Introduction to Bio-Ontologies

Barry Smith

http://ontology.buffalo.edu/smith
Outline

1. Who am I?
2. How to find your data
3. How to do biology across the genome
4. How to extend the GO methodology to clinical and translational medicine
5. Anatomy Ontologies: An OBO Foundry success story
6. The Infectious Disease Ontology
7. The Environment Ontology
1. **Who am I?**
2. How to find your data
3. How to do biology across the genome
4. How to extend the GO methodology to clinical and translational medicine
5. Anatomy Ontologies: An OBO Foundry success story
6. The Infectious Disease Ontology
7. The Environment Ontology
Who am I?

Foundational Model of Anatomy Ontology (FMA)
Common Anatomy Reference Ontology (CARO)
Protein Ontology (PRO)
Infectious Disease Ontology (IDO)
Ontology for General Medical Science (OGMS)
Plant Ontology (PO)
Biometrics Upper Ontology
NCBO: National Center for Biomedical Ontology (NIH Roadmap Center)

- Stanford Biomedical Informatics Research
- The Mayo Clinic
- University at Buffalo Department of Philosophy

http://bioportal.bioontology.org
<table>
<thead>
<tr>
<th>TERM NAME</th>
<th>MATCHED IN</th>
<th>ONTOLOGY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wood</td>
<td>Preferred Name</td>
<td>Read Codes, Clinical Terms Version 3 (CTV3)</td>
</tr>
<tr>
<td>Wood</td>
<td>Preferred Name</td>
<td>NCI Thesaurus</td>
</tr>
<tr>
<td>Wood</td>
<td>Preferred Name</td>
<td>Galen</td>
</tr>
<tr>
<td>wood</td>
<td>Preferred Name</td>
<td>MDSS Mo</td>
</tr>
<tr>
<td>wood</td>
<td>Preferred Name</td>
<td>Environment Ontology</td>
</tr>
<tr>
<td>Wood</td>
<td>Preferred Name</td>
<td>SNOMED Clinical Terms</td>
</tr>
<tr>
<td>Wood</td>
<td>Preferred Name</td>
<td>Medical Subject Headings</td>
</tr>
<tr>
<td>wood</td>
<td>Preferred Name</td>
<td>envo_153</td>
</tr>
<tr>
<td>coconut plantation</td>
<td>Synonym</td>
<td>Environment Ontology</td>
</tr>
<tr>
<td>coconut plantation</td>
<td>Synonym</td>
<td>envo_153</td>
</tr>
<tr>
<td>forest</td>
<td>Synonym</td>
<td>Environment Ontology</td>
</tr>
<tr>
<td>forest</td>
<td>Synonym</td>
<td>envo_153</td>
</tr>
<tr>
<td>olive grove</td>
<td>Synonym</td>
<td>Environment Ontology</td>
</tr>
<tr>
<td>olive grove</td>
<td>Synonym</td>
<td>envo_153</td>
</tr>
<tr>
<td>secondary xylem</td>
<td>Synonym</td>
<td>Cereal plant gross anatomy</td>
</tr>
<tr>
<td>secondary xylem</td>
<td>Synonym</td>
<td>Plant Anatomy</td>
</tr>
<tr>
<td>secondary xylem</td>
<td>Synonym</td>
<td>Plant Anatomy</td>
</tr>
<tr>
<td>Wood material</td>
<td>Synonym</td>
<td>NCI Metathesaurus</td>
</tr>
<tr>
<td>woodland</td>
<td>Synonym</td>
<td>Environment Ontology</td>
</tr>
<tr>
<td>woodland</td>
<td>Synonym</td>
<td>envo_153</td>
</tr>
</tbody>
</table>
Preferred Name (Preferred_Name): Wood

Definitions (DEFINITION)
The hard, fibrous substance composing most of the stem and branches of a tree or shrub, and lying beneath the bark; the xylem.

