

Representation of Rare Diseases in Health Information Systems: The Orphanet Approach to Serve a Wide Range of End Users

Ana Rath,^{1*} Annie Olry,¹ Ferdinand Dhombres,^{1,2} Maja Miličić Brandt,³ Bruno Urbero,¹ and Segolene Ayme¹

¹Orphanet, Institut National de la Santé et de la Recherche Médicale (INSERM), US14, Paris, France; ²Université Pierre et Marie Curie and INSERM, U872, ICS, Paris, France; ³National Genetics Reference Laboratory, Manchester, United Kingdom

For the Deep Phenotyping Special Issue

Received 7 January 2012; accepted revised manuscript 28 February 2012.

Published online 15 March 2012 in Wiley Online Library (www.wiley.com/humanmutation).DOI: 10.1002/humu.22078

ABSTRACT: Rare disorders are scarcely represented in international classifications and therefore invisible in information systems. One of the major needs in health information systems and for research is to share and/or to integrate data coming from heterogeneous sources with diverse reference terminologies. ORPHANET (www.orpha.net) is a multilingual information portal on rare diseases and orphan drugs. Orphanet information system is supported by a relational database built around the concept of rare disorders. Representation of rare diseases in Orphanet encompasses levels of increasing complexity: lexical (multilingual terminology), nosological (multihierarchical classifications), relational (annotations—epidemiological data—and classes of objects—genes, manifestations, and orphan drugs—integrated in a relational database), and interoperational (semantic interoperability). Rare disorders are mapped to International Classification of Diseases (10th version), SNOMED CT, MeSH, MedDRA, and UMLS. Genes are cross-referenced with HGNC, UniProt, OMIM, and Genatlas. A suite of tools allow for extraction of massive datasets giving different views that can be used in bioinformatics to answer complex questions, intended to serve the needs of researchers and the pharmaceutical industry in developing medicinal products for rare diseases. An ontology is under development. The Orphanet nomenclature is at the crossroads of scientific data repositories and of clinical terminology standards, and is suitable to be used as a standard terminology.

Hum Mutat 33:803–808, 2012. © 2012 Wiley Periodicals, Inc.

KEY WORDS: rare diseases; nosology; classification; ontology; interoperability; relational database

Introduction

Rare disorders are defined in Europe as those with a prevalence lower than 5 in 10,000 people. According to the definition adopted in the Orphanet database (see below), there are around 6,000 rare disorders, including diseases, syndromes, and anomalies. Most of them are genetic in origin and almost all Mendelian disorders are rare. Because of their rarity, these disorders are scarcely represented in international classifications and therefore invisible in health information systems, contributing to their invisibility in society at large. For instance, hospital information systems in many countries use the WHO's International Classification of Diseases (ICD) in its 10th version, or even in the 9th version, to record patient information. Only around 500 rare diseases are listed in the ICD-10, and only half of these have their own specific code. As a result, clinical research on these disorders requires specific sources of information (disease registers) for epidemiological and research purposes, which are costly to set up and to maintain. The lack of specific codes for rare disorders in international classification systems is especially regrettable in light of the avalanche of new data generated by next-generation sequencing, likely to first benefit rare diseases [Lindblom and Robinson, 2011]. This lack of codes is also particularly unfortunate now that technology allows for the integration of multiple sources of information, including electronic health records, but also gene and protein databases in many species, genetic variations in humans and animal models [Webb et al., 2011], and medicinal chemistry databases such as PubChem, DrugBank, ChemBank IUPHAR, and ChEMBLdb [Blomberg et al., 2011]. The availability of open-access clinical databases in the field of rare diseases is likely to boost research, especially given that rare diseases are disease models that help to understand the physiopathology of diseases for the direct benefit of all patients. Currently, therapies developed for rare diseases represent a significant part of all innovations in the therapeutic field. In the United States in fiscal year 2011, 30% of the 35 innovative treatments approved have orphan indications (<http://www.fda.gov/downloads/AboutFDA/ReportsManualsForms/Reports/UCM278358>). The aim of this article is to give an overview on how rare diseases are represented in Orphanet, as well as the reasons for knowledge representation.

