How to code rare diseases with international terminologies ?

Ana Rath

Inserm US14, Paris, France

ana.rath@inserm.fr

Special thanks to Prof Paul Landais for his kind presentation.

Needs for terminologies

- Codify patients with a given disease
  - Diagnosis
  - Need terminologies for diseases
- Describe phenotypes
  - Genotype-phenotype correlations
  - Need terminologies for phenotypes (meaning signs and symptoms)
- Describe genotypes
  - Mutation databases
  - Need terminologies for genes and for variations
    - ✓ Won’t be discussed here

Levels of granularity

- **Disorders**
  - Purpose: coding diagnoses (i.e. medical records, patient registries)
- **Clinical manifestations**
  - Purpose: describing patients, genotype-phenotype correlations, … (i.e. assistance-to-diagnosis tools, research databases)
- **Specialized terms**
  - Fit the particular needs of a disease-focused database/registry (i.e. Phe values in PKU and related disorders)

Diseases, phenotypes

- Describe a clinical situation:
  - Making a conclusion = diagnosis
  - Describing a patient = phenotypic features
- Using language
  - To make the annotation (for ourselves)
    - ✓Uncontrolled medical language
  - To retrieve the information (for analysis)
    - ✓Controlled terminology
  - To communicate with others
    - ✓Common nomenclature, mappings
  - To exchange the information
    - ✓Common IT format for data sharing

Coding disorders

Need for coding RD

- **Make RD visible** in order to:
  - Have sound epidemiological data
  - Document the natural history of RD
  - Identify patients from health records for clinical research
  - Bring clinical data to research
- Since different systems are using different **terminologies**, the latter **should be inter-operable**
- There is a **need to have a common language** to allow for sharing clinical data between health care centres and databases and registries:
  - Patients are rare and scattered
  - Significant amounts of data are necessary to perform research

The current situation

- Most health information systems use **ICD**
  - Some ICD-9
  - Most ICD-10
- WHO' ICD-11 revision is expected for 2017
- Some countries have adopted **SNOMED CT**
- Genetic databases use **OMIM**

- In Europe, countries having national plans/strategies for RD decided to integrate the **Orphanet nomenclature** of RD, and code patients with the **ORPHA** code

Terminologies currently used

- **SNOMED CT** (Systematized Nomenclature of Medicine – clinical terms, IHTSDO); 401,200 terms
  - comprehensive clinical terminology,
  - multihierarchical ontology
  - intended for use in EHR, and to semantic interpretation of EHR
  - translated in licensed countries

- **ICD-10** (International classification of diseases- WHO), 12,451 terms
  - Monohierarchical classification of diseases
  - Intended for statistical uses (morbidity, mortality)
  - Translated and adapted in different countries

Other terminologies/resources

- **OMIM** (Online Mendelian Inheritance in Man):
  - Genetic disorders and phenotypes (regardless their rarity)
  - Organized by genes
  - English only
  - Use in (genetic) databases

Other terminologies/resources

- **MeSH** (Medical Subject Headings; NLM); 242,262 terms
  - Medical terminology intended at indexing medical publications
  - Translated
- **MedDRA** (Medical Dictionary for Regulatory activities); 73,742 terms
  - Standardised international medical terminology,
  - Used for registration, documentation and safety monitoring of medicinal products across the phases of the development cycle.
  - Translated in 10 languages
- **UMLS** (Unified Medical Language System, NLM); 2,930,638 concepts (> 11,300,000 terms)
  - Integration of terminologies, classifications and coding standards
  - Intended for biomedical information and interoperability
  - Submitted to licensing
  - Translated (partially)

How many RD are included in these terminologies?

- **ICD10**
  - **466 specific codes** matching Orphanet rare disease entities (including groups of diseases) (= EXACT mappings)
  - 431 inclusion terms matching Orphanet RD entities
  - 82 index terms matching Orphanet RD entities
  - → Total: only 979 Orphanet RD entities with an ICD-10 mention
  - But >80% of ORPHA entries have been attributed an ICD10 code
- **SNOMED CT** (from UMLS AA2013)
  - On 15,043 candidate mappings, 3,541 were EXACT (**2,883 ORPHA** entries)
- ✓ **OMIM**
  - On 6,617 total mappings, 3,388 are EXACT (**3,380 ORPHA** entries)

The ICD - 11

- Rare diseases Topic Advisory Group (RD-TAG)
  - Chair: Ségolène Aymé
  - Managing editor: Ana Rath
  - Information scientist: Bertrand Bellet
- The aim is to include RD with a specific code
- Revision process for rare diseases involved the major experts and networks in the field, worldwide

