Challenges Arising from Ontology Imports Utilized for Exploring Mechanisms of Disease

J. Allen Baron [1] and Lynn M. Schriml [1]

[1] *Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, Maryland, USA*

Abstract

The Human Disease Ontology (DO) seeks to describe the breadth and complexity of human disease and to provide a stable framework for advanced analysis. It accomplishes these goals through automated import of non-disease ontologies and the definition of logical axioms. Exploration of diseases via these imports and axioms is available by browsing DO's OWL tree or through advanced searches at DO's website, https://www.disease-ontology.org, or by downloading the doid.owl file from DO's GitHub repository. Utilizing other ontologies to define disease mechanisms contributes to DO's interoperability but has also presented challenges. Multiple examples of challenges faced while curating the DO and the approaches used to mitigate them are outlined here. These address common problems in ontology curation including ontology size, differing scopes/philosophies, unexpected changes, and ultimately the real-world difficulty of defining (medical) terms.

Keywords

disease, disease ontology, interoperability, imports, integration challenges

1. Introduction

The Human Disease Ontology (DO, https://www.disease-ontology.org) [1] encompasses the breadth of diseases, rare and common, and innovates interoperability between the DO and other ontologies through the integration of imports of non-disease ontology terms to define connections between diseases, defined by logical axioms (**Figure 1**). For example, 'has phenotype' some 'Abnormal immunoglobulin level' retrieves 25 autoimmune-related disease terms. These logical assertions connect diseases by shared attributes and thus allow viewing inferred relationships between diseases and building a more complex multi-parental disease-to-disease classification. Axioms allow for a broader understanding of disease mechanisms. Utilizing ontology imports enables exploration of disease terms through their related cell of origin, anatomical location, symptoms, phenotypes, and genetic or environmental risk factors. DO import files are generated with each DO release using ROBOT extract (http://robot.obolibrary.org/extract) [2], which syncs the import file with their source ontologies. To date, DO integrates 14 import files: anatomy, cell types, chemicals, clinical modifier (onset), disease drivers, evidence codes, food material, inheritance pattern, omim_susceptibility (genetic risk factors), phenotypes, sequence structural and functional variants, symptoms, taxonomy, and transmission methods (https://disease-ontology.org/resources/DO_Imports).

The exploration of disease mechanisms via imports using features available on DO's website and the challenges faced while integrating other ontologies due to scope, size, and unexpected changes are outlined herein.