The Orphanet Approach to Representing Knowledge on Rare Diseases to Serve End Users

Besides its mission of producing and disseminating information on rare disorders, Orphanet has built a representation of knowledge on rare disorders, which is constantly evolving in order to

*Correspondence to: Ana Rath, Orphanet, Institut National de la Santé et de la Recherche Médicale, US14, 96 rue Didot, 75014 Paris, France. E-mail: ana.rath@inserm.fr

Contract grant sponsor: European Commission, Directorate Générale in Grants Health and Consumers (grant agreement RDPortal, JA RDTF); INSERM and the "Direction Générale de la Santé" in France; and GlaxoSmithKline.

better serve its diverse population of users. This representation can be analyzed in its levels of increasing complexity: lexical (terminology), nosological (classification), relational (annotations and classes of objects integrated in a relational database), and interoperational (semantic interoperability). According to the users' needs, data on rare disorders can be obtained at each one of these levels to be used for different purposes, either from the Orphanet Website (www.orpha.net), which is continuously updated, or from a dedicated download platform, OrphaData (www.orphadata.org), which is monthly updated. An ontology of rare diseases, OntoOrpha, is under development as well.

Inventory and Terminology of Rare Diseases

An inventory of rare disorders has been maintained since 1997 in order to provide an exhaustive list of terms designating rare diseases, rare syndromes, rare morphological or biological anomalies, and particular clinical situations considered as rare conditions. In the Orphanet inventory, a disorder is defined as a clinically unique, distinct entity, whatever the number and nature of the causes (i.e., number of causative genes, different modes of inheritance, etc.). The choice of clinical criteria to define what a rare disorder is was driven by the original aim of Orphanet, that is, to guide physicians and patients dealing with rare diseases to the appropriate source of information for them.

This inventory is updated monthly in order to: (1) include newly described disorders; (2) follow the evolution of knowledge, that is, grouping separate entities that are now known to be a unique entity or, on the contrary, separating different forms of a previously unique disorder that have proven to be distinct entities; (3) update the preferred terms used to designate disorders; and (4) expand synonymy.

More precisely, when a disorder is recognized to have been separated into subforms, the latter are created while the parent disorder is kept (and hierarchically related to its subforms, see below). Deletions in the inventory of rare disorders are limited to duplicate entries in the database, and traceability of decisions is assured in the back office.

Each rare disorder has a unique identifier, called the ORPHA number, which is stable over time. This unique identifier makes it possible to always refer to the same entity, even if its preferred name has changed over time.

The inventory of rare diseases can be used in medical information systems (i.e., hospital information systems) to code patients' records, thus allowing identification of those presenting with a rare disorder.

Classifications of Rare Disorders

In order to move on to a nosological level of representation of rare disorders, groups of disorders, as well as subtypes, were added to the Orphanet terminology, and further organized into classifications. Classifications accept an unrestricted number of granularity levels, and a clinical entity can be subdivided into several clinical and/or etiological subtypes. It allows, among other things, to map to any other resource in a one-to-one manner: entries in genetic-based databases (e.g., OMIM) map to deeper levels in the hierarchy (genetic subtypes), whereas clinical-based ones (e.g., ICD) will map to higher levels in the hierarchy. As an example, Silver–Russell syndrome (a clinical entity) encompasses the following etiological subtypes:

- Silver–Russell syndrome
- Silver–Russell syndrome due to 11p15 microduplication
- Silver–Russell syndrome due to 7p11.2p13 microduplication

- Silver–Russell syndrome due to imprinting defect of 11p15
- Silver–Russell syndrome due to maternal uniparental disomy of chromosome 11
- Silver–Russell syndrome due to maternal uniparental disomy of chromosome 7

Like disorders, groups of disorders, categories, and sub-types also have unique ORPHA numbers.

Classifications are produced and updated from different sources: (1) medical and scientific publications; (2) workshops with experts groups; (3) expert validation for restricted groups of disorders; and (4) international collaborative efforts such as the revision of the ICD.

The Orphanet classification of rare diseases has adopted a clinical approach to classification, with the aim of following the organization of medical specialties that roughly follow organization into body systems. Deeper in the different sections of the Orphanet classification of rare diseases, other criteria are followed, according to the intrinsic logic of the medical management in a given specialty. For instance, in the neuromuscular section of neurological diseases, a histopathological approach is further adopted for muscular diseases, distinguishing myopathies from dystrophies, as muscle biopsy is a first step in the diagnostic workup. Etiopathological classifications of particular groups of diseases are also produced, in order to have disorders classified by mechanisms or causative genes for instance.