- WHO's target release: 2017
- Next: adoption by countries

➢ Using ICD11 to code RD will take a long time

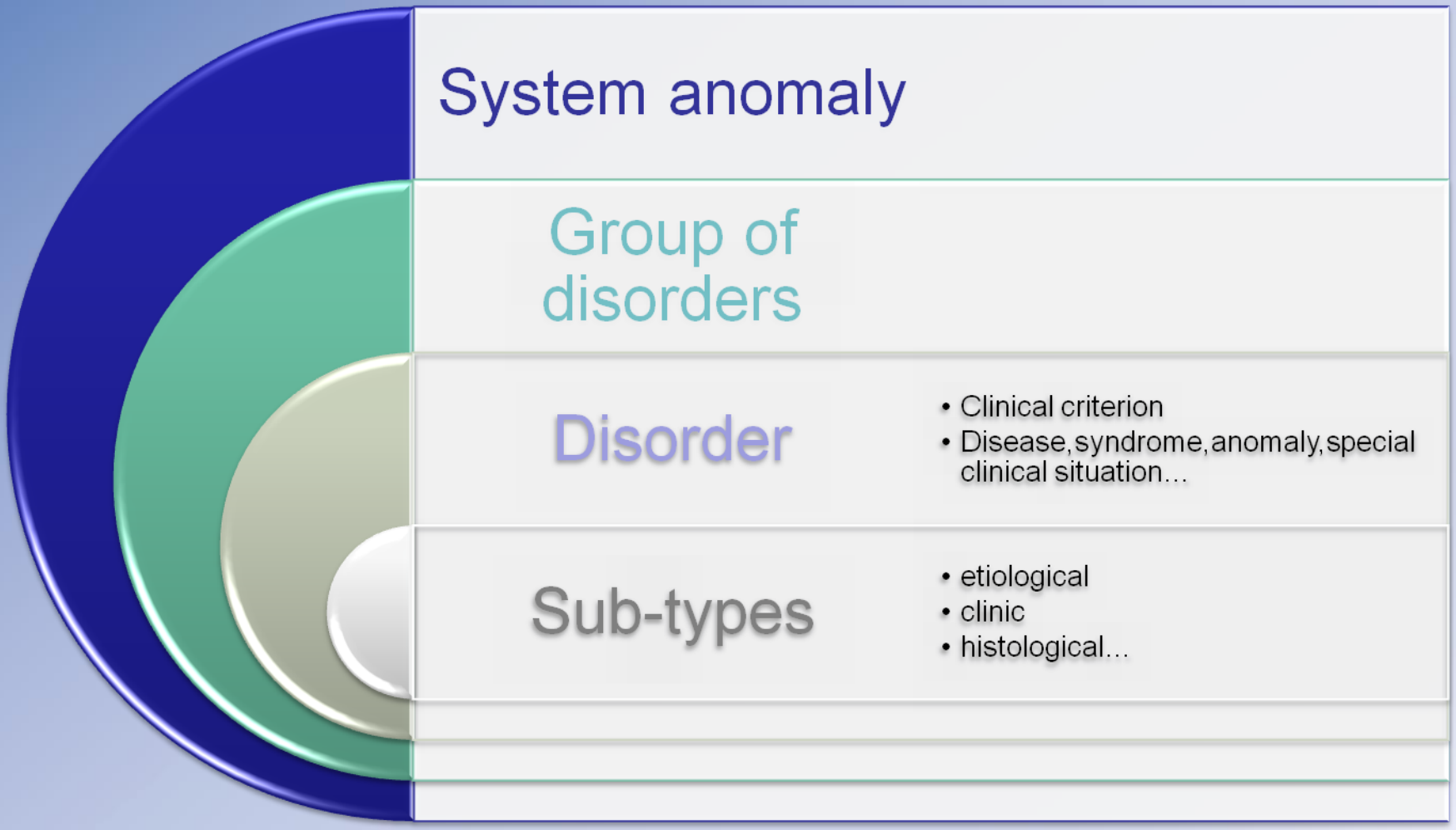
How the Orpha nomenclature and classification are produced?

Rare disorders in Orphanet

- Since 1997: Inventory of rare disorders (prev<1/2 000)
  - Mapped to OMIM
- 2005: Mapping to ICD-10
- 2007: Classification of rare disorders
- 2011: Mappings to UMLS, SNOMED CT, MeSH, MedDRA
- 2014: ORDO (Orphanet ontology of rare diseases) in collaboration with the EBI.

