D940–D946 *Nucleic Acids Research, 2012, Vol. 40, Database issue* *Published online 12 November 2011*
doi:10.1093/nar/gkr972

Disease Ontology: a backbone for disease semantic integration

Lynn Marie Schriml[1,2,*], Cesar Arze[2], Suvarna Nadendla[2], Yu-Wei Wayne Chang[1,2], Mark Mazaitis[2], Victor Felix[2], Gang Feng[3] and Warren Alden Kibbe[3,*]

[1]Department of Epidemiology and Public Health, [2]Institute for Genome Sciences, University of Maryland School of Medicine, Baltimore, MD, USA, [3]The Biomedical Informatics Center and Robert H. Lurie Comprehensive Cancer Center, Northwestern University, 676 St Clair, Suite 1200, Chicago, IL 60611, USA

Received August 3, 2011; Revised October 12, 2011; Accepted October 16, 2011

ABSTRACT

The Disease Ontology (DO) database (http://disease-ontology.org) represents a comprehensive knowledge base of 8043 inherited, developmental and acquired human diseases (DO version 3, revision 2510). The DO web browser has been designed for speed, efficiency and robustness through the use of a graph database. Full-text contextual searching functionality using Lucene allows the querying of name, synonym, definition, DOID and cross-reference (xrefs) with complex Boolean search strings. The DO semantically integrates disease and medical vocabularies through extensive cross mapping and integration of MeSH, ICD, NCI's thesaurus, SNOMED CT and OMIM disease-specific terms and identifiers. The DO is utilized for disease annotation by major biomedical databases (e.g. Array Express, NIF, IEDB), as a standard representation of human disease in biomedical ontologies (e.g. IDO, Cell line ontology, NIFSTD ontology, Experimental Factor Ontology, Influenza Ontology), and as an ontological cross mappings resource between DO, MeSH and OMIM (e.g. GeneWiki). The DO project (http://diseaseontology.sf.net) has been incorporated into open source tools (e.g. Gene Answers, FunDO) to connect gene and disease biomedical data through the lens of human disease. The next iteration of the DO web browser will integrate DO's extended relations and logical definition representation along with these biomedical resource cross-mappings.

INTRODUCTION

From ancient texts such as the Eshuma Code of Babylon in the 23rd century BC (1) to the experimental results reported in literature today, scientists have documented variation in human health in order to unravel the mystery of disease. Diagnostic evaluation, treatment and data comparisons over time and between studies can be greatly facilitated by semantically consistent annotations such as those available through the Disease Ontology (DO).

The research and clinical communities have developed and utilized a variety of vocabularies in order to systematically record mortality and morbidity classifications, to standardize clinical and event healthcare reporting, to index Medline articles or to interconnect biomedical concepts defined across hundreds of disparately developed vocabularies, coding systems, thesauri and classifications. Although these vocabularies and ontologies include disease and disease related concepts and terms, *none* of them are 'organized' around the concept of disease.

The DO was developed to create a single structure for the classification of disease which unifies the representation of disease among the many and varied terminologies and vocabularies into a relational ontology that permits inference and reasoning of the relationships between disease terms and concepts and is optimized toward annotating disease.

The DO aims to provide a clear definition for each disease within an etiological based classification of disease enabling their consistent use and application for annotating biomedical data. The DO addresses the complexity of disease nomenclature through the inclusion of MeSH, OMIM, ICD and SNOMED CT concept names and IDs. The DO web browser will provide a framework for data mining, reasoning and inference enabling the exploration of biomedical disease and gene data for ongoing research and novel discovery based on the shared representation of disease. In this report, we present the new DO database and web browser (http://disease-ontology.org) (Figure 1), a description of the DO's semantic integration activities, data updates and the DO's development directions.

*To whom correspondence should be addressed. Tel: +1 410 706 6776; Fax: +1 410 706 6756; Email: lschriml@som.umaryland.edu
Correspondence may also be addressed to Warren Alden Kibbe. Tel: (312) 503-3229; Fax: (312) 503-5388; Email: wakibbe@northwestern.edu