Full Id:
http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#Wood

Alt Definition: Fibrous plant material under the bark that is created by lateral cell division from the vascular cambium. Noted for high content cellulose, hemicellulose, and lignin in the cell walls. FDA
1. Who am I?
2. How to find your data
3. How to do biology across the genome
4. How to extend the GO methodology to clinical and translational medicine
5. Anatomy Ontologies: An OBO Foundry success story
6. The Infectious Disease Ontology
7. The Environment Ontology
<table>
<thead>
<tr>
<th><strong>Species name</strong></th>
<th>Staphylococcus aureus Rosenbach 1884 AL</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strain number</strong></td>
<td>B141</td>
</tr>
<tr>
<td><strong>Other collections number</strong></td>
<td>See other collection numbers</td>
</tr>
<tr>
<td><strong>Original substract</strong></td>
<td>Swimming pool water</td>
</tr>
<tr>
<td><strong>Country of origin</strong></td>
<td>Morocco</td>
</tr>
<tr>
<td><strong>Collector name</strong></td>
<td>Benabdellah</td>
</tr>
<tr>
<td><strong>Isolator name</strong></td>
<td>ONEP</td>
</tr>
<tr>
<td><strong>Determinator name</strong></td>
<td>ONEP (Staphylococcus aureus subsp.aureus)</td>
</tr>
<tr>
<td><strong>Redeterminator</strong></td>
<td>LMG</td>
</tr>
<tr>
<td><strong>Name as redetermined</strong></td>
<td>Staphylococcus aureus subsp.aureus (Staphylococcus aureus).</td>
</tr>
<tr>
<td><strong>Depositor name</strong></td>
<td>ONEP</td>
</tr>
<tr>
<td><strong>History</strong></td>
<td>&lt;=2000, S. Benabdellah ONEP Rabat (Staphylococcus aureus). (19</td>
</tr>
<tr>
<td><strong>Media and temperature</strong></td>
<td>Details 37°C</td>
</tr>
<tr>
<td>Species name</td>
<td>Staphylococcus aureus Rosenbach 1884 AL</td>
</tr>
<tr>
<td>------------------------------</td>
<td>-----------------------------------------</td>
</tr>
<tr>
<td>Strain number</td>
<td>B13</td>
</tr>
<tr>
<td><strong>Other collections number</strong></td>
<td>See other collection numbers</td>
</tr>
<tr>
<td>Original substract</td>
<td>Sputum</td>
</tr>
<tr>
<td>Country of origin</td>
<td>Morocco</td>
</tr>
<tr>
<td>Collector name</td>
<td>A. Alaoui</td>
</tr>
<tr>
<td>Isolator name</td>
<td>FM Rabat</td>
</tr>
<tr>
<td>Determinator name</td>
<td>FM Rabat</td>
</tr>
<tr>
<td>Redeterminator</td>
<td>LMG</td>
</tr>
<tr>
<td>Name as redetermined</td>
<td>Staphylococcus aureus subsp.aureus</td>
</tr>
<tr>
<td>Depositor name</td>
<td>FM Rabat</td>
</tr>
<tr>
<td>History</td>
<td>&lt;2000, A. Alaoui FM Rabat (Staphylococcus aureus). (1999), hôpital human, sputum</td>
</tr>
<tr>
<td>Media and temperature</td>
<td>37°C</td>
</tr>
</tbody>
</table>
**Borrelia burgdorferi**

**Isolation:** tick, *Ixodes dammini*, New York

<table>
<thead>
<tr>
<th>Organism:</th>
<th>Borrelia burgdorferi</th>
</tr>
</thead>
<tbody>
<tr>
<td>Designations:</td>
<td>031</td>
</tr>
<tr>
<td>Depositor:</td>
<td>AG Burgdorfer</td>
</tr>
<tr>
<td>Biosafety Level:</td>
<td>2</td>
</tr>
<tr>
<td>Growth Conditions:</td>
<td>ATCC medium 1914: Revised BS K medium Temperature: 37°C</td>
</tr>
</tbody>
</table>

Permits/Forms: In addition to the MTA mentioned above, other ATCC and/or regulatory permits may be required for the transfer of this ATCC material. Anyone purchasing ATCC material is ultimately responsible for obtaining the permits. Please [click here](#) for information regarding the specific requirements for shipment to your location.

