Disease Ontologies

Olivier Bodenreider, US National Library of Medicine, Bethesda, Maryland, USA,
obodenreider@mail.nih.gov

Background

Disease ontologies can be traced back to the 17th century, when health authorities in London used a standard list of about 200 causes of death to compile accurate health statistics known as the Bills of Mortality (Bodenreider 2008). This list was later integrated into the International Classification of Diseases, the 11th revision of which is currently under active development. Among the several hundredths of biomedical ontologies currently available, a few dozen provide coverage of diseases. A selection of these ontologies is presented in this brief review.

Definition

Biomedical ontologies

Biomedical ontologies represent the properties of biomedical entities and their relations to other biomedical entities. As such, biomedical ontologies ontologies are artifacts used to represent and share knowledge about the biomedical domain (Bodenreider and Stevens 2006). More specifically, biomedical ontologies tend to focus on definitional knowledge (i.e., what is always true of biomedical entities), as opposed to assertional knowledge, usually found in knowledge bases. Ontologies also differ from terminologies, whose focus is purely naming, and from thesauri, in which knowledge is usually organized for a specific purpose (e.g., information retrieval). Despite these differences, the term "ontology" is often used loosely, as an umbrella name for these various kinds of artifacts.

Disease ontologies

Disease ontologies are biomedical ontologies providing coverage for the domain of diseases, disorders, illness, etc. The degree to which these words are synonymous is subject to debate (Ceusters and Smith 2010), but disease ontologies generally cover conditions understood or suspected of being a deviation from a healthy status and, less frequently, diagnostic criteria for these conditions. Some disease ontologies also cover the manifestations of these conditions, i.e., the signs and symptoms associated with them. However, the relations between conditions and their manifestations are usually not recorded in ontologies, as such relations are not definitional for most diseases. In fact, except for the so-called pathognomonic manifestations, there is a probabllistic, not systematic relation between a manifestation and a condition. Phenotypes are the observable characteristics of organisms resulting from the genetic makeup of a particular organism. There is partial overlap between phenotypes and disease manifestations, and phenotypes may be covered by disease ontologies.

The principal use of disease ontologies is to support the annotation of diseases in

biomedical datasets, including the curation of knowledge bases, the clinical documentation of electronic health records and the indexing of the biomedical literature. Disease ontologies are also used for aggregation purposes (e.g., for grouping myocardial infarction and mitral stenosis under cardiovascular diseases), as well as for clinical decision support (e.g., drugs contra-indicated with asthma should not be prescribed to patients diagnosed with specific forms of asthma, such as seasonal asthma and occupational asthma).

Existing disease ontologies and their characteristics

In this section, we review 17 ontologies, which roughly qualify as disease ontologies according to the definition above. The list of ontologies is shown in Table 1, along with a URL from which more information can be obtained. A list of salient characteristics for disease ontologies is presented in Table 2. Finally, for each ontology, we provide a brief description and a list of characteristics summarized in Table 3, with additional notes in Table 4.

- **Disease Ontology** (DO): Controlled terminology originally created for annotation purposes as part of the NuGene project at Northwestern University. Still under development.
- **Diagnostic and Statistical Manual of Mental Disorders** (DSM): Standard classification of mental disorders in the United States, developed by the American Psychiatry Association and used by a wide range of mental health professionals across clinical settings.
- **Human Phenotype Ontology** (HPO): Controlled vocabulary for the phenotypic features encountered in human hereditary and other diseases, used for the annotation of the genetic diseases listed in OMIM. Developed by a consortium including Charité Hospital (Berlin) and the University of Cambridge (UK).
- **International Classification of Diseases** (ICD): Classification from the World Health Organization (WHO) family of health classifications, with many local adaptations. ICD9-CM, developed by the Center for Medicare & Medicaid Services (CMS) for use in the US, includes clinical modifications. Broad coverage of diseases and health problems.
- **International Classification of Primary Care** (ICPC): Classification of reasons for encounter, diagnoses or problems, and process of care. Developed by the World Organization of Family Doctors (Wonca). Coverage of diseases and health problems at the level of detail required for primary care.
- **Infectious Disease Ontology** (IDO): Set of ontologies for specific infectious diseases, including malaria, influenza and tuberculosis, sharing a core ontology. Covers entities relevant to both biomedical and clinical aspects of most infectious diseases. Developed by the Infectious Disease Ontology Consortium.
- **Logical Observation Identifiers Names and Codes** (LOINC): Set of names and codes for laboratory and other clinical observations (elements of clinical phenotypes). Developed at the Regenstrief Institute. Coverage restricted to clinical observations.
- **MEDCIN**: Developed by Medicomp Systems, MEDCIN is a vocabulary for clinical documentation and a knowledge base for clinical decision support. It provides coverage for elements including symptoms, medical history, physical examination,

tests and diagnoses.