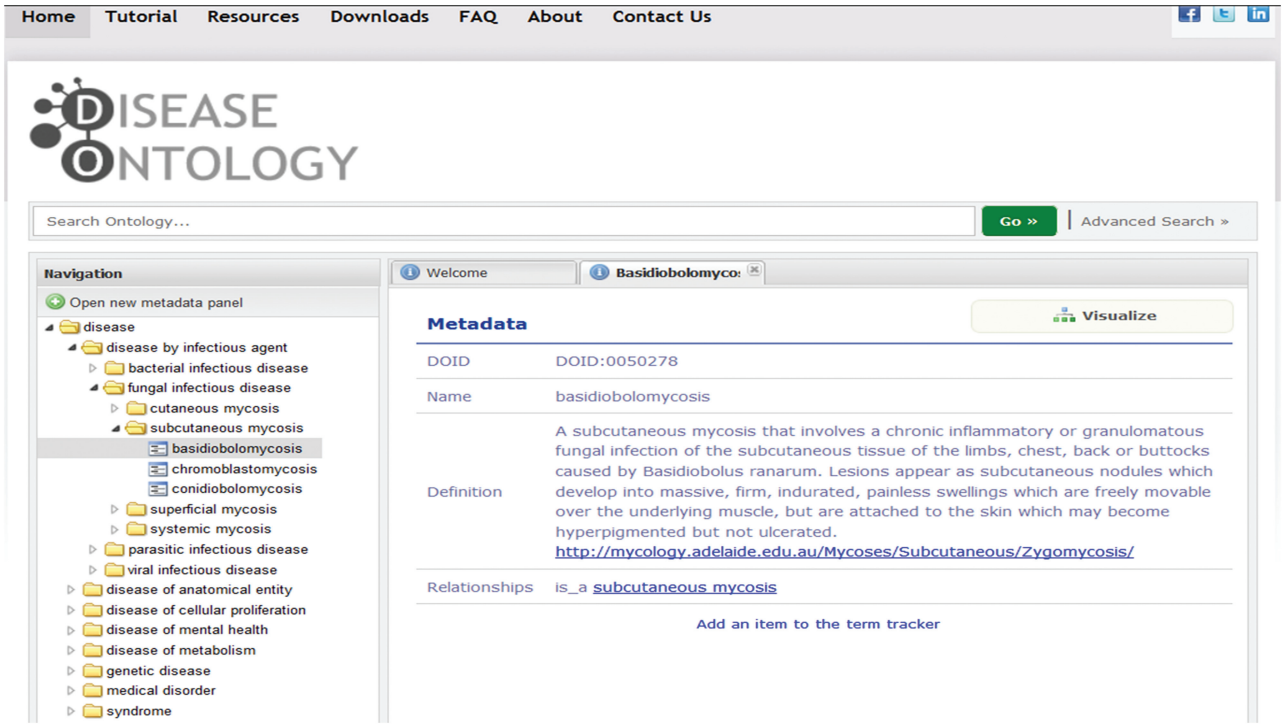

Figure 1. DO web interface with search, navigation and display functions. The disease tree view displays the DO's hierarchical structure and the placement of top level parent nodes expandable to view subgraphs. The fungal infectious disease subgraph with its direct child terms are shown. Term Metadata is displayed for selected terms from the tree view.

SCOPE OF THE DO

The DO is an open source ontological description of human disease, organized from a clinical perspective of disease etiology and location. Providing the classification framework for a disease 'Rosetta Stone' was a driving use case for starting the Disease Ontology in 2004 (2,3).

The initial builds of DO in 2003 and 2004 used ICD-9 as the foundational vocabulary. These early versions were extensively reorganized by process, system affected and cause (genetic disorders, infectious diseases, metabolic disorders). Further revisions improved with the re-organization of DO based on UMLS disease concepts in conjunction with term concept mappings to SNOMED CT and ICD-9.

The DO has become a community-driven, open and extensible framework for capturing human disease knowledge through direct and indirect semantic relationships. The DO enables the exploration of datasets and data resources through disease mappings available in clinical, gene and genome study metadata. This exploration leverages the semantic richness embedded in the DO. DO's directed acyclic graph (DAG) present terms linked by computable relationships in a hierarchy (e.g. brain glioblastoma multiforme is_a brain glioma, and brain glioma is_a brain cancer) organized by interrelated subtypes (e.g. Brill-Zinsser disease is_a epidemic typhus, and epidemic typhus is_a typhus). The DO is organized into eight main nodes to represent cellular proliferation, mental health, anatomical entity (e.g. cardiovascular system disease), infectious agent (e.g. anthrax), metabolism and genetic diseases along with medical disorders and syndromes anchored by traceable, stable identifiers (DOIDs).

