdfs/orphacom/cahiers/docs/EN/OrphanetRDFactsheet.pdf [accessed Feb 24, 2026]

3. European Parliament: Directorate-General for Parliamentary Research Services, Schmiegelow K, Fehn K, Westman D, Clemente Pesudo I, Atanasovska D, Vucina Pedersen N. Addressing challenges to European multi-country collaboration models for rare diseases. European Parliament; 2024. Available from: https://data.europa.eu/doi/10.2861/51755 [accessed June 15, 2025]

4. Ferreira CR. The burden of rare diseases. Am J Med Genet A 2019 June;179(6):885–892. PMID:30883013

5. European Parliament: Directorate-General for Internal Policies of the Union, Kamphuis B, De Jongh T, Bastiaanssen V. Tackling rare diseases – Challenges, opportunities and gaps for action on rare diseases in the European Union. European Parliament; 2024. Available from: https://data.europa.eu/doi/10.2861/263904 [accessed June 15, 2025]

6. Richter T, Nestler-Parr S, Babela R, Khan ZM, Tesoro T, Molsen E, Hughes DA, International Society for Pharmacoeconomics and Outcomes Research Rare Disease Special Interest Group. Rare Disease Terminology and Definitions-A Systematic Global Review: Report of the ISPOR Rare Disease Special Interest Group. Value Health J Int Soc Pharmacoeconomics Outcomes Res 2015 Sept;18(6):906–914. PMID:26409619

7. Khosla N, Valdez R. A compilation of national plans, policies and government actions for rare diseases in 23 countries. Intractable Rare Dis Res 2018 Nov 30;7(4):213–222. PMID:30560012

1014 8. Haendel M, Vasilevsky N, Unni D, Bologa C, Harris N, Rehm H, Hamosh A, Baynam G,
1015 Groza T, McMurry J, Dawkins H, Rath A, Thaxon C, Bocci G, Joachimiak MP, Köhler S,
1016 Robinson PN, Mungall C, Oprea TI. How many rare diseases are there? Nat Rev Drug
1017 Discov 2020 Feb;19(2):77–78. PMID:32020066

1018 9. Chan AYL, Chan VKY, Olsson S, Fan M, Jit M, Gong M, Zhang S, Ge M, Pathadka S, Chung
1019 CCY, Chung BHY, Chui CSL, Chan EW, Wong GHY, Lum TY, Wong ICK, Ip P, Li X. Access
1020 and Unmet Needs of Orphan Drugs in 194 Countries and 6 Areas: A Global Policy
1021 Review With Content Analysis. Value Health Elsevier; 2020 Dec 1;23(12):1580–1591.
1022 PMID:33248513

1023 10. Wainstock D, Katz A. Advancing rare disease policy in Latin America: a call to action.
1024 Lancet Reg Health Am 2023 Feb;18:100434. PMID:36844013

1025 11. Abozaid GM, Kerr K, McKnight A, Al-Omar HA. Criteria to define rare diseases and
1026 orphan drugs: a systematic review protocol. BMJ Open 2022 July 29;12(7):e062126.
1027 PMID:35906057

1028 12. Hayashi S, Umeda T. 35 years of Japanese policy on rare diseases. Lancet Lond Engl
1029 2008 Sept 13;372(9642):889–890. PMID:18790305

1030 13. European Parliament and Council of the European Union. Regulation (EC) No 141/2000
1031 of the European Parliament and of the Council of 16 December 1999 on orphan
1032 medicinal products. Publications Office of the European Union; 2000 Jan. Available
1033 from: https://eur-lex.europa.eu/eli/reg/2000/141/oj/eng [accessed June 15, 2025]

1034 14. United States Congress. Orphan Drug Act, 21 U.S.C. § 360bb(a)(2). United States
1035 Government Publishing Office; 2023. Available from:
1036 https://www.govinfo.gov/app/details/USCODE-2023-title21/USCODE-2023-title21-
1037 chap9-subchapV-partB-sec360bb [accessed Aug 3, 2025]

1038 15. Abozaid GM, Kerr K, Alomary H, Al-Omar HA, McKnight A. Global insight into rare
1039 disease and orphan drug definitions: a systematic literature review. BMJ Open British
1040 Medical Journal Publishing Group; 2025 Jan 1;15(1):e086527. PMID:39863413

