EUCERD/Eurogentest Workshop
on
Cross-Referencing of Terminologies

Summary Report

SUMMARY OF THE EUCERD/EUROGENTEST WORKSHOP ON THE CROSS-REFERENCING OF TERMINOLOGIES

This workshop was organised in the scope of Work Package 5 of the EUCERD Joint Action (N° 2011 22 01) and Eurogentest2, a Coordination Action funded under the European Commission's 7th Framework Programme.

The **objective** of the workshop was to provide the community with standards to be used in order to achieve interoperability between databases, for two main purposes: making rare diseases visible in health information systems and proposing standards to the clinical research community for the description of phenotypes. This is crucial to interpret genomic rearrangements as well as future high-throughput sequence data.

The **first day** was dedicated to discussing the classification systems suitable for coding rare diseases diagnosis. The classifications of diseases which were considered were ICD11 (currently in Beta phase), the Orphanet classification, OMIM and SNOMED CT. The draft of ICD11 which is currently available indicates that many RD are likely to be incorporated into ICD11 but probably not all of them, that the structure of the classification is not satisfactory, that the revision process in not well enough defined so as to ensure that experts' views are taken into account and that a permanent update process will keep ICD11 up to date. The SNOMED CT disease terminology is intended to be used in electronic health records to code the health status of patients. It is the most comprehensive terminology in the world. It does not focus on RD. The Orphanet polyhierarchy classification system is entirely dedicated to RD and benefits from the contribution of many experts around the world. It is appropriately funded to ensure continuity. OMIM is the standard coding system for genetic phenotypes widely used for that purpose. The four classification systems have been cross-referenced by the Orphanet team as to provide the community with tools to navigate from one to another. The mapping outcome is visible on the Orphanet website and available for extraction on www.orphadata.org.

The expert group proposed:
1- Continue impacting on ICD11 as much as possible but with limited hope that the new version will meet the needs of the RD community;
2- Set up an active collaboration with SNOMED CT as to ensure that missing codes are considered for incorporation, considering that SNOMED CT is on the track to become a de facto standard terminology for electronic health information;
3- Recommend that Orphanet and OMIM codes are to be accepted as the standards of the rare disease community which means that any database for RDs should have either Orphanet codes or MIM numbers or both. The scientific community has been using MIM codes for a very long time and will continue to need the "splitter" perspective OMIM offers. But using the Orphanet nosology for the structure, which OMIM does not offer in this form, is really useful for clinicians and ontologists. To work with a combination of both is ideal.
4- To continue cross-referencing OMIM and Orpha codes with the standard terminologies (ICD and SNOMED CT), as it is a quality-control exercise for all of these terminologies and as it is necessary for navigation from one classification to another.

The **second day** was dedicated to the relevant terminologies in use by different communities to describe phenomes. They were presented and their strengths and weaknesses discussed. The terminologies which were considered were PhenoDB (2846 terms), London Dysmorphology Database (LDDB; 1318 terms), Orphanet (1243 terms), Human Phenotype Ontology (9895 terms, 22/08/2012), Elements of Morphology (AJMG; 423 terms), ICD10 (1230 terms), as well as medical terminologies in use: Unified Medical Language System Metathesaurus (UMLS; 7,957,179 distinct concept terms), Systematised Nomenclature of Medicine -- Clinical Terms (SNOMED CT ; >311,000 concepts), Medical Subject Headings (MeSH ; 26,853 concepts) and Medical Dictionary for Regulatory Activities (MedDRA, 69,389 concepts). The Orphanet team established a strategy to compare them to find commonalities and differences, using ONAGUI as a tool to pick up exact matches. The non-exact matches were verified manually by an expert. Some results were presented in this document to prepare the discussion between experts. Two strategies were discussed: either to build a consensual terminology using the concepts shared by existing terminologies to describe phenotypes, or publish the existing terminologies with their cross-referencing, as it is done between disease terminologies to ease the interoperability between databases without disturbing the habits of the different groups of users.

After discussion it was agreed on that:

1- Given the multitude of needs and applications in the field of rare diseases, it is not currently realistic or even desirable to have one terminology for all applications. Prominent terminologies have different focuses and user bases. The Orphanet thesaurus of signs and symptoms is intended for use by clinicians who are not necessarily geneticists. The Human Phenotype Ontology has been developed to enable computational analysis of human disease manifestations. PhenoDB is intended to enable quick entry of phenotypic features by clinicians (or health care providers). The Elements of Morphology is a glossary of state of the art definitions for phenotypic features. Generalist systems such as the ICD and SNOMED CT do not currently comprise many of these features, meaning that they are coded, if at all, using more general concepts;

2- The expert group will agree upon a core set of about 2 000 terms that represent the major phenotypic abnormalities encountered in persons with rare diseases which will be cross-matched with the available terminologies. This core set of terms will be recommended for use in any new information system intended to collect phenotypic data, either for research or clinical purpose. They will be published together with definitions;

3- The core set of phenotypic terms will be set up by comparing the different terminologies, considering that terms used by the majority of them are likely to constitute the candidates for standard terms. The Orphanet team will carry out the preparatory work and the expert group will act as reviewers and decision makers to ensure that there is a good coverage of all body systems for which descriptors are needed;

4- This set of terms will be proposed for inclusion in SNOMED CT and ICD-11;

5- As there is a need to continuously revise the proposal, the expert group proposes to set up an International Consortium of Phenotype Terminologies. Therefore, the core set of terms will be named ICPT codes.