The DO project continues to improve and expand the representation of all human disease with the addition of new DO terms as needed for curation, term requests and collaborative development. Rare diseases, for example, are currently underrepresented in DO. Curatorial efforts are underway to deepen DO's representation and to expand our standard is_a relations in the DO logical definition (HumanDO_xp.obo) file. The additional logical definition file format connects disease terms with related ontological concepts (e.g. anatomy, phenotype, disorder, cell type). The HumanDO_xp.obo file is available from DO's SourceForge site and includes additional relationships for 931 DO terms.

The DO provides ongoing documentation via the DO wiki (http://diseaseontology.sf.org), DO Facebook (http://www.facebook.com/group.php?gid=130516806961828), DO LinkedIn (http://www.linkedin.com/groups?gid=3078180&trk=anetsrch_name), DO twitter postings (http://twitter.com/#!/diseaseontology) and the DO website (http://disease-ontology.org/about).

Ontological disease definition

An ontological definition of disease enables each type (or class) of disease to be singularly classified in a formalized structure. The ontological distinction of disorder, disposition and disease as a realized disposition have been clarified by the development of the upper level

organizational Basic Formal Ontology (BFO, http://www.ifomis.uni-saarland.de/bfo/) and the Ontology of General Medical Science (OGMS, http://www.acsu.buffalo.edu/~ag33/ogms.html) along with discussion of ontological realism for mental disease (4) and the treatment of disease and diagnosis (5). Encompassing clinical descriptors of disease, the Disease Ontology has clarified DO's ontological scope with the adoption of the OGMS ontological definition of disease, 'A disposition (i) to undergo pathological processes that (ii) exists in an organism because of one or more disorders in that organism'. Within this context, DO describes the attributes of disease as manifested in individuals.

DO SEMANTIC INTEGRATION

The breadth of immune system, bone, mental, genetic and infectious disease subtrees in DO have been broadened through collaborative efforts with the DO team improving DO to meet the needs of our community. The DO project has provided the ontological framework for uniform data management and consistent annotation of human disease terms in biomedical databases and ontologies.

DO terms and their DOIDs have been utilized to annotate disease concepts in several major biomedical resources. The Rat Genome Database (6) (RGD) annotates their rat and mouse gene records and rat QTLs that are animal models of human disease with DO's human disease terms. The Immune Epitope Database (IEDB) (7) epitope records are annotated with 168 DO terms. Annotation of the GeneWiki's (8) gene records with 2983 candidate DO annotations is underway. Experimental expression records (9611) at the EBI's Array Express (9) have been annotated with DO terms representing an extensive resource for understanding the relationships between diseases and gene function.

DO continues to be utilized by a growing set of biomedical ontologies as a standard representation of disease. For example, the NCBO's Neuroscience Information Network (NIF) Standard ontology [NIFSTD, (10)] has integrated DO's representation of 252 mental disorders and neurological diseases. Feedback provided by NIF subject matter experts continues to improve DO's disease representation.

DO CONTENT AND STRUCTURE

The DO is logically structured into major types of disease to enable guided expansion of the ontology. The DO is being enhanced through the continued efforts to improve our representation of textual definitions (1822 textual definitions, 22% of DO terms, DO version 3, revision 2510). The DO's stable HumanDO.obo file provides the basis to advance DO's representation of the complex relationships between disease, disorder and phenotype. DO has begun to expand our set of cross-product relations linking DO terms to orthogonal ontologies with the annotation of disease attributes (e.g. symptom, phenotype, anatomical or cellular location and pathogenic agent) with 932 logical definitions in the DO's logical definition file (HumanDO_xp.obo) to the Foundational Model of Anatomy (FMA) (11), Human Phenotype Ontology (HP) (12), NCBI organismal classification vocabulary (13), Transmission Process ontology, Symptom Ontology (14), PATO (15), GO (16) and Cell Type ontology (17). Expansion of DO's set of relations in the HumanDO_xp.obo file (transmitted_by, results_in_formation_of, reslts_in, realized_by_suppression_with, part_of, located_in, has_symptom, has_material_basis_in, derives_from and composed_of) (18) will expand the DO's ability to define these complex relationships.

Linking disease terminologies

DO's extensive cross-mapping and inclusion of concepts from the standard clinical and medical terminologies [MeSH (19), ICD (20), OMIM (21) and NCI thesaurus (22)] into an ontological classification of disease (23) provides a rich resource for semantically connecting phenotypic, gene and genetic information related to human disease. Linking health information and patient's electronic health records will be further enhanced through the planned harmonization of ICD and SNOMED CT terminologies and classification (http://www.who.int/classifications/AnnouncementLetter.pdf).

