Harmonising phenomics information for a better interoperability in the rare disease field

Sylvie Maiella[1], Annie Olry[1], Valérie Lanneau[1], Bruno Donadille[1], Halima Lourghi[1], Sebastian Köhler[2], Tudor Groza[5], Charlotte Rodwell[1], Simon Jupp[4], Peter N Robinson[6], Michael Brudno[3], Marc Hanauer[1], Helen Parkinson[4], Ana Rath[1].

1INSERM, US14 – Orphanet, Paris, France; 2. Institute for Medical Genetics and Human Genetics, Charité Universitätsmedizin Berlin, Germany;
3. The Hospital for Sick Children, Toronto, Canada; 4. European Bioinformatics Institute, European Molecular Biology Laboratory (EMBL-EBI), Hinxton, United Kingdom;
5. Garvan Institute of Medical Research, Darlinghurst, Sydney, Australia 6. The Jackson Laboratory, USA

Purpose.

Rare disease (RD) research is a field of medicine increasingly reliant on information technology, with the advent of low-cost whole-genome sequencing revolutionising the discovery of genetic causes of disorders. Detailed phenotypic data, combined with genomic data, have an enormous potential to accelerate the identification of clinically actionable prognostic or therapeutic implications and to improve our understanding of RD. The harmonisation of phenomics information, including disorders and phenotypic traits that are stored in different contexts in a non-standardised way, is a cornerstone for producing sound data to foster research.

HIPBI-RD («Harmonising phenomics information for a better interoperability in the rare disease field ») is a three-year project starting in 2016 funded via the E-Rare 3 ERA-NET. The project aims to provide the community with an integrated, RD-specific bio-informatics ecosystem that will harmonise the way phenomics information is stored in databases and patient files worldwide, and thereby contribute to interoperability. This ecosystem will consist of a suite of tools and ontologies, optimised to work together, made available through commonly used software repositories (Fig.1).

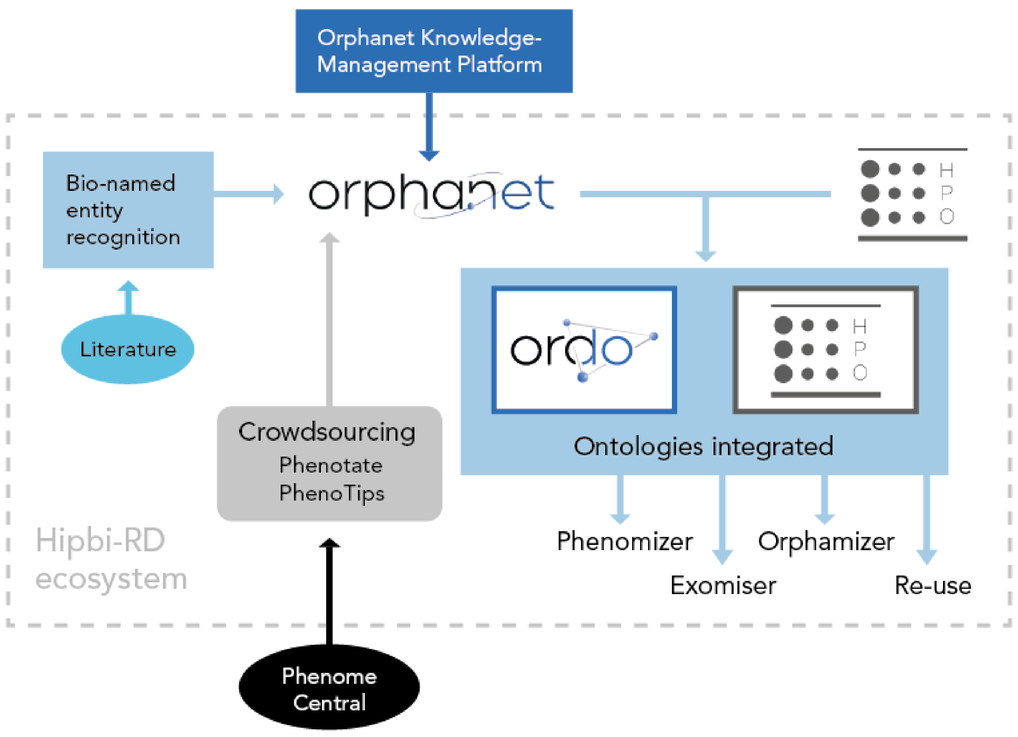

Figure 1 The proposed Hipbi-RD ecosystem: a suite of ontologies and tools optimized to work together.

Methods

This project builds on three resources largely adopted by the RD community

- Orphanet
- The Orphanet Rare Disease Ontology (ORDO)
- The Human Phenotype Ontology (HPO)
- PhenoTips and PhenomeCentral

Results

Rare disorders listed in Orphanet are being annotated with HPO phenotypic traits. The alignment is characterized by frequency (obligate, very frequent, frequent, occasional, very rare or excluded) and whether the annotated HPO term is a major diagnostic criterion or a pathognomonic sign of the rare disease. As of April 2018, 3,082 Orphanet disorder entities have been annotated with 5,487 relevant HPO terms together with information on frequency of occurrence of each of these phenotypic features. Fig. 2). This represents 49% coverage of current Orphanet disorder entities. The creation of an HPO annotation (total of 64,661 annotations are available) needs medical expertise, and is based on a literature review and consultation with experts on each RD.