- **MedDRA**: Created by a pharmaceutical industry trade group, the Medical Dictionary for Regulatory Activities (MedDRA) is a medical terminology used to classify adverse event information associated with the use of medications, vaccines and medical devices, especially for reporting to regulatory agencies.
- **Medical Subject Headings** (MeSH): Controlled vocabulary developed by the U.S. National Library of Medicine for the indexing and retrieval of the biomedical literature, especially in the MEDLINE bibliographic database. Broad coverage including diseases.
- **Pathbase pathology ontology** (MPATH): Ontology of mutant and transgenic mouse pathology phenotypes used for the annotation of Pathbase, a repository of histopathology images. Developed by the Pathbase European Consortium.
- **Mammalian Phenotype Ontology** (MPO): Controlled vocabulary for the annotation of mammalian phenotypes, currently used for the annotation of phenotypic data in mouse and rat databases. Developed at the Jackson Laboratory. Coverage restricted to phenotypes.
- **NCI Thesaurus** (NCIt): Controlled vocabulary developed by the National Cancer Institute to support the integration of information related to cancer research. Broad coverage including diseases.
- **National Drug File-Reference Terminology** (NDF-RT): Reference terminology for medications, providing information including pharmacologic class, therapeutic intent, mechanism of action, and physiologic effect. Produced by the U.S. Department of Veterans Affairs, Veterans Health Administration (VHA). Coverage of diseases through their relations to drugs (therapeutic intent).
- **Online Mendelian Inheritance in Man** (OMIM): Knowledge base on human genetic diseases developed at John Hopkins University and available through the NCBI Entrez system. Coverage restricted to genetic diseases.
- **Phenotypic Quality Ontology** (PATO): Ontology of phenotypic qualities, intended for use in a number of applications, primarily defining composite phenotypes and phenotype annotation. Coverage restricted to phenotypes.
- **SNOMED CT**: The largest clinical terminology, maintained by the International Health Terminology Standard Development Organization (IHTSDO) for use in electronic health records and adopted by fifteen countries to date. Broad coverage including diseases.

Ontology repositories

Many of the disease ontologies listed above are present in ontology repositories (Table 5), which offer a convenient way of integrating disease resources annotated to different ontologies. The **Unified Medical Language System** (UMLS) is a terminology integration system developed by the U.S. National Library of Medicine. The UMLS establishes a correspondence among terms from different terminologies for a given biomedical entity. It integrates a number of the terminologies presented above, as well as many other biomedical terminologies. Developed by the National Center for Biomedical Ontology (NCBO), The **BioPortal** is another such repository, which offers mapping among terms from different

ontologies. The BioPortal provides systematic coverage of the ontologies from the Open Biomedical Ontologies (OBO) family, as well as many other ontologies. The NCBO also indexes resources, such as clinical trials and gene expression databases, in reference to ontology entities from the BioPortal.

Table 1. List of disease ontologies

DO	*Disease Ontology -* http://diseaseontology.sourceforge.net/
DSM	*Diagnostic and Statistical Manual of Mental Disorders -* http://www.psych.org/
HPO	*Human Phenotype Ontology -* http://www.human-phenotype-ontology.org/
ICD	*International Classification of Diseases -* http://www.who.int/classifications/icd/en/
ICPC	*International Classification of Primary Care -* http://www.globalfamilydoctor.com/wicc/
IDO	*Infectious Disease Ontology -* http://www.infectiousdiseaseontology.org/
LOINC	*Logical Observation Identifiers Names and Codes -* http://loinc.org
MEDCIN	*MEDCIN -* http://www.medicomp.com/
MedDRA	*Medical Dictionary for Regulatory Activities -* http://www.meddramsso.com/
MeSH	*Medical Subject Headings -* http://www.nlm.nih.gov/mesh/
MPATH	*Mouse Pathology Ontology -* http://www.pathbase.net/
MPO	*Mammalian Phenotype Ontology -* http://www.informatics.jax.org/searches/MP_form.shtml
NCI Thes.	*NCI Thesaurus -* http://ncit.nci.nih.gov/
NDF-RT	*National Drug File-Reference Terminology -* http://evs.nci.nih.gov/ftp1/NDF-RT/
OMIM	*Online Mendelian Inheritance in Man -* http://www.ncbi.nlm.nih.gov/omim/
PATO	*Phenotypic Quality Ontology -* http://obofoundry.org/wiki/index.php/PATO:Main_Page
SNOMED CT	*SNOMED CT -* http://www.ihtsdo.org/