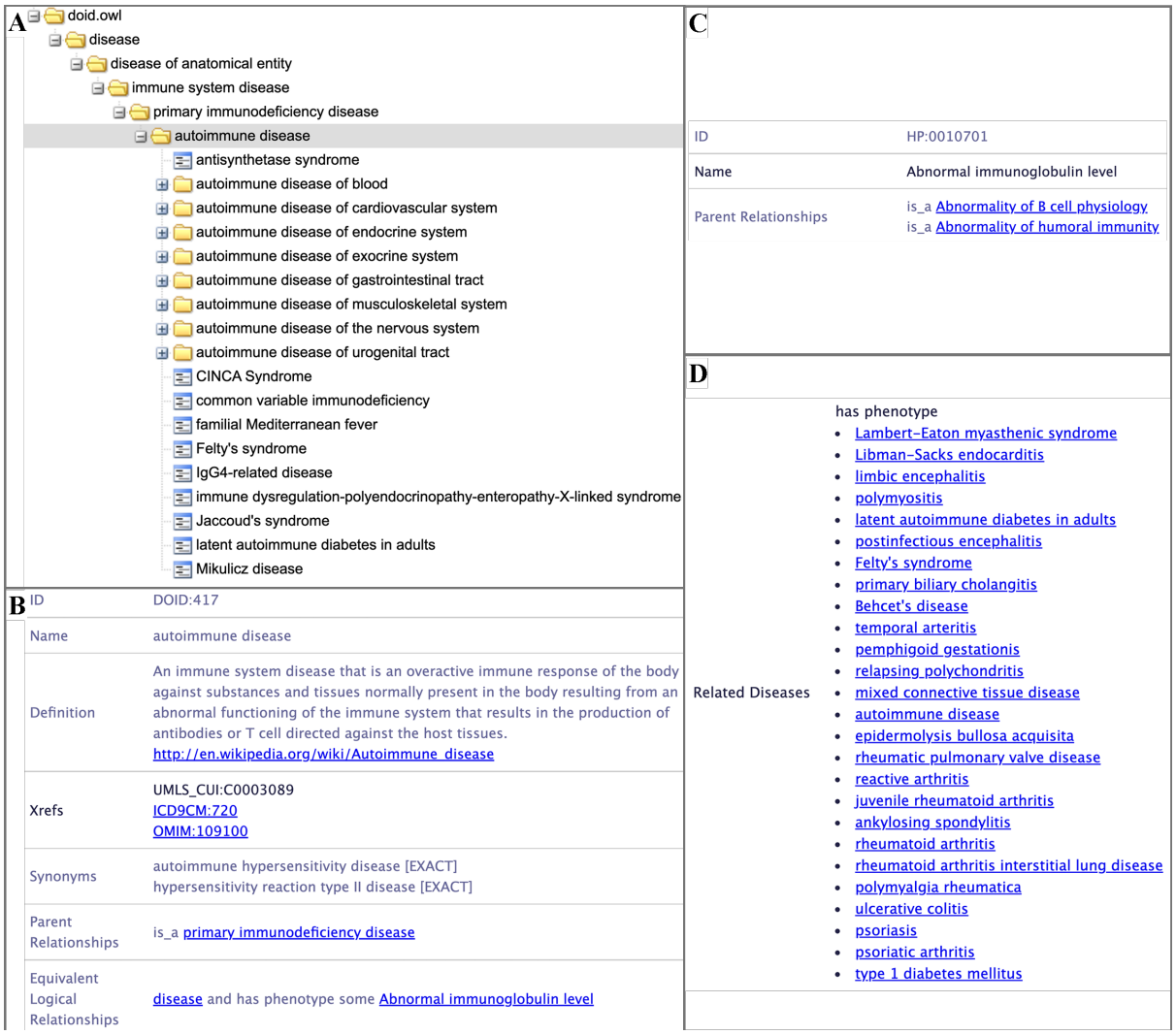

Figure 1: Human Disease Ontology logical axioms. (A) DO's OWL tree, doid.owl, hierarchical view (B) autoimmune disease (DOID:417) defined by an Equivalent Axiom (C) HPO term - DO's phenotype import (D) DO diseases defined by 'has phenotype' some 'Abnormal immunoglobulin level' logical definition.

2. Exploring Disease Mechanisms

The DO's OBO and OWL trees provide a multi-parentage view of disease classification. DO's OWL tree enables exploration of disease via DO's imports, thus enabling exploration of disease through related biomedical information (e.g., age of onset, anatomical location, cell type of origin); environmental driver (infectious agents, chemicals, exposures, food allergies); genetic drivers (susceptibility traits, sequence variants); and clinical variants (phenotypes and symptoms). These novel views of disease provide a unique perspective, such as exploring related diseases or disease mechanisms through the lens of common disease drivers (genetic, biological, chemical or ecological), phenotypes, symptoms, mode of inheritance, or infectious agent.

Explore DO's axioms through the Advanced Search option on the DO's website (https://www.disease-ontology.org). Explore disease metadata by queries for Name (disease name), Synonym, Definition, Subset (e.g., DO_cancer_slim or DO_AGR_slim), DOID, Alternate ID, Xrefs (clinical vocabularies: OMIM, NCIthesaurus, ICD, SNOMED, GARD, Orphanet [3–8]) or Relation. The recently added Relation drop

down options support querying of DO's axioms via RO relation terms: "adjacent to", "derives from", "disease has basis in", "has allergic trigger", "has material basis in", "has origin", "has symptom", "has phenotype", "disease has location" or "transmitted by".

3. Import Challenges

Most practical challenges to integration arise from the scope of 'import ontologies'. For example, the Human Phenotype Ontology (HPO) [9] classifies phenotypic features, including some disease entities, with a disease being defined as a phenotypic feature when it is a feature of another disease. Full integration of these 'disease phenotypes' would result in undesirable duplication of disease terms within the DO. To mitigate this problem, the import of HPO's 'Phenotypic Abnormality' branch explicitly includes only primary phenotypic features in the import by excluding disease entities with a custom 'exclude' slim (HPO term list).