Cross References:
- GenBank: M90084: Borrelia burgdorferi 22 kD antigen.
- GenBank: M34710: Borrelia burgdorferi flagellin (fla) gene, 5' end.
- GenBank: M52203: Borrelia burgdorferi small subunit 16S rDNA sequence.
- GenBank: U76405: Borrelia burgdorferi vmp-like VbsE (vbsE1) gene, complete CDS.
- GenBank: S75873: 60 kDa antigen [Borrelia coriaceae, C653, ATCC 35210, Genomic, 1873 nt].
- GenBank: U01894: Borrelia burgdorferi 031 outer surface protein C (ospC) gene, complete cDS.
- GenBank: U78406: Borrelia burgdorferi putative VLS recombination cassettes
  - Vis2-Vis1b (vis) gene, complete sequence.

Type Strain: yes (0669) [32662] [49799]
The Infinite Monkey (Fortuitous Interoperability) strategy to resolve data silos
How to find your data?

How to find and integrate other people’s data?

How to reason with data when you find it?

How to understand the significance of the data you collected 3 years earlier?

Part of the solution must involve consensus-based, standardized terminologies and coding schemes.
NIH Mandates for Sharing of Research Data

Investigators submitting an NIH application seeking $500,000 or more in any single year are expected to include a plan for data sharing

(http://grants.nih.gov/grants/policy/data_sharing)
Making data (re-) usable through standards

• Standards provide
  – common structure and terminology
  – single data source for review (less redundant data)

• Standards allow
  – use of common tools and techniques
  – common training
  – single validation of data
Problems with standards

• Standards involve considerable costs of re-tooling, maintenance, training, ...
• Not all standards are of equal quality
• Bad standards create lasting problems
Ontology success stories, and some reasons for failure

Linked Open Data in the Semantic Web
Term Search
Search for a term across multiple ontologies

obesity

Select ontologies to search
Type here to select ontologies or leave blank to use all

Search

<table>
<thead>
<tr>
<th>TERM NAME</th>
<th>MATCHED IN</th>
<th>ONTOLOGY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obesity</td>
<td>Preferred Name</td>
<td>ICD10</td>
</tr>
<tr>
<td>Obesity</td>
<td>Preferred Name</td>
<td>Common Terminology Criteria for Adverse Events</td>
</tr>
<tr>
<td>obesity</td>
<td>Preferred Name</td>
<td>Neuro Behavior Ontology</td>
</tr>
<tr>
<td>obesity</td>
<td>Preferred Name</td>
<td>CRISP Thesaurus, 2006</td>
</tr>
<tr>
<td>Obesity</td>
<td>Preferred Name</td>
<td>Read Codes, Clinical Terms Version 3 (CTV3)</td>
</tr>
<tr>
<td>obesity</td>
<td>Preferred Name</td>
<td>Bleeding History Phenotype</td>
</tr>
<tr>
<td>obesity</td>
<td>Preferred Name</td>
<td>eVOC (Expressed Sequence Annotation for Humans)</td>
</tr>
<tr>
<td>Obesity</td>
<td>Preferred Name</td>
<td>NCI Thesaurus</td>
</tr>
<tr>
<td>Obesity</td>
<td>Preferred Name</td>
<td>Galen</td>
</tr>
<tr>
<td>Obesity</td>
<td>Preferred Name</td>
<td>NIF Dysfunction</td>
</tr>
</tbody>
</table>

etc.
The more ontology is successful, the more it fails

- As ontologies (controlled vocabularies) become easier to create, and to use
- more and more ontologies are constructed
- thereby recreating the very silo problems ontologies were designed to solve