1041 16. Wang CM, Whiting AH, Rath A, Anido R, Ardigò D, Baynam G, Dawkins H, Hamosh A, Le
1042 Cam Y, Malherbe H, Molster CM, Monaco L, Padilla CD, Pariser AR, Robinson PN,
1043 Rodwell C, Schaefer F, Weber S, Macchia F. Operational description of rare diseases: a
1044 reference to improve the recognition and visibility of rare diseases. Orphanet J Rare Dis
1045 2024 Sept 11;19(1):334. PMID:39261914

1046 17. Valdez R, Ouyang L, Bolen J. Public Health and Rare Diseases: Oxymoron No More. Prev
1047 Chronic Dis 2016 Jan 14;13:E05. PMID:26766846

1048 18. United Nations General Assembly. Addressing the challenges of persons living with a
1049 rare disease and their families: resolution. United Nations; 2021. Available from:
1050 http://digitallibrary.un.org/record/3953832 [accessed June 15, 2025]

1051 19. Jetté N, Quan H, Hemmelgarn B, Drosler S, Maass C, Moskal L, Paoin W, Sundararajan
1052 V, Gao S, Jakob R, Ustün B, Ghali WA, IMECCHI Investigators. The development,
1053 evolution, and modifications of ICD-10: challenges to the international comparability of
1054 morbidity data. Med Care 2010 Dec;48(12):1105–1110. PMID:20978452

1055 20. Rath A, Olry A, Dhombres F, Brandt MM, Urbero B, Ayme S. Representation of rare
1056 diseases in health information systems: The orphanet approach to serve a wide range
1057 of end users. Hum Mutat 2012;33(5):803–808. PMID:22422702

1058 21. Aymé S, Bellet B, Rath A. Rare diseases in ICD11: making rare diseases visible in health
1059 information systems through appropriate coding. Orphanet J Rare Dis 2015 Mar
1060 26;10:35. PMID:25887186

1061 22. Aymé S, Rath A, Bellet B. WHO International Classification of Diseases (ICD) Revision
1062 Process: incorporating rare diseases into the classification scheme: state of art.
1063 Orphanet J Rare Dis 2010 Oct 19;5(Suppl 1):P1. doi: 10.1186/1750-1172-5-S1-P1

1064 23. World Health Organization. Rare diseases (frequently asked questions). WHO – Int
1065 Classif Dis. Available from: https://www.who.int/standards/classifications/frequently-
1066 asked-questions/rare-diseases [accessed June 15, 2025]

1067 24. SNOMED International. INSERM and SNOMED International release SNOMED CT to
1068 Orphanet map supporting representation and use. SNOMED Int News. 2021. Available
1069 from: https://www.snomed.org/news/inserm-and-snomed-international-release-
1070 snomed-ct-to-orphanet-map-supporting-representation-and-use [accessed June 15,
1071 2025]

1072 25. Commission Expert Group on Rare Diseases (CEGRD), European Commission.
1073 Recommendation on ways to improve codification for rare diseases in health
1074 information systems. 2014 Nov. Available from:
1075 https://health.ec.europa.eu/publications/recommendation-ways-improve-codification-
1076 rare-diseases-health-information-systems_en [accessed June 15, 2025]

1077 26. Choquet R, Maaroufi M, de Carrara A, Messiaen C, Luigi E, Landais P. A methodology for
1078 a minimum data set for rare diseases to support national centers of excellence for
1079 healthcare and research. J Am Med Inform Assoc JAMIA 2015 Jan;22(1):76–85.
1080 PMID:25038198

1081 27. Council of the European Union. Council Recommendation of 8 June 2009 on an action
1082 in the field of rare diseases. Publications Office of the European Union; 2009. Available
1083 from: https://eur-lex.europa.eu/legal-
1084 content/EN/TXT/?uri=celex%3A32009H0703%2802%29 [accessed June 15, 2025]

1085 28. Orphanet, Inserm. Orphanet – Knowledge on rare diseases and orphan drugs.
1086 Orphanet. Available from: http://www.orpha.net/ [accessed July 30, 2025]

1087 29. Orphanet, Inserm. Procedural document: Orphanet nomenclature and classification of
1088 rare diseases. 2023 June. Report No.: 05. Available from:
1089 https://www.orpha.net/pdfs/orphacom/cahiers/docs/GB/eproc_disease_inventory_R1
1090 _Nom_Dis_EP_05.pdf [accessed Aug 3, 2025]

1091 30. Orphanet, Inserm. Procedural document: Rare disease nomenclature in English. 2017
1092 Apr. Report No.: 1. Available from:
1093 https://www.orpha.net/pdfs/orphacom/cahiers/docs/GB/eproc_Disease_naming_rules
1094 _in_English_PR_R1_Nom_01.pdf [accessed Aug 3, 2025]