Overview of the Orphanet content production process

Disorders are organised according to their typology, based on clinical criteria



A multidimensional classification

- Orphanet classifications by medical specialty based on international literature and experts advice

View classifications by disease or by group of diseases

Rare metabolic disease
Metabolic disease involving complex molecules
Peroxisomal disease
Adrenoleukodystrophy, X-linked
Adrenoleukodystrophy, X-linked, cerebral form
Adrenomyeloneuropathy

Rare neurologic disease
Neurometabolic disease
Adrenoleukodystrophy, X-linked
Adrenoleukodystrophy, X-linked, cerebral form
Adrenomyeloneuropathy

Rare neurologic disease
Rare epilepsy
Metabolic diseases with epilepsy
Peroxisomal disease
Adrenoleukodystrophy, X-linked
Adrenoleukodystrophy, X-linked, cerebral form
Adrenomyeloneuropathy

Rare neurologic disease
Leukodystrophy
Adrenoleukodystrophy, X-linked
Adrenoleukodystrophy, X-linked, cerebral form
Adrenomyeloneuropathy

Rare endocrine disease
Rare adrenal disease
Primary adrenal insufficiency
Chronic primary adrenal insufficiency
Genetic chronic primary adrenal insufficiency
Adrenoleukodystrophy, X-linked
Adrenoleukodystrophy, X-linked, cerebral form

Fanconi anemia

Orpha number	:	ORPHA84	ICD-10	:	D61.0
Synonym(s)	:	Fanconi pancytopenia	OMIM	:	227645 [↗] 227646 [↗] 227650 [↗] 300514 [↗] 600901 [↗] 603467 [↗] 609053 [↗] 609054 [↗] 610832 [↗] 613390 [↗] 613951 [↗] 614083 [↗] 614083 [↗] 614087 [↗] 615272 [↗]
Prevalence	:	1-9 / 1 000 000			
Inheritance	:	Autosomal recessive X-linked recessive			
Age of onset	:	Childhood	UMLS	:	C0015625
			MeSH	:	D005199
			MedDRA	:	10055206
			SNOMED CT	:	30575002

Identity card

pancytopenia with bone marrow failure, variable congenital malformations and predisposition to develop hematological or solid tumors.

Recent determination of the carrier frequency gave an estimate of more than 1/200, with an expected prevalence at birth of at least 1/160,000. In certain populations, the carrier frequency is much higher, due to founder mutations. Until now, more than 2,000 cases have been reported in the literature.

In 2/3 of patients, the first signs of FA are congenital malformations that may involve the skeleton, skin, uro-genital, cardio-pulmonary, gastrointestinal and central nervous systems. Limb anomalies are unilateral or bilateral, the latter being frequently asymmetrical. Minor anomalies can also be present such as low height and weight, microcephaly and/or microphthalmia. Skin pigmentation abnormalities and hypoplastic thenar eminence are frequent. Almost 20% of patients have ear malformations with or without hearing loss. Congenital malformations may vary in a family. When congenital malformations are not prominent, diagnosis may be delayed until the onset of bone marrow failure (BMF), which occurs at a median age of 7 years. Hematologic abnormalities may occur at a younger age and, more rarely, in adults, with 90% of patients developing BMF by 40 years of age. Patients may develop acute myeloid leukemia, often preceded by myelodysplastic syndrome. Patients are also highly predisposed to solid tumors, of the head and neck or anogenital regions. Short stature is often secondary to hormonal deficiencies. Fertility is almost totally impaired in males, and is highly disturbed in half of females. Pregnancy is often complicated.

FA is due to mutations in genes involved in DNA repair and genomic stability. Fifteen genes representing 15 complementation groups have been identified.

Given the high heterogeneity in genetic causation and clinical phenotype, and the pathogenic mechanism of FA, diagnosis relies on the evaluation of chromosomal breakage induced by diepoxybutane (DEB) or mitomycin C (MMC).

FA clinical manifestations overlap with many malformation syndromes (Dubowitz, Seckel, Holt-Oram, Baller-Gerold, thrombocytopenia-absent radius, Nijmegen breakage syndromes, VACTERL association, dyskeratosis congenita; see these terms) and diagnosis of FA is often delayed until a patient develops BMF or malignancies. FA should be considered in the differential diagnosis of all young patients with BMF of unknown etiology. Other cancer predisposition syndromes (Bloom, Rothmund-Thomson or Werner syndromes; see these terms) or syndromes with pancytopenia (Diamond-Blackfan anemia, immune pancytopenia, Pearson or Shwachman-Diamond syndromes; see these terms) should be considered.

Prenatal diagnosis is feasible with a DEB-induced chromosomal breakage assay or by molecular study when the mutation is known.