How to solve this problem?
1. Who am I?
2. How to find your data
3. **How to do biology across the genome**
4. How to extend the GO methodology to clinical and translational medicine
5. Anatomy Ontologies: An OBO Foundry success story
6. The Infectious Disease Ontology
7. The Environment Ontology
How to do biology across the genome?
Biomedical Ontology in PubMed
By far the most successful: GO (Gene Ontology)
GO:0008152 : metabolism (34935)
  ▪ GO:0006066 : alcohol metabolism (1043)
  ▪ GO:0006081 : aldehyde metabolism (67)
  ▪ GO:0009308 : amine metabolism (2232)
  ▪ GO:0006519 : amino acid and derivative metabolism (2320)
    ▪ GO:0006575 : amino acid derivative metabolism (659)
      ▪ GO:0018902 : 1,3-dichloro-2-propanol metabolism (0)
      ▪ GO:0018871 : 1-aminocyclopropane-1-carboxylate metabolism (1)
      ▪ GO:0019471 : 4-hydroxyproline metabolism (35)
      ▪ GO:0046442 : aerobactin metabolism (0)
      ▪ GO:0042398 : amino acid derivative biosynthesis (351)
      ▪ GO:0042219 : amino acid derivative catabolism (70)
      ▪ GO:0009448 : aminobutyrate metabolism (14)
      ▪ GO:0006576 : biogenic amine metabolism (278)
      ▪ GO:0009692 : ethylene metabolism (38)
      ▪ GO:0046516 : hypusine metabolism (7)
      ▪ GO:0046418 : nopaline metabolism (0)
      ▪ GO:0046419 : octopine metabolism (1)
biological process

physiological process

is_a

cellular process

is_a

cellular physiological process

is_a

cell cycle

is_a

meiotic cell cycle

is_a

M phase

part_of

M phase of meiotic cell cycle

part_of

cytokinesis

is_a

cytokinesis after meiosis I
the GO works through annotation of data.
three types of data

what cellular component?

what molecular function?

what biological process?
Gene Ontology Consortium

WormBase
Gramene
FlyBase
Rat Genome Database
DictyBase
Mouse Genome Database
The Arabidopsis Information Resource
The Zebrafish Information Network
Berkeley Drosophila Genome Project
Saccharomyces Genome Database
...

Benefits of GO

1. rooted in basic experimental biology
2. links people to data and to literature
3. links data to data
   • across species (human, mouse, yeast, fly ...)
   • across granularities (molecule, cell, organ, organism, population)
4. links medicine to biological science
5. cumulation of scientific knowledge in algorithmically tractable form
Sjöblöm T, et al. analyzed 13,023 genes in 11 breast and 11 colorectal cancers using functional information captured by GO identified 189 genes as being mutated at significant frequency and thus as providing targets for diagnostic and therapeutic intervention.

1. Who am I?
2. How to find your data
3. How to do biology across the genome
4. **How to extend the GO methodology to clinical and translational medicine: Open Biomedical Ontologies**
5. Anatomy Ontologies: An OBO Foundry success story
6. The Infectious Disease Ontology
7. The Environment Ontology
The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration

Barry Smith1, Michael Ashburner2, Corneliu Rosse3, Jonathan Bard4, William Bug5, Werner Ceusters6, Louis J Goldberg7, Karen Eilbeck8, Amelia Ireland9, Christopher J Mungall10, the OBI Consortium11, Neocles Leontis12, Philippe Rocca-Serra9, Alan Ruttenberg13, Susanna-Assunta Sansone9, Richard H Scheuermann14, Nigam Shah15, Patricia L Whetzel16 & Suzanna Lewis10

The value of any kind of data is greatly enhanced when it exists in a form that allows it to be integrated with other data. One approach to integration is through the annotation of multiple bodies of data using common controlled vocabularies or ‘ontologies’. Unfortunately, the very success of this approach has led to a proliferation of ontologies, which itself creates obstacles to integration. The Open Biomedical Ontologies (OBO) consortium is pursuing a strategy to overcome this problem. Existing OBO ontologies, including the Gene Ontology, are undergoing coordinated reform, and new ontologies are being created on the basis of an evolving set of shared principles governing ontology development. The result is an expanding family of ontologies designed to be interoperable and logically well formed and to incorporate accurate representations of biological reality. We describe this OBO Foundry initiative and provide guidelines for those who might wish to become involved.