The Orphanet classification is multihierarchical; each entry is classified in one or more classifications and in one or more sections of a single classification (multiple parentage). For example, Bardet–Biedl syndrome is classified as a developmental anomaly syndrome and as an endocrine disease, and, among endocrine diseases, as a genetic form of obesity and as a syndrome with hypogonadotropic hypogonadism. Furthermore, Bardet–Biedl syndrome is classified as a ciliopathy, and its siblings in this mechanistic classification are totally different than those in the clinical classifications (Table 1).

Relational Database of Rare Disorders

Rare disorder inventories and classifications are the core of a rich, complex relational database, encompassing other classes of entities, namely, a nomenclature of genes involved in rare disorders, a thesaurus of clinical manifestations, an inventory of pharmacological compounds (orphan drugs and orphan designations), and a directory of medical services (expert centers and diagnostic tests), patient organizations, and research activities (research projects, clinical trials, patients registries, mutation registries, and biobanks).

Furthermore, each rare disorder is annotated with its epidemiological data (prevalence, age of onset, and age of death) and its mode of inheritance for inheritable diseases. Each disorder is linked to one or more texts produced by Orphanet (expert-reviewed abstracts, emergency guidelines, and lay people encyclopedia), or by others (review articles, clinical practice guidelines, gene cards, practical genetics articles, etc.)

This relational database is constantly evolving to better represent knowledge. Ongoing changes include qualifying gene-disorder relationships (i.e., causality, susceptibility, and major role in phenotype). Disorders are qualified as diseases, syndromes, anomalies, particular clinical situations, groups of disorders, etiological subtypes, clinical subtypes, and histological subtypes.

Interoperability and Cross-Referencing

One of the major needs in health information systems and for research is to be able to share and/or to integrate data coming

Table 1. Comparison Between Siblings of Bardet–Biedl Syndrome Depending on the Classification Criteria*

Nephronophthisis-related ciliopathy	Syndromic retinitis pigmentosa	Genetic syndromic obesity
Bardet–Biedl syndrome	Bardet–Biedl syndrome	Bardet–Biedl syndrome
Alström syndrome	Alström syndrome	Alström syndrome
Ellis–van Creveld syndrome	Cohen syndrome	Cohen syndrome
Jeune syndrome	Usher syndrome	Prader–Willi syndrome
Joubert syndrome	Joubert syndrome with ocular defect	Borjesson–Forssman–Lehma syndrome
Joubert syndrome with oculorenal defect	Joubert syndrome with oculorenal defect	MEHMO syndrome
AR nephronophthisis	Cockayne syndrome	MOMO syndrome
Senior–Loken syndrome	Kearns–Sayre syndrome	Maternal UDP14

*These lists are not exhaustive.

from heterogeneous sources with diverse reference terminologies. This requires mapping between the terminologies in use in order to allow syntactic and semantic interoperability, that is, being able to achieve translation from one terminology into another, and to share a common conceptual context. The mapping process also allows terminology curation because the mapping itself, and its evaluation, may reveal inaccuracies in both the source and target terminologies, such as wrong synonyms, redundancies, or errors in concept hierarchies.

Orphanet undertook collaboration with the National Genetic Reference Laboratories (Manchester, UK) [Miličić Brandt et al., 2011] in order to map the Orphanet nomenclature of rare diseases with UMLS, SNOMED CT, MeSH, and MedDRA. Automated mappings are reviewed by independent reviewers, and they are qualified as “exact,” “narrower term–broader term,” “broader term–narrower term,” or “inexact.” All but inexact mappings will be displayed on the Orphanet Website in the near future, giving access to the target terminologies, and will be updated and checked once a year. Massive extraction of mappings via the OrphaData platform will be also available.

The Orphanet nomenclature is already mapped to the ICD-10 terms, although these mappings are not qualified. Orphanet is represented in the rare diseases Topic Advisory Group (RD-TAG) working on the revision of the ICD to produce its 11th version (ICD-11). Integration of rare diseases in the ICD-11 will allow one-to-one mapping between an ORPHA code and an ICD-11 concept, which is expected to significantly increase the visibility of rare diseases in medical information systems. The Orphanet nomenclature is also cross-referenced with OMIM.