DO identifies, integrates and connects synonymous disease concepts in MeSH, SNOMED CT, OMIM and ICD9CM and DO based on each disease term's UMLS Concept Unique Identifiers (CUIs). DO updates vocabulary mappings twice yearly from an extraction of term CUI's from the ULMS MRCONSO.RRF vocabulary mapping file (Table 1). Through this process, 91% (7845) of DO terms (August, 2011) are mapped to UMLS CUIs. This represents a 7% reduction of UMLS mappings since the May 2010 DO-UMLS mapping reflecting DO's increased utilization of logical definitions to define complex disease relationships which has decreased the number of unique DOIDs. For example, the DO defines adenocarcinoma as a type of (is_a relationship) carcinoma that is derived from epithelial cells which originate in glandular tissue. The DO defines gallbladder adenocarcinoma as a type of gallbladder carcinoma. These two sets of relationships represent a single

Table 1. DO UMLS CUI ID mappings

Vocabulary	Vocabulary IDs		DO IDs	
	May 2010	August 2011	May 2010	August 2011
OMIM	2304	1389	1594	2330
SNOMED CT	20 985	14 313	8054	5155
NCI thesaurus	7249	4761	7067	4858
MeSH	3932	3032	3921	3116
ICD9CM	6403	2971	5757	3325

A total of 7845 of 8588 DOID's (91%) in DO version 3, revision 2490 were mapped to the 2011 UMLS CUIs in August 2011. The number of unique vocabulary IDs mapped is given in the center column and the number of DO terms mapping to other terminologies through the CUI mapping file is presented on the right. Note that a single DO term may have multiple matches in a given terminology. The decrease in SNOMED CT mappings is a reflection of the increased use of logical definitions in DO.

parentage for each term as defined in the HumanDO.obo file and visualized in the DO web browser. Multiple parentage (multiple is_a relationships) inherited from the UMLS vocabularies have been greatly reduced in the current version of the DO. Curatorial efforts are ongoing to represent secondary parentage with the creation of cross-reference definitions (logical definitions) in the HumanDO_xp.obo logical definition file. Logical definitions provide the opportunity to define the relationship between a type of organ cancer (e.g. gallbladder adenocarcinoma) and tumor's cell type (e.g. adenoma) as a type of adenoma or to define the anatomical location of a disease (gallbladder adenocarcinoma is located_in the gallbladder.

We are investigating a technical solution to enable the connection of external references through a synthetic term derived from the logical definitions. DO's set of OMIM cross-references have been validated through manual review (Spring 2011) of each disease term. OMIM cross-references have been added to DO through this process to raise the count of mapped DO-OMIM records to 1630.

DO WEB INTERFACE

DO browser database

The DO web-browser was constructed using Web 2.0 and semantic web technologies. At the forefront of these is the Neo4j graph database server (http://neo4j.org/). Neo4j provides several robust and fast mechanisms to retrieve individual nodes or to traverse a set of nodes moving between each via their relationships that otherwise would require complex join operations in a relational database.

Built into Neo4j are optimized functions for retrieving the (all, shortest, user-defined) path between two terms—very common and useful in visualizing term relationships in an ontology. The DO browser leverages the RESTful API of Neo4j (http://components.neo4j.org/neo4j-server/snapshot/rest.html) to retrieve nodes and their associated properties via HTTP ajax calls. This enables power-users familiar with the Neo4j RESTful API to request data from the Disease Ontology database 'programmatically' and facilitates integration into external projects. Currently any user may retrieve metadata for a specific DO term by making use of our REST metadata API by constructing a HTTP request in the following format: http://www.disease-ontology.org/api/metadata/<DOID>

An example scenario would be retrieving the metadata for the term 'transient cerebral ischemia': http://www.disease-ontology.org/api/metadata/DOID:224

This query would return a JSON packet containing all the metadata for this term including parents, children, definition, xrefs, synonym, name, alternate IDs and DO identifier. In the future we hope to increase the number of API commands available to cover operations such as searching and the path between two nodes.

DO terms are modeled in the Neo4j database with each node of the graph being a unique term containing the following properties: Name, DOID, Definition, Synonym(s), Alternate ID(s), Subset(s), Cross-Reference(s) and Relationship. The edges of the graph database represent relationships between terms in the ontology and have a relationship type, positioning the DO browser to enable the exploration of term connections by relationships other than 'is_a' as the number of logical definitions in DO expands.