In the search for what is biologically and clinically significant in the swarms of data being generated by today’s high-throughput technologies, a common strategy involves the creation and analysis of ‘annotations’ linking primary data to expressions in controlled, structured vocabularies, thereby making the data available to search and to algorithmic processing. The most successful such endeavor, measured both by numbers of users and by reach across species and granularities, is the Gene Ontology (GO). There exist over 11 million annotations relating gene products described in the UniProt, Ensembl and other databases to terms in the GO, of which half a million have been manually verified by specialist curators in different model-organism communities on the basis of the analysis of experimental results reported in 52,000 scientific journal articles (http://www.ebi.ac.uk/GOA/). Data related to some 180,000 genes have been manually annotated in this way, an endeavor now being refined and systematized within the Reference Genomes Project (US National Institutes of Health, National Human...
<table>
<thead>
<tr>
<th>Ontology</th>
<th>Scope</th>
<th>URL</th>
<th>Custodians</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell Ontology (CL)</td>
<td>cell types from prokaryotes to mammals</td>
<td>obo.sourceforge.net/cgi-bin/detail.cgi?cell</td>
<td>Jonathan Bard, Michael Ashburner, Oliver Hofman</td>
</tr>
<tr>
<td>Chemical Entities of Biological Interest (ChEBI)</td>
<td>molecular entities</td>
<td>ebi.ac.uk/chebi</td>
<td>Paula Dematos, Rafael Alcantara</td>
</tr>
<tr>
<td>Common Anatomy Reference Ontology (CARO)</td>
<td>anatomical structures in human and model organisms</td>
<td>(under development)</td>
<td>Melissa Haendel, Terry Hayamizu, Cornelius Rosse, David Sutherland,</td>
</tr>
<tr>
<td>Foundational Model of Anatomy (FMA)</td>
<td>structure of the human body</td>
<td>fma.biostr.washington.edu</td>
<td>JLV Mejino Jr., Cornelius Rosse</td>
</tr>
<tr>
<td>Functional Genomics Investigation Ontology (FuGO)</td>
<td>design, protocol, data instrumentation, and analysis</td>
<td>fugo.sf.net</td>
<td>FuGO Working Group</td>
</tr>
<tr>
<td>Gene Ontology (GO)</td>
<td>cellular components, molecular functions, biological processes</td>
<td><a href="http://www.geneontology.org">www.geneontology.org</a></td>
<td>Gene Ontology Consortium</td>
</tr>
<tr>
<td>Phenotypic Quality Ontology (PaTO)</td>
<td>qualities of anatomical structures</td>
<td>obo.sourceforge.net/cgi-bin/detail.cgi?attribute_and_value</td>
<td>Michael Ashburner, Suzanna Lewis, Georgios Gkoutos</td>
</tr>
<tr>
<td>Protein Ontology (PrO)</td>
<td>protein types and modifications</td>
<td>(under development)</td>
<td>Protein Ontology Consortium</td>
</tr>
<tr>
<td>Relation Ontology (RO)</td>
<td>relations</td>
<td>obo.sf.net/relationship</td>
<td>Barry Smith, Chris Mungall</td>
</tr>
<tr>
<td>RNA Ontology (RnaO)</td>
<td>three-dimensional RNA structures</td>
<td>(under development)</td>
<td>RNA Ontology Consortium</td>
</tr>
<tr>
<td>Sequence Ontology (SO)</td>
<td>properties and features of nucleic sequences</td>
<td>song.sf.net</td>
<td>35 Karen Eilbeck</td>
</tr>
<tr>
<td>RELATION TO TIME</td>
<td>CONTINUANT</td>
<td>OCCURRENT</td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>------------</td>
<td>-----------</td>
<td></td>
</tr>
<tr>
<td>GRANULARITY</td>
<td>INDEPENDENT</td>
<td>DEPENDENT</td>
<td></td>
</tr>
<tr>
<td>ORGAN AND ORGANISM</td>
<td>Organism (NCBI Taxonomy)</td>
<td>Anatomical Entity (FMA, CARO)</td>
<td>Organ Function (FMP, CPRO)</td>
</tr>
<tr>
<td>CELL AND CELLULAR COMPONENT</td>
<td>Cell (CL)</td>
<td><strong>Cellular Component</strong> (FMA, GO)</td>
<td>Cellular Function (GO)</td>
</tr>
<tr>
<td>MOLECULE</td>
<td>Molecule (ChEBI, SO, RnaO, PrO)</td>
<td><strong>Molecular Function</strong> (GO)</td>
<td></td>
</tr>
</tbody>
</table>
Community / Population Ontology

- family, clan
- ethnicity
- religion
- diet
- social networking
- education (literacy ...)
- healthcare (economics ...)
- household forms
- demography
- public health
- ...

