Standardising biological trait representation with the Ontology of Biological Attributes (OBA) – Abstract

Arwa Ibrahim[1,*], Ray Stefancsik[1], Vinicius de Souza[1], Anita R. Caron[1], Nicolas Matentzoglu[2], James McLaughlin[1] and Zoë M. Pendlington[1]

[1]European Bioinformatics Institute (EMBL-EBI), Hinxton, Cambridgeshire CB10 1SD, UK
[2]Semanticly, Athens, Greece

Abstract

The Ontology of Biological Attributes (OBA) is a standardised, species-independent, computationally-amenable representation of phenotype traits relevant to biological, clinical, environmental and life science applications. While many current phenotype ontologies describe abnormal phenotypes relative to some wild-type state, OBA represents neutral traits or attributes. OBA has a central data integration role and employs Dead Simple Ontology Design Patterns (DOS-DPs) templates for the automated classification of traits, drawing on terms from domain-specific reference ontologies in a post-compositional approach. This computation based on logical inferences and automated reasoners promotes interoperability as multiple links to these ontologies enables powerful querying, knowledge graph integration and inference. In addition, due to its rich axiomatisation, a central OBA use case is to provide structure to weakly axiomatised ontologies like the Vertebrate Ontology (VT) and the Experimental Factor Ontology (EFO). The logical axioms in OBA provide a previously missing bridge that can link Mendelian phenotypes with genome-wide association studies (GWAS) and measurable traits. The GWAS Catalog incorporates EFO in its curation workflows and there is an ongoing OBA mapping effort of the EFO measurement branch. The Open Targets Platform, which identifies potential therapeutic drug targets, is a major consumer of GWAS data, so OBA classes imported into EFO can facilitate computational drug target identification. Also, there is an increasing demand from the GWAS Catalog for new trait terms to annotate summary statistics sometimes containing hundreds or thousands of GWAS studies. The template-based curation workflow of OBA means that the addition of new terms is highly scalable which is important in the context of the generation of large amounts of scientific data, reducing the need for manual curation and eliminating human error. OBA is a component of the Unified Phenotype Ontology (uPheno 2) and this integration allows for the grouping of phenotypic effects across biological attributes. It is an Open Biological and Biomedical Ontologies (OBO) Foundry compliant ontology and uses the Ontology Development Kit (ODK) to run releases, quality control checks and external dependencies. It currently uses eleven DOS-DP templates that cover many anatomical, metabolite and cellular attributes needed for data integration with the possibility to expand based on stakeholder and community use cases.

Keywords

OBO ontologies, traits, attributes, phenotypes, ontology design patterns, DOS-DP, data integration, interoperability, mappings

CEUR
Workshop
Proceedings

ceur-ws.org
ISSN 1613-0073

CEUR-WS.org/Vol-3939/abstract3.pdf