Modular Extensions to the ChEBI Ontology

Janna Hastings*, Paula de Matos, Adriano Dekker, Marcus Ennis, Venkatesh
Muthukrishnan, Steve Turner, Gareth Owen and Christoph Steinbeck

Cheminformatics and Metabolism, European Bioinformatics Institute, Hinxton, UK

ABSTRACT

ChEBI is an ontology of chemical entities of biological interest. It includes a structure-based and a role-based classification of chemical entities including molecules, ions and salts. This poster showcases recent developments and forthcoming enhancements to the ontology, focusing on modular extensions of the content. These extensions are being developed in support of error detection and consistency checking and to improve the capability for automated reasoning about the biological context of chemicals.

1 INTRODUCTION

ChEBI is an ontology of chemical entities of biological interest (de Matos *et al.*, 2010). It includes a structure-based and a role-based classification of chemical entities, such as molecules, ions and salts. Within the structure-based classification, chemical entities are organised according to shared structural features, such as skeletons and functional groups. In addition to taxonomic classification, structure-based relationships capture biologically relevant associations between compounds, such as tautomers, enantiomers and conjugate bases and acids. Within the role classification, the various ways that chemical entities can be active in biological and chemical contexts are organised, such as by inhibiting the activity of enzymes, acting as hormones, being used as insecticides, and being used as solvents. Chemical entities are linked to roles with the *has role* relationship.

ChEBI includes just over 28,000 annotated entities as of the May 2012 release. The ontology is available for browsing at `http://www.ebi.ac.uk/chebi/` and is fully downloadable in several formats including OBO and OWL.

2 RECENT DEVELOPMENTS

ChEBI recently completed a large annotation push to make the ontology fully *is-a complete*, that is, to ensure that each entity has at least one taxonomic parent. (Previously, due to historical legacy, there were some entities that were only classified using structural relationships such as *has functional parent*.) Furthermore, ChEBI is now aligned with BFO, as described in (Hastings *et al.*, 2011) and work is well underway in alignment with the Gene Ontology (The Gene Ontology Consortium, 2000). Another recent development is the provision of a new graph-based visualisation of the ontology.

3 EXTENSIONS FOR ERROR DETECTION

ChEBI is being enhanced with modular extensions to the ontology in support of error detection and consistency checking. These extensions are being annotated in separate OWL module files which extend the main ontology file that is generated on a monthly basis

*To whom correspondence should be addressed: hastings@ebi.ac.uk

from the core database content. The mapping to BFO is made available as one such modular extension, `chebi-in-bfo.owl`.

Another extension, available as `chebi-disjoints.owl`, includes axioms about disjointness between entities in the ontology. Many of the classes in the chemical entity ontology are *not* disjoint, because chemical classification is *compositional* (Hastings *et al.*, 2012). Annotating those pairs of classes that are disjoint thus allows for better error detection, as a reasoner is able to raise an error if anything is accidentally classified beneath both of the two classes that should be disjoint. In an ontology the size of ChEBI, such automated assistance with the detection of errors is essential. For example, *group* (CHEBI:24433) is disjoint from *molecular entity* (CHEBI:23367). Introducing this axiom allowed us to detect and correct an incorrect classification of the ion *3-D-glucuronosyl-N2,6-disulfonato-beta-D-glucosamine(3-)* (CHEBI:58150) as a group.

4 THE BIOLOGICAL CONTEXT OF CHEMICALS

ChEBI is also being extended in order to explicitly represent the biological context of the chemicals as semantic relationships to biological entities in external ontologies, as described in (Batchelor *et al.*, 2010). One example of such is the explicit annotation of *species* and *tissue* for metabolites (natural products). Another is the explicit annotation of *diseases* in the case of drugs used to treat the diseases and in the case of metabolites that are markers for certain diseases. Cross-references and annotations of ChEBI identifiers in biological databases will be used as sources for this effort, which will then be checked by ChEBI curators.

Relationships from the RO and other ontologies from the OBO Foundry effort will be used in all cases.

REFERENCES

Batchelor, C., Hastings, J., and Steinbeck, C. (2010). Ontological dependence, dispositions and institutional reality in chemistry. In A. Galton and R. Mizoguchi, editors, *Proceedings of the 6th Formal Ontology in Information Systems conference*, Toronto, Canada.

de Matos, P., Alcántara, R., Dekker, A., Ennis, M., Hastings, J., Haug, K., Spiteri, I., Turner, S., and Steinbeck, C. (2010). Chemical Entities of Biological Interest: an update. *Nucleic Acids Research*, **38**, D249–D254.

Hastings, J., de Matos, P., Dekker, A., Ennis, M., Haug, K., Josephs, Z., Owen, G., Turner, S., and Steinbeck, C. (2011). Recent developments in the ChEBI ontology. In *Proceedings of the International Conference on Biomedical Ontology (ICBO), CEUR-WS volume 833*.

Hastings, J., Magka, D., Batchelor, C., Duan, L., Stevens, R., Ennis, M., and Steinbeck, C. (2012). Structure-based classification and ontology in chemistry. *Journal of Cheminformatics*, **4**(1), 8.

The Gene Ontology Consortium (2000). Gene ontology: tool for the unification of biology. *Nat. Genet.*, **25**, 25–9.