37
<table>
<thead>
<tr>
<th>GRANULARITY</th>
<th>RELATION TO TIME</th>
<th>CONTINUANT</th>
<th>OCCURRENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORGAN AND ORGANISM</td>
<td></td>
<td>INDEPENDENT</td>
<td>DEPENDENT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Organism (NCBI Taxonomy)</td>
<td>Anatomical Entity (FMA, CARO)</td>
</tr>
<tr>
<td>CELL AND CELLULAR COMPONENT</td>
<td>Cell (CL)</td>
<td>Cellular Component (FMA, GO)</td>
<td>Cellular Function (GO)</td>
</tr>
<tr>
<td>MOLECULE</td>
<td>Molecule (ChEBI, SO, RnaO, PrO)</td>
<td>Molecular Function (GO)</td>
<td>Molecular Process (GO)</td>
</tr>
<tr>
<td>Granularity</td>
<td>Relation to Time</td>
<td>Continuant</td>
<td>Occurrent</td>
</tr>
<tr>
<td>-------------</td>
<td>-----------------</td>
<td>------------</td>
<td>-----------</td>
</tr>
<tr>
<td><strong>Organ and Organism</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Independent</td>
<td>Family, Community, Deme, Population</td>
<td>DEPENDENT</td>
<td>Phenotypic Quality (PaTO)</td>
</tr>
<tr>
<td>Dependent</td>
<td>Organ Function (FMP, CPRO)</td>
<td>Biological Process (GO)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Organism (NCBI Taxonomy)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Anatomical Entity (FMA, CARO)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cell and Cellular Component</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cell (CL)</td>
<td>Cell Component (FMA, GO)</td>
<td>Cellular Function (GO)</td>
<td></td>
</tr>
<tr>
<td><strong>Molecule</strong></td>
<td>Molecule (ChEBI, SO, RnaO, PrO)</td>
<td><strong>Molecular Function</strong> (GO)</td>
<td>Molecular Process (GO)</td>
</tr>
</tbody>
</table>

http://obofoundry.org
<table>
<thead>
<tr>
<th>Complex of Organisms</th>
<th>Continuant</th>
<th>Occurrent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Independent</td>
<td>Dependent</td>
</tr>
<tr>
<td>Organ and Organism</td>
<td>Organism (NCBI Taxonomy)</td>
<td>Anatomical Entity (FMA, CARO)</td>
</tr>
<tr>
<td>Cell and Cellular Component</td>
<td>Cell (CL)</td>
<td>Cellular Component (FMA, GO)</td>
</tr>
<tr>
<td>Molecule</td>
<td>Molecule (ChEBI, SO, RnaO, PrO)</td>
<td></td>
</tr>
<tr>
<td>Granularity</td>
<td>Continuant</td>
<td>Occurrent</td>
</tr>
<tr>
<td>-------------</td>
<td>------------</td>
<td>-----------</td>
</tr>
<tr>
<td><strong>Complex of Organisms</strong></td>
<td>Family, Community, Deme, Population</td>
<td>Population Phenotype, Population Process</td>
</tr>
<tr>
<td><strong>Organ and Organism</strong></td>
<td>Organism (NCBI Taxonomy), (FMA, CARO)</td>
<td>Organ Function (FMP, CPRO), Phenotypic Quality (PaTO), Biological Process (GO)</td>
</tr>
<tr>
<td><strong>Cell and Cellular Component</strong></td>
<td>Cell (CL), Cell Component (FMA, GO)</td>
<td>Cellular Function (GO)</td>
</tr>
<tr>
<td><strong>Molecule</strong></td>
<td>Molecule (ChEBI, SO, RnaO, PrO)</td>
<td>Molecular Function (GO), Molecular Process (GO)</td>
</tr>
</tbody>
</table>