The Genes nomenclature is also cross-matched with external knowledge databases such as HGNC, UniProt, Genatlas, and OMIM. These mappings allow for interfacing with additional scientific resources, which use one or more of these nomenclatures. To achieve this, partnerships are being established with several EBI resources (Reactome, Ensembl, for instance), IUPHAR, and further collaborations are expected, allowing interconnections between heterogeneous but complementary resources around rare diseases.

The Reasons for Moving from an Inventory of Rare Diseases to a Multihierarchical, Interoperable, Classification System, and to an Ontology of Rare Diseases

ORPHANET (www.orpha.net) is an information portal on rare diseases and orphan drugs jointly established in 1997 by the Institut National de la Santé et de la Recherche Médicale (INSERM; the French National Institute of Health and Medical Research) and the French General Directorate for Health in order to provide healthcare

professionals and the general public with valuable information that can be used to improve diagnosis and care. In this respect, a multilingual information portal was created, composed of an inventory of rare diseases and an online free-access encyclopedia as well as a directory of expert clinics, medical laboratories, research projects, clinical trials, registries, biobanks, drugs in development and on the market, and patient organizations [Ayme et al., 1998]. It became a European project in 2000 with funding from the European Commission (EU). It is now an EU Joint Action with a network of almost 40 countries contributing to the collection of data and to the translations. It is under the responsibility of the INSERM, which has a specific unit fully dedicated to information and services in the field of rare diseases and orphan drugs. In 2010, on average, 10,000 visitors were registered per day with 1/3 of them being doctors, 1/5 other healthcare professionals, and 1/3 patients, viewing over 1,000,000 pages per month. Researchers (8%), pharmaceutical industry (6%), and health policy makers (1%) are the other categories of users (<http://www.orpha.net/orphacom/cahiers/docs/GB/ActivityReport2010.pdf>).

The Orphanet information system is supported by a relational database built around the concept of distinct clinical entity that we call a disorder in this article. Every piece of information is linked to a disorder or a group of disorders (Fig. 1).

From 1997 to 2007, the inventory of rare diseases was a flat list, each piece of information being linked to the ad hoc list of diseases. This was not a problem for the information on pharmaceutical products, which are always developed for one single disease or a very limited number of diseases, but became a problem when linking expert clinics to specific diseases. Some clinics were highly specialized, such as expert centers for cystic fibrosis or for Gaucher disease, whereas others, equally relevant, covered a larger group, such as rare pulmonary diseases in childhood (which includes cystic fibrosis) or clinics for lysosomal storage diseases (which includes Gaucher disease). The problem did not concern linking the expert resources appropriately, but rather enabling Website users to retrieve the appropriate information. Those querying for “lysosomal storage disease” would not find the Gaucher clinics. This was also a problem for the encyclopedia as some review articles covered a group of diseases while others dealt with a specific disease. It became evident that a hierarchical representation of rare diseases was necessary to better serve end users by allowing them to query at any level, from very specific subtypes to large entities and still have an answer. However, the population of users of the Website was heterogeneous, including experts, nonexpert health professionals, patients, families, researchers, healthcare managers, and bioinformaticians. Their representation of the field differed, as well as their expectations and needs. We started by collecting classifications published in text books and peer-reviewed articles, and adapted these classifications to our own identified needs. We needed a classification of diseases reflecting clinical pathways, and for it to be suitable for retrieving