Table 2. List of salient characteristics for disease ontologies

component	the disease ontology is a component of a broader ontology
specialized	the disease ontology only covers a specific group of diseases
human	the disease ontology mainly covers human diseases
OBO	the disease ontology is part of the Open Biomedical Ontologies (OBO) family of ontologies
clinical	the disease ontology is mainly used in clinical practice
definitions	the disease ontology includes definitions (textual or logical)
translations	the disease ontology is available in other languages than English
publicly available	the disease ontology is publicly available
x-ref	the disease ontology has cross-references to other disease ontologies (natively or through the UMLS)

Table 3. Some characteristics of 17 disease ontologies (see Table 3 for the definitions of the characteristics)

	Comp-onent	Spec-ialized	Human	OBO	Clinical	Definitions	Trans-lations	Publicly available	X-ref
DO	no	no	yes	yes	no	textual	no	yes	native
DSM	no	yes	yes	no	yes	none*	no	no	UMLS
HPO	no	yes	yes	yes	no*	textual	no	yes	native
ICD	no	no	yes	no	yes	none	yes	no	UMLS
ICPC	no	yes*	yes	no	yes	none	yes	yes	native to ICD, UMLS
IDO	no	yes	yes	yes	no	textual	no	yes	none
LOINC	no	yes	yes	no	yes	logical*	yes	yes	UMLS
MEDCIN	yes	no	yes	no	yes	none	no	no	UMLS
MedDRA	no	yes*	yes	no	yes	none	yes	no	UMLS
MeSH	yes	no	yes*	no	no	textual	yes	yes	UMLS
MPATH	no	no	no	yes	no	textual	no	yes	none
MPO	no	no	no	yes	no	textual	no	yes	none
NCI Thes.	yes	yes*	yes	no	yes*	logical, textual	no	yes	native, UMLS
NDF-RT	yes	no	yes	no	yes	logical*	no	yes	UMLS
OMIM	no	yes	yes	no	yes	textual*	no	yes	UMLS
PATO	no	no	yes*	yes	no	logical, textual	no	yes	none
SNOMED CT	yes	no	yes	no	yes	logical	yes	yes*	UMLS

Table 4. Additional notes on Table 3 items

DSM	provides diagnostic criteria for many mental disorders
HPO	PhenExplorer is a clinical diagnostic tool based on HPO annotations of OMIM diseases
ICPC	primary care can be considered a specialty
IDO	IDO borrows concepts from other OBO ontologies
LOINC	although LOINC does not use description logics (DL), its organization is close to DL representation
MedDRA	MedDRA is not restricted to any medical specialties, but focuses on adverse events
MeSH	MeSH covers, but isnot limited to human diseases
NCI Thes.	NCIt essentially covers cancers and cancer-related diseases; used in clinical research
NDF-RT	weak logical definitions (primitive classes)
OMIM	OMIM contains extensive narrative descriptions more than definitions
PATO	PATO represents all kinds of phenotypes, including in humans
SNOMED CT	freely available for use in the IHTSDO member countries

Table 5. Ontology repositories providing coverage of disease entities

UMLS	*Unified Medical Language System -* https://uts.nlm.nih.gov
BioPortal	*NCBO BioPortal -* http://bioportal.bioontology.org/

Acknowledgments

References

Bodenreider, O. (2008). "Biomedical ontologies in action: role in knowledge management, data integration and decision support." Yearb Med Inform: 67-79.

Bodenreider, O. and R. Stevens (2006). "Bio-ontologies: current trends and future directions." Brief Bioinform **7**(3): 256-274.

Ceusters, W. and B. Smith (2010). "Foundations for a realist ontology of mental disease." J Biomed Semantics **1**(1): 10.