The taxonomic breadth of an ontology, such as the cross-species anatomy ontology UBERON [10], also challenges integration to the human-specific disease ontology. For example, terms in UBERON classifying anatomical parts may be labelled with a human or non-human name, such as 'pes' or 'manus'. Pes is defined as the "zoological term for the distal portion of the hind limb of tetrapod animals", which is synonymous with the human anatomical term "foot". Fortunately, the human anatomical terms are synonyms of the non-human anatomical terms and updating the label while retaining the ID in the uberon_import.owl file is a simple solution for this issue.

Scope has also proved to be a challenge at the intersection of ontologies. Recent work to improve the DO has focused on the addition of terms to define environmental drivers of disease. These terms add the opportunity for greater disease inference, in much the same way as cell type, phenotype, and symptoms, but they are not diseases and are outside the scope of the DO. These new terms are types of environmental stressors, which is a concept captured by the Exposure Ontology (ExO) [11] and adopted by the Environmental Conditions, Treatments and Exposures Ontology (ECTO) [12]. However, as the list of potential stressors could be unending it is beyond the scope of ExO and ECTO to include each stressor; they aim to represent upper-level terms only. Therefore, the new DISDRIV (Disease Drivers) import has been created in collaboration with ExO and ECTO to fulfill the need for specific, leaf-node terms and submitted to the OBO Foundry as a new application ontology.

Along with scope, the size of import ontologies is a practical concern. The DO incorporates a number of large ontologies, including HPO, Sequence Types and Features Ontology (SO) [13], Chemical Entities of Biological Interest (ChEBI) [14], whose inclusion in full would make DO files unnecessarily large and slow logical reasoning. Instead, the terms relevant to DO have been identified and captured through the creation of custom import slims (see DO's GitHub import directory, https://github.com/DiseaseOntology/HumanDiseaseOntology/tree/main/src/ontology/imports). These are updated and aligned with the DO at each DO data release.

On occasion, unexpected changes in ontologies imported by DO have proved challenging. A good example of this was a major restructuring and renaming of terms that occurred in SO. During the subsequent DO update the SO import caused many of the previously SO-defined axioms to disappear. These axioms had to be redefined. Neither axioms nor slims can track these kinds of changes. To identify and mitigate similar issues in the future, an axiom report was created that lists all axioms and tracks changes in axioms per release.

4. Acknowledgements

The Disease Ontology team thanks our biomedical ontology collaborators, developers and curators across the Alliance of Genome Resources and Model Organism Databases for their ongoing coordination of human disease terms. This work is funded by the NIH/NHGRI Grant: U41 HG008735.

5. References

[1] L.M. Schriml, E. Mitraka, J. Munro, B. Tauber, M. Schor, L. Nickle, V. Felix, L. Jeng, C. Bearer, R. Lichenstein, K. Bisordi, N. Campion, B. Hyman, D. Kurland, C.P. Oates, S. Kibbey, P. Sreekumar, C. Le, M. Giglio, C. Greene, Human Disease Ontology 2018 update: classification, content and workflow expansion, Nucleic Acids Research. 47 (2019) D955–D962. https://doi.org/10.1093/nar/gky1032.

[2] R.C. Jackson, J.P. Balhoff, E. Douglass, N.L. Harris, C.J. Mungall, J.A. Overton, ROBOT: A Tool for Automating Ontology Workflows, BMC Bioinformatics. 20 (2019) 407. https://doi.org/10.1186/s12859-019-3002-3.

[3] J.S. Amberger, A. Hamosh, Searching Online Mendelian Inheritance in Man (OMIM): A Knowledgebase of Human Genes and Genetic Phenotypes, Current Protocols in Bioinformatics. 58 (2017) 1.2.1-1.2.12. https://doi.org/10.1002/cpbi.27.

[4] N. Sioutos, S. de Coronado, M.W. Haber, F.W. Hartel, W.-L. Shaiu, L.W. Wright, NCI Thesaurus: A semantic model integrating cancer-related clinical and molecular information, J. of Biomedical Informatics. 40 (2007) 30–43. https://doi.org/10.1016/j.jbi.2006.02.013.

[5] World Health Organization, ed., The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines, World Health Organization, Geneva, 1992.