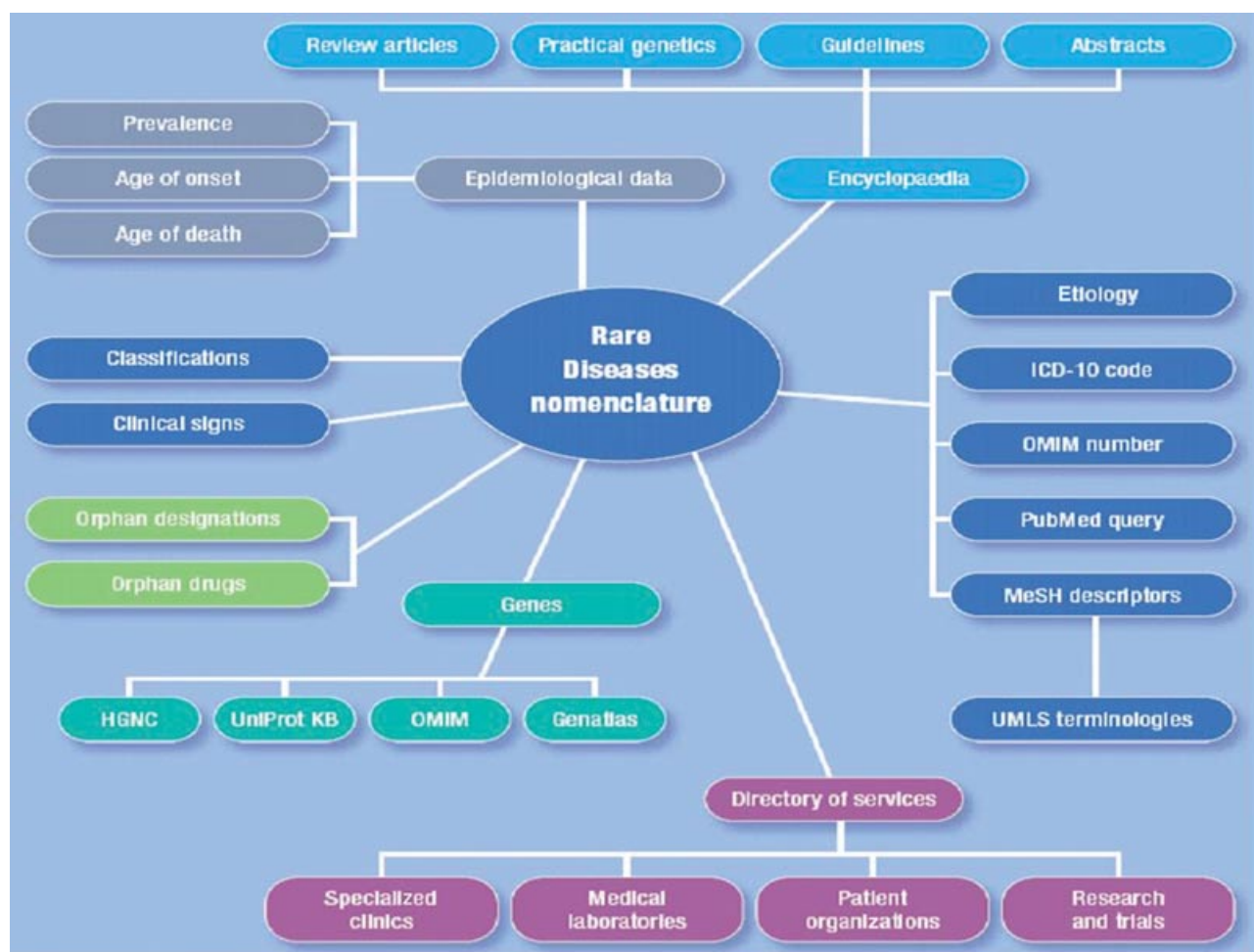


Figure 1. Schematic representation of the Orphanet relational database centered on disorders.

information on expert clinics. This was a relatively easy exercise for diseases in which the phenotype corresponds to a single medical specialty, that is, 3,231 of the 5,954 diseases in Orphanet, although their position in the internal hierarchy could also be difficult. The question of differentiating between diseases expressed in childhood to those expressed in adulthood had to be considered as most expert clinics either see exclusively children or adults. For the rest of the 2,723 diseases, it was impossible to assign them to a clear medical specialty as the phenotype included manifestations managed by different specialists. It was necessary to determine which specialty clinics were the main entry point for these phenotypes, and which ones were secondary. This resulted in the establishment of a multi-hierarchical ad hoc classification. The same approach was taken to build another hierarchy for the research activities. A logic based on mechanisms and etiologies was used to build up new classifications. This multihierarchical, multipurpose classification system was the basis of the fourth version of Orphanet released in 2008.

After the launch of this new version, Orphanet started to receive requests from researchers, both from the academic and from the industry sector, who were interested in making use of our dataset. The dialogue with them, and the comments received from Website users through our annual user satisfaction survey, led to a shift in our approach toward a representation of the knowledge for bioinformatics and the semantic Web. It prompted us to increase mapping between

the Orphanet nomenclature of rare disorders, including the inventory of genes involved in rare disorders, and other resources (i.e., medical terminologies and scientific databases). Interoperability of the Orphanet nomenclature of rare disorders with other biomedical terminologies together with the cross-referencing of the genes nomenclature with external databases contributed to the creation of a crossroads system between knowledge databases/ontologies using different terminologies and having very different perspectives (genes annotation, proteins annotation, fundamental pharmacology, physiological pathways, etc.). It was intended to provide researchers with access to integrated resources of interest in the field of rare diseases.