Visualization

User interface. The DO browser was designed with a focus on presenting all the ontology tree, query results, DO term metadata and visualization on a single page with multiple tabs that allows for any metadata, search results or visualizations to persist while the ontology is further explored.

The interface consists of HTML and CSS for controlling the layout, sizing, fonts and color scheme. The tree and visualization components are delivered through ajax using the ExtJS (http://www.sencha.com/products/extjs/) and jquery libraries (http://jquery.com/) for added GUI elements and functionality. The full-text of the DO (name, synonym, definition, xrefs, DOIDs) is included in the Apache Lucene (http://lucene.apache.org/java/docs/index.html) index. The Lucene indexing allows users to search all or any fields as all text of the ontology is run through an analyzer that removes common English stop words and tokenizes the text allowing for flexible queries that return results sets containing partial hits.

The layout of the DO Browser can be sub-divided into three distinct sections, as seen in the 'DO Tutorial' (http://disease-ontology.org/tutorial/). The 'Search Panel' provides all the necessary tools to execute basic or advanced queries on the DO. The 'Navigation Panel' contains a interactive tree-based model enabling navigation and exploration of subtrees. The ontology can be traversed by a single-click of the arrow found to the left of each term or a double-click of a term that is denoted with a folder icon. Once expanded any children for a given term will be rendered into the tree and can be likewise expanded until a leaf node is encountered. The Navigation Panel tree is refreshed when a term is selected from search results or a Metadata Panel. The 'Content Panel' tabs house search results, term metadata and graphic visualization of terms and term relationships. Visualization of nodes (Figure 2) in the DO Browser can be accessed through the 'Visualization' button found in the 'Metadata Panel'. Invoking the visualization feature of a term will create a new tab that will house an interactive canvas upon which the target term and any children or parents will be rendered and explored. By default, terms rendered on the canvas in a visualization panel attempt to arrange themselves in a layout that prevents any overlapping. Terms that have any associated parents or children will be colored green and can be expanded by a single click. Upon selecting a term from either the Navigation Panel tree or a set of search results a new 'Metadata Panel tab' is created to display available term metadata including DOID, Name, Definition, Xrefs, Alternateids, Synonyms, Relationships and a link to the DO term tracker. Where available cross-references (xrefs) and definitions will contain links out to the relevant resource.

