USEFULNESS OF ONTOLOGIES FOR RARE DISEASES

Manuel Posada, Verónica Alonso and Estrella López Martín

Institute of Rare Diseases Research (IIER-ISCIII)
Madrid, Spain

RD-Connect Workshop Data Linkage and Ontologies
ISS, Rome, 24-25th September 2015
☐ No conflict of interest is linked to this presentation
INTERNATIONAL DATA SHARING

Research enhancement

Better knowledge of RD

New therapies

Improvement of patients life’s quality
INTEROPERABILITY

“Capability of two or more systems or components to exchange information and to use the exchanged information”
BIOBANKS
(Sample data)

PATIENTS
REGISTRIES
(Clinical data)

INTEROPERABILITY
TOOLS

GUID
CDEs
SOPs
ELSI

ONTOLOGIES

REPOSITORIES
FOR –OMICS DATA
(Genomic & other -omics data)

DATA FLOW

DATA FLOW

DATA FLOW
...WHAT ARE THE ONTOLOGIES?
An ontology describes the concepts in some knowledge domain, some of their properties and how the concepts relate to each other.

Each concept (theoretical item) is represented by one unique reference (URI, real item), which is computer-readable. The references are associated to human-readable terms, definitions and synonyms.
Protein HDAC1 interacts with gene Parvb
or
La proteína HDAC1 interacciona con el gen Parvb
or
(different synonyms, languages, etc)

http://purl.uniprot.org/uniprot/Q13547
http://conceptwiki.org/index.php/Concept:e6559...
http://bio2rdf.org/geneid:29780
...What is the usefulness of ontologies in the rare diseases field?

- Every term has a unique reference provided by ontologies (standardization)
- Mappings specify the relation between datasets (linked data)

Integration of linkable data
CLASSIFICATION OF BIOMEDICAL ONTOLOGIES USEFUL IN THE RD FIELD

STRATEGY FOR THE SEARCHING OF ONTOLOGIES

It consisted of a multi-stage process:

**STEP 1**

The two leading repositories of biomedical ontologies, the Biportal and the OBO Foundry, were consulted for collecting a preliminary inventory of existing biomedical ontologies, which was composed by 377 ontologies.

**STEP 2**

After the revision of the inventory by two different researchers, a list formed by the selected biomedical ontologies was obtained. Our searching process was focused on phenotype ontologies, but we also included some other types of ontologies as support tools. This list contained 71 records.

**STEP 3**

Afterwards, the selected biomedical ontologies were classified into the following groups: phenotype, clinic, classifications & nomenclatures, disorder/disease-specific, anatomy, patients-related data, pharmacy, epidemiology, health indicators, miscellaneous and discarded.
CLASSIFICATION OF BIOMEDICAL ONTOLOGIES USEFUL IN THE RD FIELD

- PHENOTYPE
- CLINIC
- CLASSIFICATIONS AND NOMENCLATURES
- DISORDER/DISEASE-SPECIFIC ONTOLOGIES
- ANATOMY
- PATIENTS-RELATED DATA
- PHARMACY
- EPIDEMIOLOGY
- HEALTH INDICATORS
- OTHER
PHENOTYPE ONTOLOGIES USEFUL IN THE RD FIELD

- **Human Phenotype Ontology (HPO):** The HPO contains more than 10,000 terms and provides a structured and controlled vocabulary for the phenotypic features encountered in human hereditary and other diseases. The HPO itself does not describe individual disease entities but, rather, the phenotypic abnormalities associated with them.

- **Orphanet Ontology (ORDO):** The ORDO is an ontology of rare diseases whose content attempts disorders, subtypes, clinical signs and genes. It also includes epidemiology data related to rare diseases in Europe.

- In practical terms, HPO and ORDO cover different aspects and their use is perfectly complementary.
EXAMPLE 1

ATAXIA_USE_CASE_ISCIII

- Woman, 49 years old (age at diagnosis).
- Diagnosis: Spinocerebellar Ataxia Type 3 (SCA3), also known as Machado-Joseph Disease.
- Patient case registered in the Spanish National Rare Diseases Registry.
EXAMPLE 1