FA is usually an autosomal recessive disorder but X-linked transmission may occur.

- Classification(s) (7)
- Gene(s) (16)
- Publications in PubMed [↗]
- Other website(s) (8)

Health care resources for this di

- Expert centres (331)
- Diagnostic tests (73)
- Patient organisations (34)
- Orphan drug(s) (1)

Research activities on this disease

- Research projects (34)
- Clinical trials (1)
- Registries/biobanks (39)
- Networks (16)

Orphanet Reports series

- Prevalence
- Orphan drugs in Europe

Getting involved /informed

- Read the newsletter
- Read OJRD [↗]
- Register your activity

The documents contained in this web site are presented for information purposes only. The material is in no way intended to replace professional medical care by a qualified specialist and should not be used as a basis for diagnosis or treatment.

Additional information (genes, classifications, PubMed, websites)

Abstract in 7 languages

Orphanet nomenclature & classification are monthly updated

- **Demands** : literature survey, experts, expert ressources, classifications...
- **Decisions** on:
  - Creation of new entries (new described entities/lacking entries)
  - Modification (reorganization) of entries
  - Obsolescence/deprecation of entries (i.e. double entries ; « moved to » entries)
  - Revision of classifications
  - For each entry:
    - ✓ Nomenclature: preferred term and synonyms; key-words if needed
    - ✓ Type of phenome: group/disease-syndrome/clinical subtype/etiological subtype/ « moved » to entry/historical entity/non rare disease…
    - ✓ Classification
    - ✓ Information attached to it: type of text, epidemiological data available, genes, OMIM numbers

Naming rules for Orphanet entries

- The ***preferred term*** is the primary disease identification. Usually, it is the most generally accepted name in the medical community. This can be defined by :
  - a published consensus
  - expert advice
  - compelling predominance of the name in medical literature

- ***Synonyms*** are perfect equivalents of the preferred terms, except that they do not fit so well the defining criteria of a preferred term.
  - Abbreviations (initialisms, acronyms) are included when actually used in literature.
  - Subentities must not be included among synonyms

Translation of preferred terms and synonyms should follow the same rules, and adapted to the local language situation (most widely used, medical acceptance, inclusion of all relevant synonyms)

Mappings to other terminologies

- Disorders mapped to OMIM (manually)
- Disorders mapped to ICD-10 (manually)
- Disorders mapped to UMLS, MeSH, SNOMED CT, MedDRA (semi-automatically)
- Mappings are qualified (exact ; narrow-to-broad ; broad-to-narrow)
- Information on the validation status is noted
- Updates depending on the target terminology
  - Monthly (ICD10, OMIM)
  - Twice a year (UMLS, SNOMED CT, MeSH, MedDRA)

Qualifying mappings

E	exact mapping (the terms and the concepts are equivalent)
NTBT	narrower term maps to a broader term
BTNT	broader term maps to a narrower term
W	incorrect mapping (two different concepts)
NTBT/E	narrower term maps to a broader term because of an exact mapping with a synonym in the target terminology
BTNT/E	broader term maps to a narrower term because of an exact mapping with a synonym in the target terminology
W/E	incorrect mapping (two different concepts) but syntactically exact mapping to a synonym or a preferred term in the target terminology
ND	not yet decided/unable to decide
The following are attributed to ICD10 codes only :	
Specific code	The term has its own code in the ICD10
Inclusion term	The term is included under a ICD10 category and has not its own code
Index term	The term is oncluded in ICD10 index and refers to one more general code
Attributed code	The term does not exist in ICD10 and a code was attributed by Orphanet

Mapping Orphanet to OMIM

- Follow-up of OMIM monthly updates
  - New OMIM entries: if relevant, mappings are manually done and qualified
  - Modified OMIM entries: the current mapping is revised… but this is not always easy.

Some examples

OMIM 612337 “Chromosome 1q43-q44 deletion syndrome” maps exactly to ORPHA36367 “Distal monosomy 1q”.