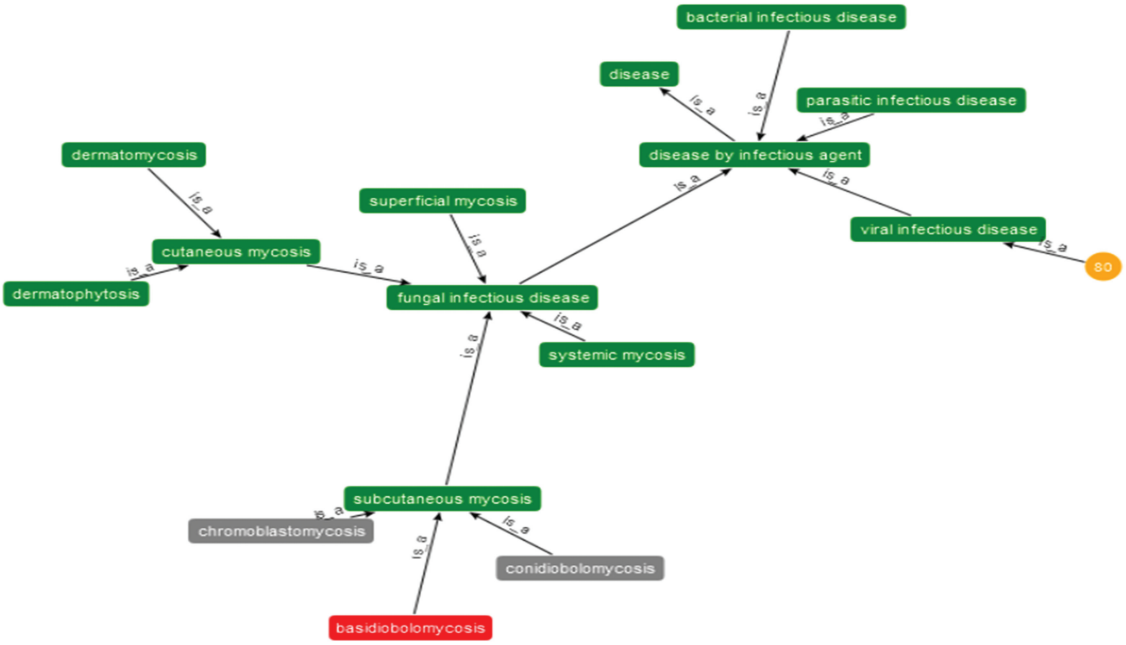

Figure 2. Term visualization in the DO web-browser. The 'Visualize' button on the Metadata page opens a graphical view of DO. Clicking on this button will open a new tab that will display the target node of the visualization (e.g. basidiobolomycosis) [red box], parent node [green box] and sibling leaf nodes [gray box]. Nodes with five or more children are represented by a gold circle containing the number of children. Clicking on a node in the graph will expand the view.

The 'Search Panel' provides 'full-text' contextual searching functionality against all metadata fields (Basic Search) and an Advanced Search that allows for the user to generate targeted and complex Boolean queries against specific fields of the ontology. The Advanced Search dialog box facilitates queries with the option to 'Match All' (AND) or 'Match Any' (OR) of the query terms provided. Each search generates a distinct panel allowing for persistence of result sets.

Comparison with alternative ontology visualization services. The DO web browser was developed with expanded search and display capabilities for the DO. DO's non-flash interactive graphic visualization and full text metadata searching is not available from current ontology resources [e.g. EBI's Ontology Lookup Service (24) and NCBO's BioPortal (25)] and provides accessibility with mobile devices such as the iPad. Implementation of full-text searching provides users with full access to the depth and richness available in the Disease Ontology. Alternative services limit their searches against the name or synonym field. Furthermore the DO Browser provides the ability to create complex searches. The DO web browser uniquely provides links to definitions sources and Xref links to NCI, OMIM, ICD, MeSH and SNOMED CT vocabulary terms. These features provide DO users with a true semantic linkage between disease concepts based on concept identifiers rather than a text based matching. Cross-browser and cross-platform support was a strong design point for the Disease Ontology Browser and reflects that the site does not make use of any third-party plugins to render content.

DO SOFTWARE DEVELOPMENT PROJECTS

Concurrent with the Disease Ontology development, the DO group has developed the FunDO and GeneAnswers data access and exploration tools. The Functional Disease Ontology (FunDO) Web application (http://django.nubic.northwestern.edu/fundo/) (3) can be used to measure the internal consistency of DO as well as the ability of DO to functionally annotate a gene list with disease. FunDO takes a list of genes and finds relevant diseases based on statistical analysis of the Disease Ontology annotation database. 'GeneAnswers' (http://www.bioconductor.org/packages/2.5/bioc/html/GeneAnswers.html) is a reusable bioconductor software package encompassing reproducible disease-gene pathway models that can be utilized directly by researchers or incorporated into other biomedical resources (26). GeneAnswers has been downloaded 2000 times by >1000 members of the bioinformatics community since July 2010. 'DOGA' (Human Disease Gene Annotation Database) is a beta version tool for examining disease-gene annotations through data available in Gene Wiki or in the NCBI GeneRIFs (http://doga.nubic.northwestern.edu).

FUTURE DEVELOPMENT

The Disease Ontology's representation of human disease is being advanced through the inclusion of cross-references to orthogonal concepts defined by logical definitions in the HumanDO_xp.obo file. Connecting related ontological concepts, augmenting DO's relationship types and visualizing integrated disease mapping between

biomedical resources will broaden the utility of the Disease Ontology and DO web-browser.

Extension of the DO paradigm

The DO paradigm is extensible for the creation of non-human organism disease ontologies. Disease defined by etiology and the affected body system are universally applicable principles for describing organism-specific pathologies in model organism, livestock or plants. Defining diseases in DO, we have developed guiding principles to facilitate consistent classification of disease terms. For instance, a disease is classified first by etiology, if known, with a singular is_a relation. The location of the affected body system is annotated in the disease definition and then linked by the relation 'located_in' to the corresponding FMA term. A disease of unknown etiology with well defined localization is defined by the affected body system. The DO style Guide (http://do-wiki.nubic.northwestern.edu/index.php/Style_Guide) outlines the curatorial guiding principles for DO. Disease defined by etiology and the affected body system are universally applicable principles for describing organism-specific pathologies in model organism, livestock or plants with categorizations of infectious (viral, bacterial, fungal, parasitic), inherited and acquired disease (cancer, metabolism and mental health disease). Cross-references to models of human disease in DO, as defined by the model organism databases, would define the disease to model relationship.