ATAXIA & FAMILIAL SPASTIC PARAPARESIS PATIENT REGISTRY

- Personal Information
- Family History
- Diagnosis-related Issues
- Phenotypic Traits
- Genetic Data
<table>
<thead>
<tr>
<th>PHENOTYPIC TERMS (REGISTRY)</th>
<th>HPO CODE</th>
<th>PHENOTYPIC TERMS (HPO)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cerebellar system affected</td>
<td>HP:0001317</td>
<td>Abnormality of the cerebellum</td>
</tr>
<tr>
<td>Pyramidal system affected</td>
<td>HP:0007256</td>
<td>Abnormality of pyramidal motor function</td>
</tr>
<tr>
<td>Brain stem affected</td>
<td>HP:0002363</td>
<td>Abnormality of brainstem morphology</td>
</tr>
<tr>
<td>Sensory central system affected</td>
<td>HP:0011730</td>
<td>Abnormality of central sensory function</td>
</tr>
<tr>
<td>Polyneuropathy</td>
<td>HP:0001271</td>
<td>Polyneuropathy</td>
</tr>
<tr>
<td>Demyelinating polyneuropathy</td>
<td>HP:0011402</td>
<td>Demyelinating sensory neuropathy</td>
</tr>
<tr>
<td>Axonal polyneuropathy</td>
<td>HP:0003390</td>
<td>Sensory axonal neuropathy</td>
</tr>
<tr>
<td>Mixed polyneuropathy</td>
<td>HP:0007002</td>
<td>Motor axonal neuropathy</td>
</tr>
<tr>
<td>Second motor neuron signs</td>
<td>HP:0007327</td>
<td>Mixed demyelinating and axonal polyneuropathy</td>
</tr>
<tr>
<td>Extrapyramidal system affected</td>
<td>HP:0002071</td>
<td>Abnormality of extrapyramidal motor function</td>
</tr>
<tr>
<td>Visual system affected</td>
<td>HP:0000504</td>
<td>Abnormality of vision</td>
</tr>
<tr>
<td>Hearing system affected</td>
<td>HP:0000364</td>
<td>Hearing abnormality</td>
</tr>
<tr>
<td>Cognitive/Behavioural impairment</td>
<td>HP:0100543</td>
<td>Cognitive impairment</td>
</tr>
<tr>
<td>Epilepsia</td>
<td>HP:0011097</td>
<td>Epileptic spasms</td>
</tr>
<tr>
<td>Walk</td>
<td>HP:0002355</td>
<td>Difficulty walking</td>
</tr>
<tr>
<td>Autonomous walk</td>
<td>Not found</td>
<td>Not found</td>
</tr>
<tr>
<td>Use of walking stick</td>
<td>Not found</td>
<td>Not found</td>
</tr>
<tr>
<td>Use of wheelchair</td>
<td>Not found</td>
<td>Not found</td>
</tr>
<tr>
<td>Bedridden</td>
<td>Not found</td>
<td>Not found</td>
</tr>
<tr>
<td>Oculomotor signs</td>
<td>HP:0006860</td>
<td>Abnormality of eye movements</td>
</tr>
<tr>
<td>Other signs (free text)</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
EXAMPLE 2

Five years old child (non diagnosed case) with the following phenotype:

- Hypothyroidism
- Hyperpigmentation
- Intellectual disability
- Delayed speech and language development
- Delayed motor development
- Short stature
- Microcephaly
- Microretrognathia
- Trigonocephaly
- Macrotia
- Low IGF-1 blood levels
- Low Fe²⁺ blood levels
SpainRDR: Issues regarding the use of ontologies in registry data curation
SpainRDR: Issues regarding the use of ontologies in registry data curation

- Limiting factors for the integration of phenotype ontologies into the RD registries architecture: **Practical examples** extracted from a real case of the IIER Undiagnosed Disease Program (IIER-UDP)

**Five years old child with the following phenotype:**
- Hypothyroidism
- Intellectual disability
- Delayed speech and language development
- Delayed motor development
- Short stature
- Hyperpigmentation
- Microcephaly
- Microporophthalmia
- Trigonocephaly
- Macrotia
- Low IGF-1 blood levels
- Low Fe^{2+} blood levels

**Temporal evolution of a phenotype (limiting factor)**

**Accuracy in descriptions of phenotypic features (limiting factor)**
Useful Resources
Undiagnosed Cases Program – SpainUDP, IIER: Software to help to the diagnosis

Phenolyzer

Others...

PhenomeCentral

Others...

PhenoTIPS

Phenomizer
Undiagnosed Cases Program
SpainUDP - IIER

EQUIPO MULTIDISCIPLINAR
- Clínica
- Genética
- Biología celular y molecular
- Anomalías congénitas
- Bioinformática

SOFTWARES DE AYUDA AL DIAGNÓSTICO
Present vs Absent Features

**CUTANEOUS**
- Hyperpigmentation of the skin
- Generalized hyperpigmentation
- Irregular hyperpigmentation
- Hypopigmentation of the skin
- Capillary hemangiomas
- Vascular skin abnormality
- Nevus
- Absent skin pigmentation
- Deep-set nails
- Thin nail
- Woolly hair
- Concave nail
- Fragile nails
- Hyperconvex thumb nails
- Partial albinism
- Lumbosacral hirsutism
- Lack of skin elasticity
- Piebaldism
- Hyperpigmentation in sun-exposed areas
- Epidermal nevus
- Ichthyosis
- Abnormal blistering of the skin
- Progressive hyperpigmentation
- Generalized hypopigmentation

**CARDIOVASCULAR**
- Defect in the atrial septum
- Ventricular septal defect
- Complete atrioventricular canal defect
- Coartation of aorta
- Tetralogy of Fallot
- Cardiomyopathy
- Arrhythmia

**RESPIRATORY**
- Congenital diaphragmatic hernia
- Abnormality of the lung

**MUSCULOSKELETAL**
- Skeletal dysplasia
- Increased susceptibility to fractures
- Lower limb undergrowth
- Upper limb undergrowth
- Camptodactyly
- Finger
- Toe

Quick phenotype search:
Enter keywords and choose from the suggested ontology terms.
Use of tools for helping in the diagnostic process.

**CLINICAL SITE (RESEARCH)**

- Patient (case)
- CLINICIAN
- **a)** Electronic Medical Record (EMR)
- **b)** Clinical Record (paper, non-electronic document)
- HISTORICAL CLINICAL INFORMATION
- PROSPECTIVE CLINICAL INFORMATION

**REGISTRY SITE**

- REGISTERING ACTIVITIES
- RESEARCH ACTIVITIES
- DATA STORING
- CDEs
  - Standardization
  - Mappings
  - Abstraction
- **a)** Undiagnosed cases
- **b)** Confirmed diagnosis
- Undiagnosed cases program
- **b1)** Sub-phenotypes identifiable in literature
- **b2)** Sub-phenotypes not so well defined yet

**DATA COLLECTION**

- QC
SOPs

ELSI

Phenotype ontologies

- Standardized
- Linkable

CDEs

SOPs

- Harmonized

GUID

ELSI

- Universally identified

PHENOTYPIC DATA

PATIENTS REGISTRY

Integration of standardized, linkable, harmonized and universally identified PHENOTYPIC DATA with genomic and biosamples data
Thank you for your attention