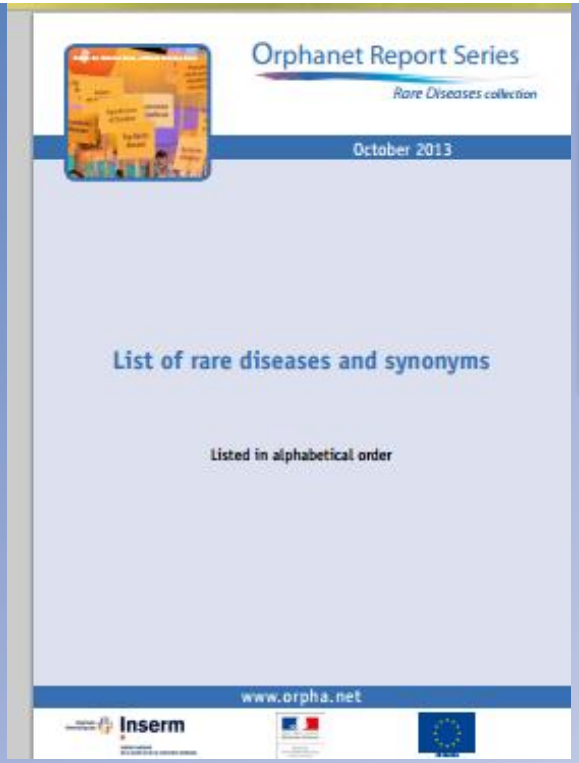
October 2013: OMIM 612337 becomes “Mental retardation, autosomal dominant 22; MRD22” and Deletion 1q43-q44 is included together with mutations in ZBTB18 gene

Mapping to ORPHA36367 is changed to NTBT

Some examples

- OMIM 606369 "Epileptic encephalopathy, typical Lennox-Gastaut" mapped exactly to ORPHA2382 “Lennox-Gastaut syndrome”.
- In September, 2013, the concept and the name of the OMIM 606369 change.
  - The preferred term becomes "Macrocephaly and epileptic encephalopathy",
  - It describes a single publication (2001) which does not correspond any more to Lennox-Gastaut syndrome
  - There is currently no OMIM entry for Lennox-Gastaut
- OMIM 606369 has been unlinked from ORPHA2382.