1095 31. Orphanet, Inserm. Procedural document: Orphanet Rare Diseases Nomenclature
1096 Production in National Language. 2017 Sept. Report No.: 1. Available from:
1097 https://www.orpha.net/pdfs/orphacom/cahiers/docs/GB/eproc_Rare_disease_Nomen
1098 clature_Production__national-language.pdf [accessed Aug 3, 2025]

32. Orphanet, Inserm. Procedural document: Collaboration with networks of expertise for the revision of the Orphanet nomenclature and classification of rare diseases. 2024 Sept. Report No.: 1.1. Available from: https://www.orpha.net/pdfs/orphacom/cahiers/docs/GB/eproc_Collaboration_networks_R1_Nom_Rev_Exp_EP_10.pdf [accessed Aug 3, 2025]

33. Angin C, Mazzucato M, Weber S, Kirch K, Abdel Khalek W, Ali H, Maiella S, Olry A, Jannot A-S, Rath A. Coding undiagnosed rare disease patients in health information systems: recommendations from the RD-CODE project. Orphanet J Rare Dis 2024 Jan 27;19:28. PMID:38280999

34. Orphanet, Inserm. Procedural document: Linearisation rules for rare diseases. 2014 Sept. Available from: https://www.orpha.net/pdfs/orphacom/cahiers/docs/GB/Orphanet_linearisation_rules.pdf [accessed Aug 3, 2025]

35. Orphanet, Inserm. Procedural document: Orphanet ICD-10 Coding Rules for Rare Diseases. 2023 June. Report No.: 3. Available from: https://www.orpha.net/pdfs/orphacom/cahiers/docs/GB/Orphanet_ICD10_coding_rules_R1_Nom_ICD_EP_07.pdf [accessed Aug 3, 2025]

36. World Health Organization. International statistical classification of diseases and related health problems. 10th revision (ICD-10). 2010 edition. Geneva: World Health Organization; Available from: https://icd.who.int/browse10/2010/en ISBN:978-92-4-154834-2

1120 37. World Health Organization. ICD-11: International Classification of Diseases, 11th
1121 Revision. ICD-11 Mortal Morb Stat. Available from: https://icd.who.int/en [accessed
1122 June 15, 2025]

1123 38. U.S. National Library of Medicine. Overview of SNOMED CT. US Natl Libr Med – Health
1124 IT. U.S. National Library of Medicine; 2014. Available from:
1125 https://www.nlm.nih.gov/healthit/snomedct/snomed_overview.html [accessed July
1126 25, 2025]

1127 39. SNOMED International. What is SNOMED CT. SNOMED Int. Available from:
1128 https://www.snomed.org/what-is-snomed-ct [accessed July 25, 2025]

1129 40. Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. OMIM.org: Online
1130 Mendelian Inheritance in Man (OMIM®), an online catalog of human genes and genetic
1131 disorders. Nucleic Acids Res 2015 Jan 28;43(Database issue):D789–D798.
1132 PMID:25428349

1133 41. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical
1134 terminology. Nucleic Acids Res 2004 Jan 1;32(Database issue):D267–D270.
1135 PMID:14681409

1136 42. Lipscomb CE. Medical Subject Headings (MeSH). Bull Med Libr Assoc 2000
1137 July;88(3):265–266. PMID:10928714

1138 43. Brown EG, Wood L, Wood S. The Medical Dictionary for Regulatory Activities
1139 (MedDRA). Drug Saf 1999 Feb 1;20(2):109–117. PMID:10082069

1140 44. National Center for Advancing Translational Sciences (NCATS), National Institutes of
1141 Health. Genetic and Rare Diseases Information Center (GARD). Genet Rare Dis Inf Cent.
1142 Available from: https://rarediseases.info.nih.gov/ [accessed June 15, 2025]