One of the use cases of the Orphanet database is that of the inventory and classifications of rare diseases that can be used in medical information systems (i.e., hospital information systems) to code patients' health records, thus allowing for the identification of those presenting with a rare disorder. As an example, it has been used since 2006 in France by CEMARA, a clinical database collecting patient information from 55 French centers of expertise (and their network) on rare diseases [Messiaen et al., 2008; Landais et al., 2010]. They allow for retrieval of information by group of diseases or by medical domain. They also allow for coding of patients with a category code when there is no specific diagnosis given as yet.

A suite of tools was developed in order to allow extraction of massive datasets giving a view of the relational database. These views can

be used by bioinformaticians to answer complex questions, such as what is the distribution of fundamental research by medical domain or by disease, how is funding for research distributed by disease, what is the landscape of centers of expertise in Europe, and so on. Some of these views are freely available at OrphaData (www.orphadata.org), which was launched in June 2011 and some are available on request, after signature of a material transfer agreement. Massive extraction of mapping between the Orphanet inventory of rare disorders and genes and other external resources will also be available in 2012 via the OrphaData platform. OrphaData is intended to better serve the needs of health information systems, those of researchers, as well as those of the pharmaceutical industry in developing medicinal products for patients with rare diseases.

Evolution to a Real “Ontology of Rare Diseases”

The growing complexity of the Orphanet knowledge base, in terms of maintenance and quality control, as well as in terms of representation and interoperability (see above), called for the construction of an ontology of rare disorders. The aims of building up the Orphanet ontology of rare diseases (OntoOrpha), besides the relational database and connecting to it, is to allow for appropriate knowledge management (edition, curation, validation, and quality control) in a simpler and more accurate manner. This ontology also aims to facilitate semantic interoperability with other ontologies, terminologies, and databases, and to provide decision support [Bodenreider, 2008]. Ontology-based automated reasoning ensures consistency across the knowledge model (compliance with the logical rules underlying the knowledge model) and allows new meaningful relationship generation (or to express implicit relationships between classes of objects in the ontology [Dhombres et al., 2011]). This is possible because the ontology offers a formalization of the domain semantics (including concepts of classification, group of disorders, disorder, malformation syndrome, clinical subtype, etiological subtype, gene, manifestations, and many others, and including relationships between these concepts: “manifestation of,” “frequent manifestation of,” “gene of,” “causal gene of,” etc.) that can be interpreted by computers. Each concept (disorders and genes) of OntoOrpha is annotated with external references and codes: ICD-10, HGNC, UniProt, OMIM, and GenAtlas. In 2012, new references will be added: SNOMED CT, MeSH, MedDRA, and UMLS.

The Orphanet ontology of rare diseases is currently under construction by a multidisciplinary team including domain experts (MDs and biologists), information scientists, computer scientists, and knowledge engineers. A beta-version of OntoOrpha has been released in BioPortal (www.bioportal.bioontology.org/ontologies/1586) and is freely accessible. It enables users to follow the development of the under-construction ontology of rare diseases, but it does not, for the moment, provide access to the whole Orphanet database content.

Discussion

Representing knowledge on diseases to serve different categories of users (e.g., physicians, researchers, and decision makers) is challenging because each particular use corresponds to a particular semantic view, including a particular definition of a disease. There are in fact multiple ways to define what a disease or a disorder is. What we call a “disorder” is a representation of a pathological process or state according to a point of view adopted in a particular context (universe of knowledge) and with a particular aim. A “disorder” encompasses multiple dimensions that can be classified, among others, in clin-

Table 2. Examples of Different Definitions of a Same Entity According to Its Different Dimensions: Congenital Adrenal Hyperplasia