[6] T.S. De Silva, D. MacDonald, G. Paterson, K.C. Sikdar, B. Cochrane, Systematized nomenclature of medicine clinical terms (SNOMED CT) to represent computed tomography procedures, Comput. Methods Prog. Biomed. 101 (2011) 324–329. https://doi.org/10.1016/j.cmpb.2011.01.002.

[7] Q. Zhu, D.-T. Nguyen, I. Grishagin, N. Southall, E. Sid, A. Pariser, An integrative knowledge graph for rare diseases, derived from the Genetic and Rare Diseases Information Center (GARD), Journal of Biomedical Semantics. 11 (2020) 13. https://doi.org/10.1186/s13326-020-00232-y.

[8] A. Rath, A. Olry, F. Dhombres, M.M. Brandt, B. Urbero, S. Ayme, Representation of rare diseases in health information systems: The orphanet approach to serve a wide range of end users, Human Mutation. 33 (2012) 803–808. https://doi.org/10.1002/humu.22078.

[9] S. Köhler, M. Gargano, N. Matentzoglu, L.C. Carmody, D. Lewis-Smith, N.A. Vasilevsky, D. Danis, G. Balagura, G. Baynam, A.M. Brower, T.J. Callahan, C.G. Chute, J.L. Est, P.D. Galer, S. Ganesan, M. Griese, M. Haimel, J. Pazmandi, M. Hanauer, N.L. Harris, M.J. Hartnett, M. Hastreiter, F. Hauck, Y. He, T. Jeske, H. Kearney, G. Kindle, C. Klein, K. Knoflach, R. Krause, D. Lagorce, J.A. McMurry, J.A. Miller, M.C. Munoz-Torres, R.L. Peters, C.K. Rapp, A.M. Rath, S.A. Rind, A.Z. Rosenberg, M.M. Segal, M.G. Seidel, D. Smedley, T. Talmy, Y. Thomas, S.A. Wiafe, J. Xian, Z. Yüksel, I. Helbig, C.J. Mungall, M.A. Haendel, P.N. Robinson, The Human Phenotype Ontology in 2021, Nucleic Acids Research. 49 (2021) D1207–D1217. https://doi.org/10.1093/nar/gkaa1043.

[10] M.A. Haendel, J.P. Balhoff, F.B. Bastian, D.C. Blackburn, J.A. Blake, Y. Bradford, A. Comte, W.M. Dahdul, T.A. Dececchi, R.E. Druzinsky, T.F. Hayamizu, N. Ibrahim, S.E. Lewis, P.M. Mabee, A. Niknejad, M. Robinson-Rechavi, P.C. Sereno, C.J. Mungall, Unification of multi-species vertebrate anatomy ontologies for comparative biology in Uberon, Journal of Biomedical Semantics. 5 (2014) 21. https://doi.org/10.1186/2041-1480-5-21.

[11] C. Mattingly, T. McKone, M. Callahan, J. Blake, E. Cohen Hubal, ExO: An Ontology for Exposure Science, Nat Prec. (2011) 1–1. https://doi.org/10.1038/npre.2011.6321.1.

[12] Thessen Anne E., Grondin Cynthia J., Kulkarni Resham D., Brander Susanne, Truong Lisa, Vasilevsky Nicole A., Callahan Tiffany J., Chan Lauren E., Westra Brian, Willis Mary, Rothenberg Sarah E., Jarabek Annie M., Burgoon Lyle, Korrick Susan A., Haendel Melissa A., Community Approaches for Integrating Environmental Exposures into Human Models of Disease, Environmental Health Perspectives. 128 (n.d.) 125002. https://doi.org/10.1289/EHP7215.

[13] K. Eilbeck, S.E. Lewis, C.J. Mungall, M. Yandell, L. Stein, R. Durbin, M. Ashburner, The Sequence Ontology: a tool for the unification of genome annotations, Genome Biology. 6 (2005) R44. https://doi.org/10.1186/gb-2005-6-5-r44.

[14] J. Hastings, G. Owen, A. Dekker, M. Ennis, N. Kale, V. Muthukrishnan, S. Turner, N. Swainston, P. Mendes, C. Steinbeck, ChEBI in 2016: Improved services and an expanding collection of metabolites, Nucleic Acids Research. 44 (2016) D1214–D1219. https://doi.org/10.1093/nar/gkv1031.