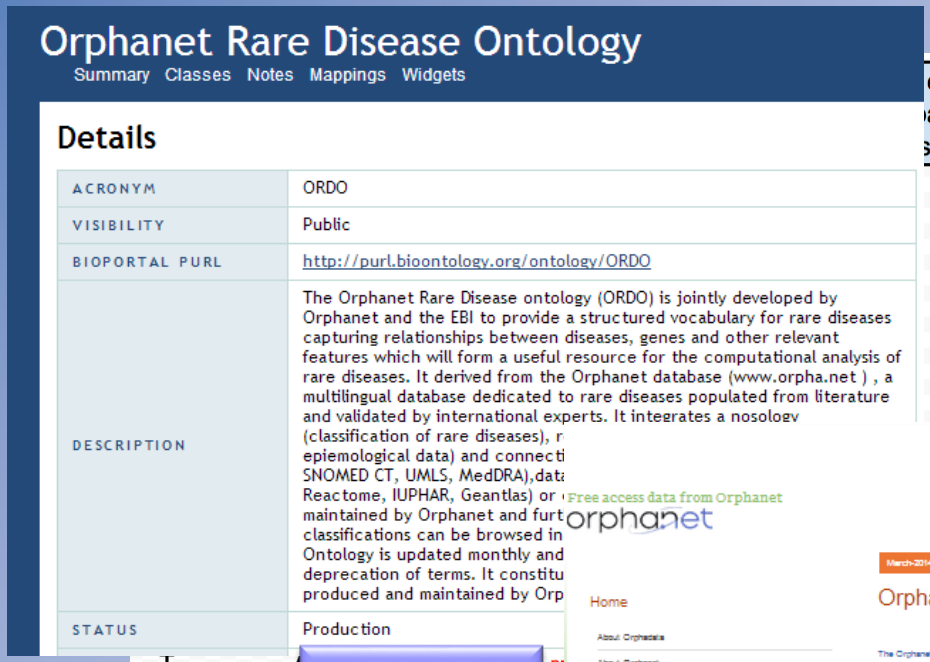
To find the Orphanet nomenclature



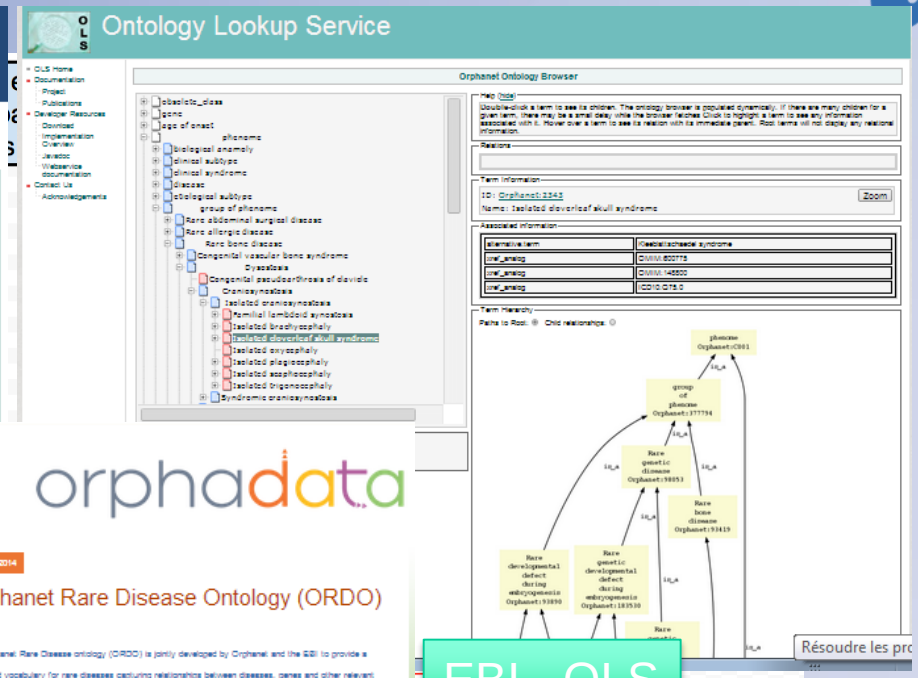
Alphabetical list of names and synonyms
with ORPHA code
Updated every 6 months
Translated.
PDF (download, print)

ORPHA nomenclature
Cross-references
Monthly updated
6 languages (7 soon)
For download
XML

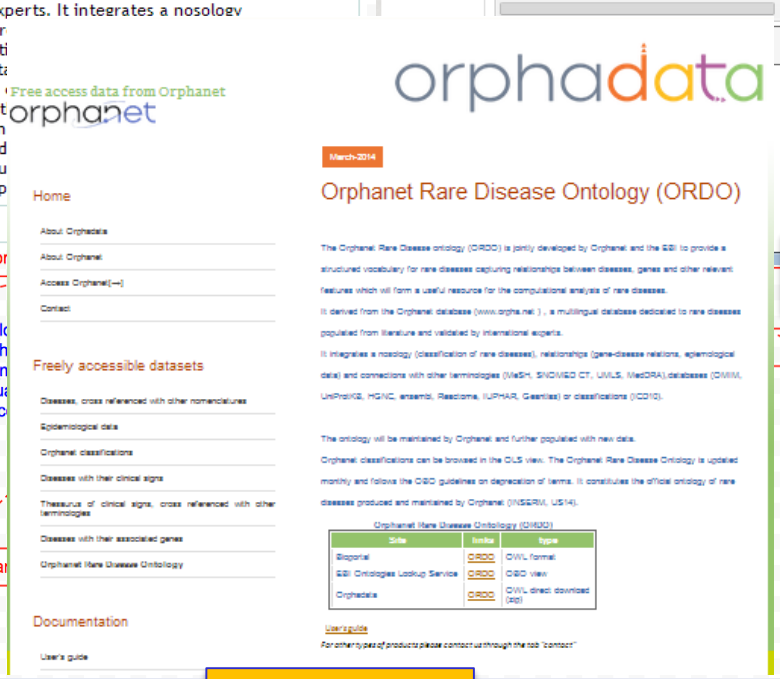
ORDO



BioPortal



EBI - OLS



Orphadata

The benefits of using ORPHA code

→ Identify rare diseases cases from health care sources

→ Connect data coming from health care to data coming from research

→ Promote international collaboration and data exchange

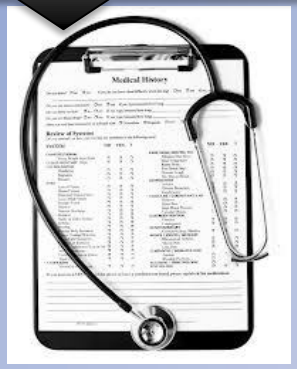
- The Orpha nomenclature provides a sound, structured, inter-operable resource for codification, and is the only nomenclature specific for rare diseases.
- ORPHA codes are never re-used
- The nomenclature and the structure are updated monthly
- Updates are provided in several IT formats to ease integration in different IT systems (xml, OWL, obo)
- **What's next?** To provide metadata to track changes between versions

Coding phenotypes

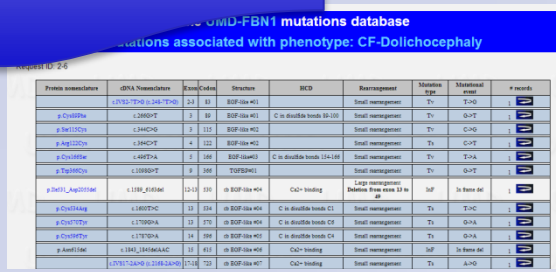
Using language: Do you mean?