AVAILABILITY

Disease ontology files are available under the Creative Commons license in three formats: the OBO formatted Disease Ontology (HumanDO.obo); the Disease Ontology file without cross-references (HumanDO_no_xrefs.obo); and an enhanced Disease Ontology file containing logical definitions to orthogonal OBO Foundry ontologies (HumanDO_xp.obo). The HumanDO.obo file is available from SourceForge (http://diseaseontology.svn.sourceforge.net/viewvc/diseaseontology/trunk/HumanDO.obo) and can be downloaded from the OBO Foundry (http://www.obofoundry.org/cgi-bin/detail.cgi?id=disease_ontology). DO is available in OWL format at the University of California at Berkeley (http://www.berkeleybop.org/ontologies/owl/DOID). The Disease Ontology web app source code will be made freely available at: https://github.com/IGS/disease-ontology. The Disease Ontology can also be browsed in EBI's Ontology Lookup Service (24) and NCBO's BioPortal (25).

ACKNOWLEDGEMENTS

The authors would like to thank the past and current members of the DO Advisory Board: Michael Ashburner, Judith Blake, Rex Chisholm, Suzanna Lewis, Maryann Martone, Chris Mungall, Alan Ruttenberg, Richard Scheuermann and Barry Smith for their invaluable guidance. We are indebted to our many contributors and collaborators. We the authors would like to thank Eric Neumann, Bernard de Bono, Peter Robinson and Bjoern Peters for their ongoing collaborative discussions.

FUNDING

National Institutes of Health—National Center for Research Resources (NCRR) (grant number R01RR025342) under the ARRA mechanism. The PIs would like to gratefully recognize the NCRR for funding this activity. Funding for open access charge: National Institutes of Health—National Center for Research Resources (NCRR) (grant number R01RR025342) under the ARRA mechanism.

Conflict of interest statement. None declared.

REFERENCES

1. Rupprecht,C.E., Smith,J.S., Fekadu,M. and Childs,J.E. (1995) The ascension of wildlife rabies: a cause for public health concern or intervention? *Emerg. Infect. Dis.*, **1**, 107–114.
2. Osborne,J.D., Flatow,J., Holko,M., Lin,S.M., Kibbe,W.A., Zhu,L.J., Danila,M.I., Feng,G. and Chisholm,R.L. (2009) Annotating the human genome with Disease Ontology. *BMC Genomics*, **10**, S1–S6.
3. Du,P., Feng,G., Flatow,J., Song,J., Holko,M., Kibbe,W.A. and Lin,S.M. (2009) From disease ontology to disease-ontology lite: statistical methods to adapt a general-purpose ontology for the test of gene-ontology associations. *Bioinformatics*, **25**, i63–i68.
4. Ceusters,W. and Smith,B. (2010) Foundations for a realist ontology of mental disease. *J. Biomed. Semantics*, **1**, 10.
5. Scheuermann,R.H., Ceusters,W. and Smith,B. (2009) Toward an ontological treatment of disease and diagnosis. *Summit Translat. Bioinforma.*, **2009**, 116–120.
6. Gene Ontology Consortium. (2010) The Gene Ontology in 2010: extensions and refinements. *Nucleic Acids Res.*, **38**, D331–D335.
7. Twigger,S.N., Shimoyama,M., Bromberg,S., Kwitek,A.E., Jacob,H.J. and RGD,Team. (2007) The Rat Genome Database, update 2007 – easing the path from disease to data and back again. *Nucleic Acids Res.*, **35**, D658–D662.
8. Vita,R., Zarebski,L., Greenbaum,J.A., Emami,H., Hoof,I., Salimi,N., Damle,R., Sette,A. and Peters,B. (2010) The immune epitope database 2.0. *Nucleic Acids Res.*, **38**, D854–D862.
9. Huss,J.W., Lindenbaum,P., Martone,M., Roberts,D., Pizarro,A., Valafar,F., Hogenesch,J.B. and Su,A.I. (2010) The Gene Wiki: community intelligence applied to human gene annotation. *Nucleic Acids Res.*, **38**, D633–D639.
10. Parkinson,H., Sarkans,U., Kolesnikov,N., Abeygunawardena,N., Burdett,T., Dylag,M., Emam,I., Farne,A., Hastings,E., Holloway,E. *et al.* (2011) ArrayExpress update – an archive of microarray and high-throughput sequencing-based functional genomics experiments. *Nucleic Acids Res.*, **39**, D1002–D1004.
11. Bug,W.J., Ascoli,G.A., Grethe,J.S., Gupta,A., Fennema-Notestine,C., Laird,A.R., Larson,S.D., Rubin,D., Shepherd,G.M., Turner,J.A. *et al.* (2008) The NIFSTD and BIRNLex vocabularies: building comprehensive ontologies for neuroscience. *Neuroinformatics*, **6**, 175–194.
12. Rosse,C. and Mejino,J.V.L. (2003) A reference ontology for biomedical informatics: the foundational model of anatomy. *J. Biomed. Inform.*, **36**, 478–500.
13. Robinson,P.N. and Mundlos,S. (2010) The human phenotype ontology. *Clin. Genet.*, **77**, 525–534.
14. Sayers,E.W., Barrett,T., Benson,D.A., Bryant,S.H., Canese,K., Chetvernin,V., Church,D.M., DiCuccio,M., Edgar,R., Federhen,S. *et al.* (2009) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **37**, D5–D15.
15. Schriml,L.M., Arze,C., Nadendla,S., Ganapathy,A., Felix,V., Mahurkar,A., Phillippy,K., Gussman,A., Angiuoli,S., Ghedin,E. *et al.* (2010) GeMInA, Genomic Metadata for Infectious Agents,