1143 45. Vasilevsky NA, Matentzoglu NA, Toro S, Flack JE, Hegde H, Unni DR, Alyea GF, Amberger
1144 JS, Babb L, Balhoff JP, Bingaman TI, Burns GA, Buske OJ, Callahan TJ, Carmody LC, Cordo
1145 PC, Chan LE, Chang GS, Christiaens SL, Daugherty LC, Dumontier M, Failla LE, Flowers
1146 MJ, Garrett HA, Goldstein JL, Gration D, Groza T, Hanauer M, Harris NL, Hilton JA,
1147 Himmelstein DS, Hoyt CT, Kane MS, Köhler S, Lagorce D, Lai A, Larralde M, Lock A,
1148 Santiago IL, Maglott DR, Malheiro AJ, Meldal BHM, Munoz-Torres MC, Nelson TH,
1149 Nicholas FW, Ochoa D, Olson DP, Oprea TI, Osumi-Sutherland D, Parkinson H,
1150 Pendlington ZM, Rath A, Rehm HL, Remennik L, Riggs ER, Roncaglia P, Ross JE, Shadbolt
1151 MF, Shefchek KA, Similuk MN, Sioutos N, Smedley D, Sparks R, Stefancsik R, Stephan R,
1152 Storm AL, Stupp D, Stupp GS, Sundaramurthi JC, Tammen I, Tay D, Thaxton CL, Valasek
1153 E, Valls-Margarit J, Wagner AH, Welter D, Whetzel PL, Whiteman LL, Wood V, Xu CH,
1154 Zankl A, Zhang XA, Chute CG, Robinson PN, Mungall CJ, Hamosh A, Haendel MA.
1155 Mondo: Unifying diseases for the world, by the world. medRxiv; 2022. p.
1156 2022.04.13.22273750. doi: 10.1101/2022.04.13.22273750

1157 46. Orphanet, Inserm. Orphadata Science – Orphanet datasets. Sci Orphadata. 2025.
1158 Available from: https://sciences.orphadata.com/ [accessed July 30, 2025]

1159 47. Orphanet, Inserm. Orphanet Nomenclature Pack for Coding. ORPHAcodes. 2025.
1160 Available from: https://www.orphacode.org/pack-nomenclature/ [accessed July 25,
1161 2025]

48. Schieppati A, Henter J-I, Daina E, Aperia A. Why rare diseases are an important medical and social issue. The Lancet Elsevier; 2008 June 14;371(9629):2039–2041. PMID:18555915

49. Kole A, Hedley V, et al. Recommendations from the Rare 2030 Foresight Study: the future of rare diseases starts today. EURORDIS-Rare Diseases Europe (Rare2030 project); 2021 Feb. Available from: https://download2.eurordis.org/rare2030/Rare2030_recommendations.pdf [accessed Aug 3, 2025]

50. Direction générale de l'offre de soins, France numérique santé (e-Santé) – Ministry of Health, France. Référentiels d'intéropérabilité sémantique : mise en euvre de terminologies de référence pour le secteur santé-social en France. Ministère des Solidarités et de la Santé, France; 2015. Available from: https://esante.gouv.fr/sites/default/files/media_entity/documents/20150504_etude_terminos_phase2_diagnostic.pdf [accessed June 15, 2025]

51. Ministerio de Salud de la Nación, Argentina. Listado de enfermedades poco frecuentes en Argentina. Argent – Salud. Available from: https://www.argentina.gob.ar/salud/pocofrecuentes/listado [accessed June 15, 2025]

52. Lamoreaux K, Lefebvre, S., Levine, D., Erler, W., Hume, T. The Power of Being Counted: A more accurate count of rare diseases and steps to getting counted. RARE-X Data Platform (RARE-X); 2022 May. Available from: https://rare-x.org/wp-content/uploads/2022/05/be-counted-052722-WEB.pdf [accessed Aug 3, 2025]

1183 53. Smith CIE, Bergman P, Hagey DW. Estimating the number of diseases – the concept of
1184 rare, ultra-rare, and hyper-rare. iScience 2022 Aug 19;25(8):104698. PMID:35856030

1185 54. International Rare Diseases Research Consortium (IRDiRC). IRDiRC Vision & Goals.
1186 IRDiRC Irdircorg. Available from: https://irdirc.org/about-us/vision-goals/ [accessed
1187 June 15, 2025]

1188 55. Boycott KM, Rath A, Chong JX, Hartley T, Alkuraya FS, Baynam G, Brookes AJ, Brudno M,
1189 Carracedo A, den Dunnen JT, Dyke SOM, Estivill X, Goldblatt J, Gonthier C, Groft SC, Gut
1190 I, Hamosh A, Hieter P, Höhn S, Hurles ME, Kaufmann P, Knoppers BM, Krischer JP,
1191 Macek M, Matthijs G, Olry A, Parker S, Paschall J, Philippakis AA, Rehm HL, Robinson
1192 PN, Sham P-C, Stefanov R, Taruscio D, Unni D, Vanstone MR, Zhang F, Brunner H,
1193 Bamshad MJ, Lochmüller H. International Cooperation to Enable the Diagnosis of All
1194 Rare Genetic Diseases. Am J Hum Genet 2017 May 4;100(5):695–705. PMID:28475856