scaphocephaly

Long narrow head

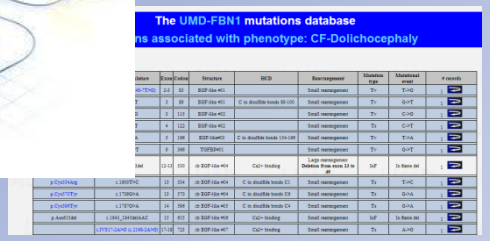


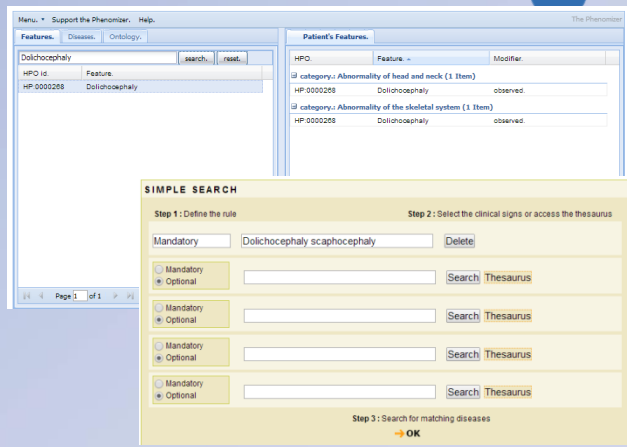
dolichocephaly



elementsofmorphology.nih.gov

Phenotypes:
Different resources, different terminologies





(e)HR:
SNOMED CT
ICD
Others?

Free text

Mutation/patient registries, databases:
HPO
LDDB
PhenoDB
Elements of morphology

Others? Free text?

Tools for diagnosis:

HPO
LDDB
Orphanet

Phenotype terminology project

- Aims:
  - **Map** commonly used clinical terminologies (Orphanet, LDDB, HPO, Elements of morphology, PhenoDB, UMLS, SNOMED-CT, MESH, MedDRA):
    - ✓ automatic map, expert validation, detection and correction of inconsistencies
  - **Find common terms** in the terminologies
  - Produce a **core terminology**
    - ✓ Common denominator allowing to share/exchange phenotypic data between databases
    - ✓ Mapped to every single terminology

Overview of project progress

- Sept 2012: start of mappings (Orphanet)
- EUGT2 – EUCERD workshop (Paris, September 2012)
  - Constitution of **the International Consortium of Human Phenotype Terminologies (ICHPT)**

PhenoDB (OMIM°

Orphanet thesaurus of signs and symptoms

HPOHuman Phneotype Ontology

LDDB
Elements of Morphology
POSSUM
SNOMED CT (IHTSDO)