a geospatial surveillance pathogen database. *Nucleic Acids Res.*, **38**, D754–D764.
16. Mungall,C.J., Gkoutos,G.V., Smith,C.L., Haendel,M.A., Lewis,S.E. and Ashburner,M. (2010) Integrating phenotype ontologies across multiple species. *Genome Biol.*, **11**, R2.1.
17. Sarntivijai,S., Ade,A.S., Athey,B.D. and States,D.J. (2007) The cell line ontology and its use in tagging cell line names in biomedical text. *AMIA Annu. Symp. Proc.*, **11**, 1103.
18. Smith,B., Ceusters,W., Klagges,B., Köhler,J., Kumar,A., Lomax,J., Mungall,C., Neuhaus,F., Rector,A.L. and Rosse,C. (2005) Relations in biomedical ontologies. *Genome Biol.*, **6**, R46.
19. Nelson,S.J., Schopen,M., Savage,A.G., Schulman,J. and Arluk,N. (2004) The MeSH translation maintenance system: structure, interface design, and implementation. In: Fieschi,M. *et al.* (eds), *Proceedings of the 11th World Congress on Medical Informatics*. IOS Press, Amsterdam, San Francisco, CA, pp. 67–69.
20. Ayme,S., Rath,A. and Bellet,B. (2010) WHO International Classification of Diseases (ICD) Revision Process: incorporating rare diseases into the classification scheme: state of art. *Orphanet J. Rare Dis.*, **5**, P1.
21. Amberger,J., Bocchini,C. and Hamosh,A. (2011) A new face and new challenges for Online Mendelian Inheritance in Man (OMIM(®)). *Hum. Mutat.*, **32**, 564–567.
22. Sioutos,N., de Coronado,S., Haber,M.W., Hartel,F.W., Shaiu,W.L. and Wright,L.W. (2007) NCI Thesaurus: a semantic model integrating cancer-related clinical and molecular information. *J. Biomed. Inform.*, **40**, 30–43.
23. Osborne,J.D., Lin,S.M., Zhu,L.J. and Kibbe,W.A. (2007) Mining biomedical data using MetaMap Transfer (MMtx) and the Unified Medical Language System (UMLS). *Methods Mol. Biol.*, **408**, 153–169.
24. Cote,R.G., Jones,P., Apweiler,R. and Hermjakob,H. (2006) The ontology lookup service, a lightweight cross-platform tool for controlled vocabulary queries. *BMC Bioinformatics*, **7**, 97.
25. Jonquet,C., Musen,M.A. and Shah,N.H. (2010) Building a biomedical ontology recommender web service. *J. Biomed. Semantics*, **1**, S1.
26. Feng,G., Du,P., Krett,N.L., Tessel,M., Rosen,S., Kibbe,W.A. and Lin,S.M. (2010) A collection of bioconductor methods to visualize gene-list annotations. *BMC Res. Notes*, **3**, 10.