Pathophysiology	Group of genetic conditions due to an anomaly in the biosynthesis of adrenal hormones, which can be either complete (classic form) or incomplete (atypical forms).
Etiology	Refers to any of several autosomal recessive diseases resulting from mutations of genes coding for enzymes, mediating the biochemical steps of production of cortisol from cholesterol by the adrenal gland (steroidogenesis).
Diagnosis (hormonal profile)	Congenital adrenal hyperplasia (CAH) is a rare disorder characterized by cortisol deficiency, with/without aldosterone deficiency, and androgen excess.
Clinical manifestations	Also called adrenogenital syndrome, it is a cause of sexual ambiguity at birth in its classic form, and of female virilization, precocious puberty, and hypofertility in its atypical forms.
Prognosis	Adrenal glands anomaly that, if untreated, lead to growth and pubertal impairment, and to a risk of potential life-threatening dehydration.
Management	Adrenal diseases requiring life-long hormone replacement therapy.
Epidemiology	Adrenal disease that is rare in Europe in its classic form, and a frequent disease in its atypical forms in some identified populations (Ashkenazi Jews, Italian, and Spanish populations).
Public health	A congenital disorder deserving newborn screening.

ical, etiological, physiopathological (mechanistic), and therapeutic and many others, each one of these categories encompassing further dimensions (Table 2). For example, from a clinical point of view, X-linked adrenoleukodystrophy can be defined either as a neurological disease or an endocrinological disease, and further, as a dementia or as an adrenal insufficiency or a cause of male infertility. . . depending on the point of view chosen (i.e., by an endocrinologist or by a neurologist). We can also look deeper into this entity to isolate, in the spectrum of its clinical manifestations, clinically relevant subforms (i.e., juvenile X-linked adrenoleukodystrophy and the adult form adrenomyeloneuropathy). These subforms can be considered as diseases per se, X-linked adrenoleukodystrophy not longer being a unique entity, that is, the “disorder” level, but a group of two distinct disorders.

A clinical entity can also be considered from an etiological point of view, such as different pathogenic agents or strains of a bacterium causing more or less the same disease, or different mutated genes resulting in a single phenotype. For the geneticist and for diagnostic purposes, each one of the genetic alterations underlying a disorder can be considered as an entity (e.g., CLN-10 is different from CLN-6, even though they both cause the late infantile form of ceroid lipofuscinosis).

In other words, depending on the context and on our interest, we will isolate an entity from a double continuum: a phenomic continuum of growing comprehensiveness, from the clinical manifestations to the category of disorders; and a nosological continuum, from the genetic mutation to the group of disorders encompassing this kind of abnormality. For example, in the phenomic continuum ranging from shortness of breath and cyanosis to interstitial lung disease, and passing through restrictive pneumopathy and idiopathic pulmonary fibrosis, we can adopt a physiopathological point of view and consider that “our” disease level is “restrictive pneumopathy,” which makes sense in a diagnostic and therapeutic perspective. In the nosological continuum, going from anomalies of *CYP21A2* to sterol-dependent hormone synthesis deficiency, we can focus our interest on a particular mutation that will be at the right level of granularity we are dealing with.

The challenge in the representation of knowledge is to build up a hierarchical and relational system that can be used by different categories of users at different levels of granularity, and to make this system able to handle connections with others, allowing for crossroads between complementary sources of knowledge. In the Orphanet relational database, four main classes of objects are interconnected: disorders (organized in a multihierarchical, multidimensional classification), genes, clinical manifestations, and orphan drugs. Each one of these classes has relationships with external resources (databases, terminologies, and ontologies) or in-house resources (inventory of clinical trials and PubMed queries, among others).

The Orphanet nomenclature is a new standard terminology to be incorporated into the UMLS, as it is complementary to the other existing ones, serving different purposes. The closest terminology is the OMIM one [Amberger et al., 2011], which pursues the task of cataloging the association between human phenotypes and their causative genes. The excellent cross-referencing of OMIM with other genomics databases makes it the number one source for this type of information. This is why Orphanet keeps track of OMIM development and interfaces its inventory of rare diseases with OMIM. The feature that differentiates Orphanet from OMIM is the proposed multihierarchical classification, which allows, for instance, all associated genes to be linked to a group of diseases. In addition, Orphanet contains rare diseases with no known genetic basis.

SNOMED CT is a sophisticated biomedical vocabulary to be used in electronic health records currently underrepresenting rare diseases. The cross-referencing of Orphanet and SNOMED CT provides the opportunity to inform SNOMED of the missing terms, contributing to better adaptation of this vocabulary to the needs of the healthcare professionals using it.