DECIPHER
IRDiRC

- ICHPT workshop (ASHG, Boston, October 2013)
  - Selection of 2,300 core terms

First list of common terms

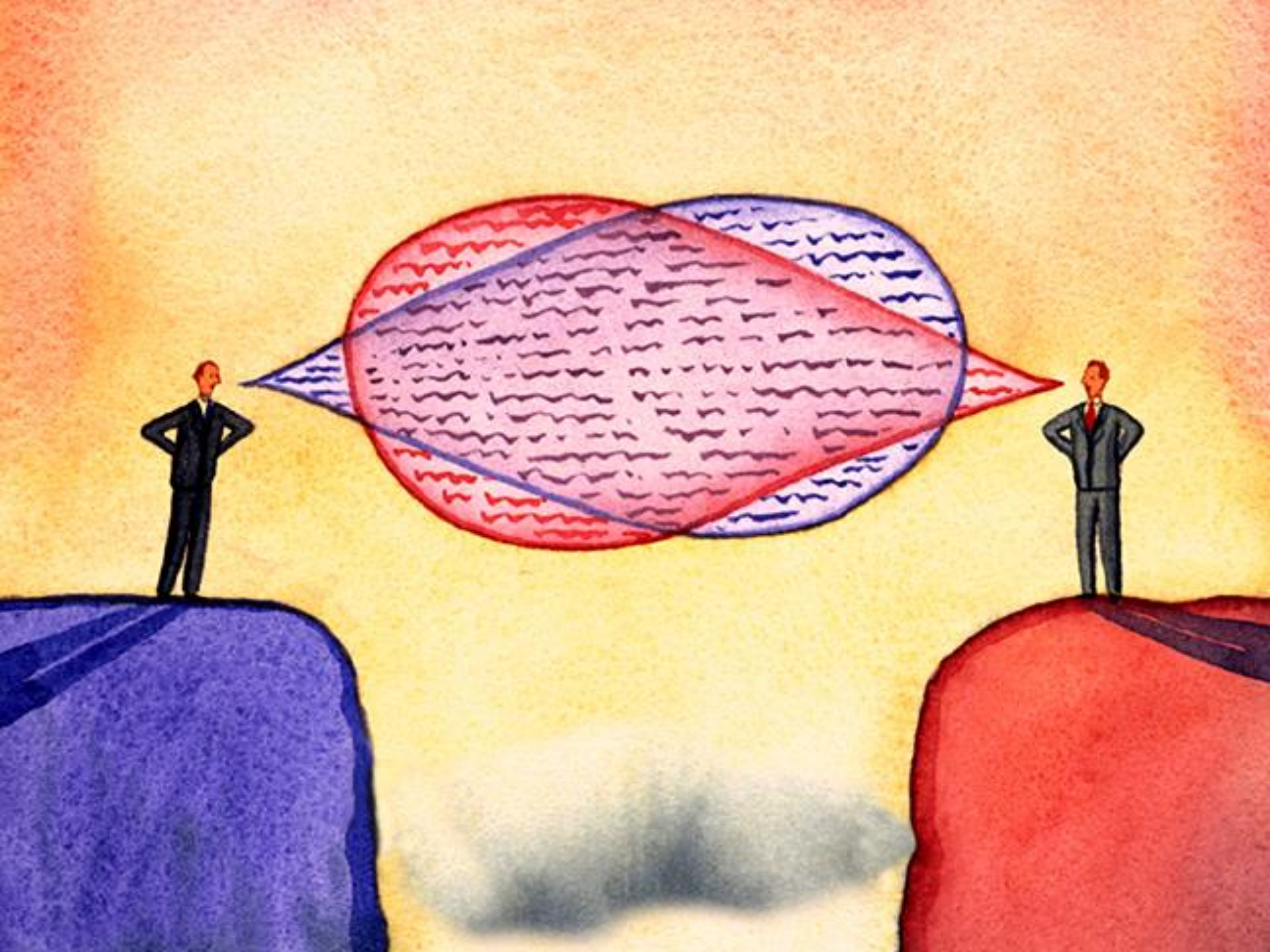
- Present in at least 3 terminologies
- Definition of rules for nomenclature
- Addition of terms present in each terminology as synonyms

	Identifier	Preferred term	Synonyms		
2	T0001	Anomaly of the skull	Skull anomalies	Cranial bones, general abnormalities	Cranium, gener
3	T0003	Cranial hyperostosis	Thick skull	Thickened skull	Dense skull
4	T0004	Basilar hyperostosis	Sclerosis of skull base		
5	T0005	Calvarial hyperostosis	Thick calvaria	Thickened calvaria	Dense calvaria
6	T0006	Decreased skull ossification	Poorly ossified skull	Ossification defects of skull	Undermineraliz
7	T0008	Decreased calvarial ossificatio	Thin calvaria	Absent ossification of calvaria	Thin calvarium
8	T0010	Anomaly of the cranial sutures	Cranial sutures, general abnormalit	Head Sutures anomalies	Abnormality of
9	T0011	Wide cranial sutures	Cranial sutures, wide	Wide cranial sutures (finding)	
10	T0012	Ridged cranial sutures	Cranial sutures, ridged		
11	T0013	Anomaly of the sella turcica	Sella turcica anomaly		
12	T0014	Large sella turcica	Sella turcica, large		
13	T0015	J-shaped sella turcica	Sella turcica, J-shaped	Shoe-shaped sella turcica	
14	T0016	Small sella turcica	Sella turcica, small		
15	T0018	Anomaly of the temporal bone	Abnormality of the temporal bone		
16	T0019	Anomaly of the mastoid proces	Mastoids, general abnormalities	Abnormality of the mastoid	
17	T0020	Small foramen magnum	Foramen magnum, small	Foramen magnum stenosis	
18	T0021	Large foramen magnum	Foramen magnum, large		
19	T0022	Delayed pneumatization of the	Delayed pneumatization of mastoids		
20	T0023	Advanced pneumatization of th	Advanced pneumatization of mastoids		

ICHPT

- Core phenotype terminology
- Common language between different vocabularies
- Completed with definitions
  - Elements of Morphology
  - HPO
  - Produced by the group
- Will be soon released in a dedicated website, hosted by
  - Visualisation
  - Download

IRDiRC
INTERNATIONAL
RARE DISEASES RESEARCH
CONSORTIUM



Thank you !