Hoom HPO-ORDO Ontological module HOOM is a module that qualifies the annotation between a clinical entity and phenotypic abnormalities according to a frequency and by integrating the notion of diagnostic criterion. In ORDO a clinical entity is either a group of rare disorders, a rare disorder or a subtype of disorder. The "phenomes" branch of ORDO has been refactored as a logical import of HPO, and the HPO-ORDO phenotype disease-annotations have been provided in a series of triples in OBAN format in which associations, frequency and provenance are modeled.

HOOM is provided as an OWL (Ontologies Web Languages) file, using OBAN, the Orphanet Rare Disease Ontology (ORDO), and HPO ontological models.

Figure 2 Disorders in Orphanet annotated with HPO terms as of April 2018

interconnection of ORDO and HPO with HOOM

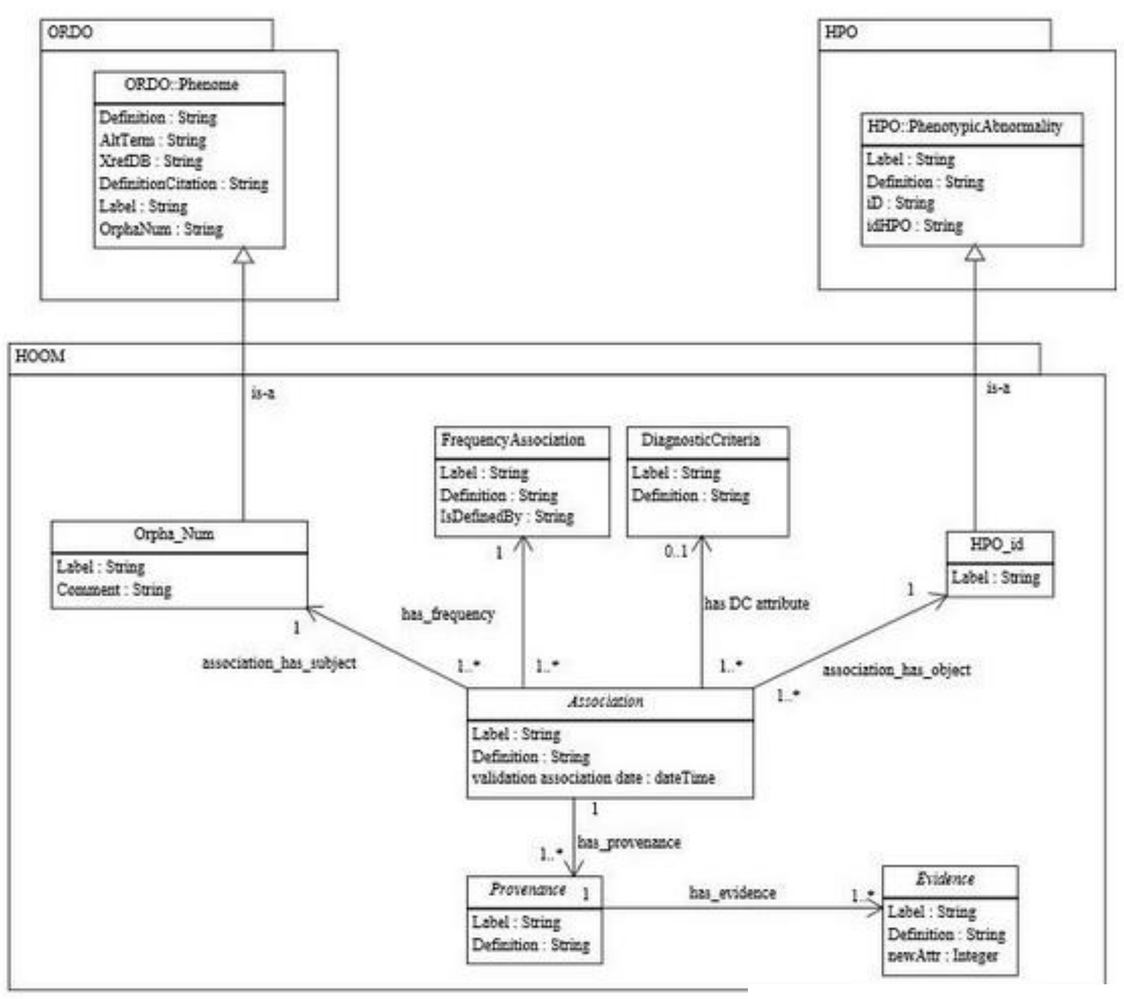

Figure 3 HOOM: HPO-ORDO ontological module

Conclusions

HOOM provides extra possibilities for researchers, pharmaceutical companies and others wishing to co-analyse rare and common disease phenotype associations, or re-use the integrated ontologies in genomic variants repositories or match-making tools. The HIPBI-RD ecosystem, once finalized, will provide a resource that will contribute to bridging genome-scale biology and a disease-centered view on human pathobiology, as harmonization of phenotypic data will contribute to the interpretation of variants identified through exome and full genome sequencing thus improving diagnostics and the delineation of RDs, as well as contributing to solving unsolved cases.

HIPBI-RD PROJECT HAS RECEIVED FUNDING FROM THE EUROPEAN UNION'S HORIZON 2020 RESEARCH AND INNOVATION PROGRAM UNDER THE ERA-NET COFUND ACTION N° 643578. IT WAS SUPPORTED BY ANR, UNDER THE FRAME OF E-RARE-3, THE ERA-NET FOR RESEARCH ON RARE DISEASES