MeSH is a National Library of Medicine (NLM) medical terminology intended to index the medical literature. As a consequence of the alignment between the Orphanet nomenclature of rare disorders and the MeSH, new specific PubMed queries will be available in Orphanet, the number of mappings between Orphanet and MeSH having increased because of the incorporation into the MeSH of many rare diseases listed by the NIH Office of Rare Diseases Research. Feedback to the NLM could be also provided in order to improve the representation of rare diseases into the MeSH.

The ICD is the international standard diagnostic classification used for epidemiological studies, healthcare management purposes, and clinical use. It also provides the basis for the compilation of national mortality and morbidity statistics by WHO Member States. The last edition, ICD-10, was adopted in 1990 and is currently under revision for adoption in 2014, by a very large group of stakeholders. The process is organized through TAGs, which organize the con-

sultation of experts and drafting of the alpha version, revising the structure and adding the clinical entities that have emerged since 1990. The establishment of a RD-TAG demonstrates the willingness of WHO to incorporate rare diseases in the next edition, using the Orphanet terminology and clinical classification system as a template. Although ICD intends to offer a multihierarchical classification of the diseases in its next electronic edition, it is likely that it will be limited to views useful to record mortality and morbidity with a clear public health perspective. Views intended to serve highly specialized care and research are less likely to be taken into account. Therefore, the Orphanet nomenclature will be aligned with the next edition of ICD from now on, easing the process of migration from ICD-10 to ICD-11 for data related to rare diseases if coded with the Orphanet nomenclature. In addition, Orphanet will offer other classifications directed more toward research and specialized care.

The Orphanet nomenclature is positioned at the crossroads of scientific data repositories and of clinical terminology standards in order to form a bridge between these two worlds and contribute to translational medicine. It is a very dynamic project that is constantly adapting to the needs of its end users, but is sufficiently stable to be used as a standard terminology.

References

- Amberger J, Bocchini C, Hamosh A. 2011. A new face and new challenges for online Mendelian inheritance in man (OMIM). *Hum Mutat* 32:564–567.
- Ayme S, Urbero B, Oziel D, Lecouturier E, Biscarat A. 1998. Information on rare diseases: the Orphanet project. *Rev Med Interne* 19:376–377.
- Blomberg N, Ecker G, Kidd R, Mons B, Williams-Jones B. 2011. Knowledge driven discovery goes semantic. *Eur Fed Med Chem Yearbook* 39–43.
- Bodenreider O. 2008. Biomedical ontologies in action: role in knowledge management, data integration and decision support. *Yearb Med Inform* 67–79.
- Dhombres F, Vandenbussche P-Y, Rath A, Olry A, Hanauer M, Urbero B, Charlet J. 2011. OntoOrpha : an ontology to support edition and audit of rare diseases knowledge in Orphanet. *Proceedings of the 2nd International Conference on Biomedical Ontology (ICBO-2011)*. Buffalo, NY, USA: Olivier Bodenreider, Maryann E. Martone, Alan Ruttenberg (eds.); pp. 241–243.
- Landais P, Messiaen C, Rath A, Le Mignot L, Dufour E, Ben Said M, Jais JP, Toubiana L, Baujat G, Bourdon-Lanoy E, Gerard-Blanluet M, Bodemer C, and others. 2010. CEMARA task force. CEMARA an information system for rare diseases. *Stud Health Technol Inform* 160:481–485.
- Lindblom A, Robinson PN. 2011. Bioinformatics for human genetics: promises and challenges. *Hum Mutat* 32:495–500.
- Messiaen C, Le Mignot L, Rath A, Richard JB, Dufour E, Ben Said M, Jais JP, Verloes A, Le Merrer M, Bodemer C, Baujat G, Gerard-Blanluet M, and others. 2008. CEMARA: a Web dynamic application within a N-tier architecture for rare diseases. *Stud Health Technol Inform* 136:51–56.
- Milićić Brandt M, Rath A, Devereau A, Ayme S. 2011. Mapping Orphanet terminology to UMLS. In: *Proceedings of the 13th Conference on Artificial Intelligence in Medicine*. Berlin, Heidelberg: Springer-Verlag; pp. 194–203.
- Webb AJ, Thorisson GA, Brookes AJ; GEN2PHEN Consortium. 2011. An informatics project and online “Knowledge Centre” supporting modern genotype-to-phenotype research. *Hum Mutat* 32:543–550.