1195 56. Lochmüller H, Le Cam Y, Jonker AH, Lau LP, Baynam G, Kaufmann P, Lasko P, Dawkins
1196 HJ, Austin CP, Boycott KM. 'IRDiRC Recognized Resources': a new mechanism to
1197 support scientists to conduct efficient, high-quality research for rare diseases. Eur J
1198 Hum Genet Nature Publishing Group; 2017 Feb;25(2):162–165. PMID:27782107

1199 57. Mazzucato M, Pozza LVD, Facchin P, Angin C, Agius F, Cavero-Carbonell C, Corrochano
1200 V, Hanusova K, Kirch K, Lambert D, Lucano C, Maiella S, Panzaru M, Rusu C, Weber S,
1201 Zurriaga O, Zvolsky M, Rath A. ORPHAcodes use for the coding of rare diseases:
1202 comparison of the accuracy and cross country comparability. Orphanet J Rare Dis 2023
1203 Sept 4;18:267. PMID:37667299

1204 58. Mazzucato M, Visonà Dalla Pozza L, Minichiello C, Toto E, Vianello A, Facchin P.
1205 Estimating mortality in rare diseases using a population-based registry, 2002 through
1206 2019. Orphanet J Rare Dis 2023 Nov 17;18:362. PMID:37978388

1207 59. Pichon T, Messiaen C, Soussand L, Angin C, Sandrin A, Elarouci N, Jannot A-S, on behalf
1208 of the BNDMR infrastructure team. Overview of patients' cohorts in the French
1209 National rare disease registry. Orphanet J Rare Dis 2023 July 3;18(1):176.
1210 PMID:37400917

1211 60. Gunne E, McGarvey C, Hamilton K, Treacy E, Lambert DM, Lynch SA. A retrospective
1212 review of the contribution of rare diseases to paediatric mortality in Ireland. Orphanet J
1213 Rare Dis 2020 Nov 4;15(1):311. PMID:33148291

1214 61. Rico J, Echevarría-González de Garibay LJ, García-López M, Guardiola-Vilarroig S,
1215 Maceda-Roldán LA, Zurriaga Ó, Cavero-Carbonell C. The interoperability between the
1216 Spanish version of the International Classification of Diseases and ORPHAcodes:
1217 towards better identification of rare diseases. Orphanet J Rare Dis 2021 Mar
1218 9;16(1):121. PMID:33750434

1219 62. Walker CE, Mahede T, Davis G, Miller LJ, Girschik J, Brameld K, Sun W, Rath A, Aymé S,
1220 Zubrick SR, Baynam GS, Molster C, Dawkins HJS, Weeramanthri TS. The collective
1221 impact of rare diseases in Western Australia: an estimate using a population-based
1222 cohort. Genet Med 2017 May;19(5):546–552. PMID:27657686

1223 63. Scanlon P, Ridler G, Say G, Kellett M, Charlesworth J, Neil A, Dickinson JL, Burdon K, Jose
1224 M, Wallis M. Measuring the impact of rare diseases in Tasmania, Australia. Orphanet J
1225 Rare Dis 2024 Oct 28;19(1):399. PMID:39468681

1226 64. European Commission (Directorate-General for Health and Food Safety). European
1227 Reference Networks. Eur Comm – Public Health. Available from:
1228 https://health.ec.europa.eu/rare-diseases-and-european-reference-
1229 networks/european-reference-networks_en [accessed July 25, 2025]

1230 65. Orphanet, Inserm. Procedural document: Creation and update of disease summary
1231 texts for the Orphanet encyclopaedia for professionals. 2021 June. Report No.: 2.
1232 Available from:
1233 https://www.orpha.net/pdfs/orphacom/cahiers/docs/GB/Orphanet_DiseaseSummaryT
1234 ext_R1_Prod_sum_EP_02.pdf [accessed Aug 3, 2025]

1235 66. Zoch M, Gierschner C, Weidner J, Lorenz S, Henke E, Reinecke I, Kallfelz M, Weber S,
1236 Thomas C, Sedlmayr M. ORPHAcode, ICD-10, SNOMED and Co.: Evaluation of
1237 terminology mappings in the field of rare diseases. Health Technol 2025 May
1238 1;15(3):611–622. doi: 10.1007/s12553-025-00980-w

1239 67. Chiu ATG, Chung CCY, Wong WHS, Lee SL, Chung BHY. Healthcare burden of rare
1240 diseases in Hong Kong – adopting ORPHAcodes in ICD-10 based healthcare
1241 administrative datasets. Orphanet J Rare Dis 2018 Aug 28;13:147. PMID:30153866

1242 68. Navarrete-Opazo AA, Singh M, Tisdale A, Cutillo CM, Garrison SR. Can you hear us now?
1243 The impact of health-care utilization by rare disease patients in the United States.
1244 Genet Med 2021;23(11):2194–2201. PMID:34183788

1245 69. Directorate-General for Health and Food Safety, European Commission. Rare diseases.
1246 Eur Comm – Public Health. 2025. Available from: https://health.ec.europa.eu/rare-
1247 diseases-and-european-reference-networks/rare-diseases_en [accessed June 15, 2025]

70. European Commission – Joint Research Centre (JRC). Set of Common Data Elements for Rare Diseases Registration. EU RD Platf Eur Comm Sci Hub. Available from: https://eu-rd-platform.jrc.ec.europa.eu/set-of-common-data-elements_en [accessed June 15, 2025]

71. eHAction — Joint Action supporting the eHealth Network. Common Semantic Strategy for Health in the European Union (eHAction Deliverable D8.2.2). European Commission; 2019 May. Available from: https://health.ec.europa.eu/document/download/92ac8823-19c4-4641-bf4d-5d9bf021a600_en [accessed Aug 3, 2025]

72. Directorate-General for Health and Food Safety. eHN Guideline on Patient Summary (Release 3.4). European Commission; 2024 Nov. Available from: https://health.ec.europa.eu/system/files/2023-10/ehn_guidelines_patientsummary_en.pdf

73. The European Health Data Space. Terms and Conditions of Use of the Code Systems used in the MVC. Available from: https://webgate.ec.europa.eu/fpfis/wikis/display/EHDSI/Terms+and+Conditions+of+Use+of+the+Code+Systems+used+in+the+MVC [accessed Feb 25, 2026]

74. EHDSI. Code Systems content managed by the Central Terminology Services. Available from: https://webgate.ec.europa.eu/fpfis/wikis/display/EHDSI/Code+Systems+content+managed+by+the+Central+Terminology+Services [accessed Feb 25, 2026]

75. X-eHealth (eXchanging electronic Health Records in a commom framework) Consortium. D5.6 – Refine PS functional specifications to account for eHN Guidelines and rare diseases. European Commission; 2022. Available from: https://www.x-ehealth.eu/wp-content/uploads/2022/11/D5.6-Refine-PS-functional-specifications-to-account-for-eHN-Guidelines-and-rare-diseases.pdf [accessed June 15, 2025]

76. Executive Board, World Health Organization. Rare diseases: a global health priority for equity and inclusion (EB156(15)). World Health Organization; 2025 Feb. Available from: https://apps.who.int/gb/ebwha/pdf_files/EB156/B156_(15)-en.pdf [accessed July 21, 2025]

77. Work Package 5, RD-ACTION Joint Action. Recommendation for routine maintenance of codification resources for rare diseases (RD-ACTION Deliverable 5.5). European Commission (Health Programme) / RD-ACTION; 2018 July. Available from: https://www.rd-action.eu/wp-content/uploads/2018/09/677024_DEL5.5_Recommendation-routine-maintenance-codification.pdf [accessed July 25, 2025]

78. RD-CODE. Specification and implementation manual of the Master file for statistical reporting with ORPHAcodes. Available from: https://www.rd-code.eu/wp-content/uploads/2022/01/826607_D5-4_Specification-and-implementation-manual-of-the-Master-file-for-statistical-reporting-with-ORPHAcodes.pdf [accessed Feb 25, 2026]

79. RD-CODE. Standard procedure and guide for coding with Orphacodes. Available from: https://www.rd-code.eu/wp-content/uploads/2022/01/826607_D5-4_Standard-procedure-and-guide-for-coding-with-Orphacodes_final.pdf [accessed Feb 25, 2026]

80. OD4RD2. Semantic interoperability of data on rare diseases - ORPHAcodes as part of the coding system landscape. Available from: https://od4rd.eu/communication-material/OD4RD2%20MS%2015%20White%20Paper_final-V02.pdf [accessed Feb 25, 2026]