http://obofoundry.org
<table>
<thead>
<tr>
<th>RELATION TO TIME</th>
<th>CONTINUANT</th>
<th>INDEPENDENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRANULARITY</td>
<td></td>
<td></td>
</tr>
<tr>
<td>COMPLEX OF ORGANISMS</td>
<td>Family, Community, Dème, Population</td>
<td>Environment of population</td>
</tr>
<tr>
<td>ORGAN AND ORGANISM</td>
<td>Organism (NCBI Taxonomy) (FMA, CARO)</td>
<td>Environment of single organism</td>
</tr>
<tr>
<td>CELL AND CELLULAR COMPONENT</td>
<td>Cell (CL) Cell Component (FMA, GO)</td>
<td>Environment of cell</td>
</tr>
<tr>
<td>MOLECULE</td>
<td>Molecule (ChEBI, SO, RnaO, PrO)</td>
<td>Molecular environment</td>
</tr>
</tbody>
</table>

http://obofoundry.org
<table>
<thead>
<tr>
<th>RELATION TO TIME</th>
<th>CONTINUANT</th>
<th>INDEPENDENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>GRANULARITY</td>
<td></td>
<td></td>
</tr>
<tr>
<td>COMPLEX OF ORGANISMS</td>
<td>Family, Community, Deme, Population</td>
<td>Environment of population</td>
</tr>
<tr>
<td>ORGAN AND ORGANISM</td>
<td>Organism (NCBI Taxonomy)</td>
<td>Environment of single organism*</td>
</tr>
<tr>
<td></td>
<td>(FMA, CARO)</td>
<td></td>
</tr>
</tbody>
</table>

* The sum total of the conditions and elements that make up the surroundings and influence the development and actions of an individual.
<table>
<thead>
<tr>
<th>GRANULARITY</th>
<th>CONTINUANT</th>
<th>OCCURRENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>RELATION TO TIME</td>
<td></td>
<td></td>
</tr>
<tr>
<td>COMPLEX OF ORGANISMS</td>
<td>Family, Community, Dème,</td>
<td>Population Process</td>
</tr>
<tr>
<td></td>
<td>Population</td>
<td></td>
</tr>
<tr>
<td>ORGAN AND ORGANISM</td>
<td>Organism (NCBI Taxonomy)</td>
<td>Phenotypic Quality (PaTO)</td>
</tr>
<tr>
<td></td>
<td>(FMA, CARO)</td>
<td></td>
</tr>
<tr>
<td>CELL AND CELLULAR</td>
<td>Cell (CL)</td>
<td>Plant Growth and</td>
</tr>
<tr>
<td>COMPONENT</td>
<td>Cell Component (FMA, GO)</td>
<td>Developmental Stage</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(GO)</td>
</tr>
<tr>
<td>MOLECULE</td>
<td>Molecule (ChEBI, SO,</td>
<td>Molecular Process (GO)</td>
</tr>
<tr>
<td></td>
<td>RnaO, PrO)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Molecular Function (GO)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

http://obofoundry.org
Goal of the OBO Foundry

all biomedical research data should cumulate to form a single, algorithmically processable, whole

The ontology is open and available to be used by all.

The ontology is instantiated in, a common formal language and shares a common formal architecture.

The developers of the ontology agree in advance to collaborate with developers of other OBO Foundry ontology where domains overlap.
CRITERIA

- The developers of each ontology commit to its maintenance in light of scientific advance, and to soliciting community feedback for its improvement.

- They commit to working with other Foundry members to ensure that, for any particular domain, there is community convergence on a single controlled vocabulary.
The OBO Foundry is a collaborative experiment involving developers of science-based ontologies who are establishing a set of principles for ontology development with the goal of creating a suite of orthogonal interoperable reference ontologies in the biomedical domain. The groups developing ontologies who have expressed an interest in this goal are listed below, followed by other relevant efforts in this domain.

In addition to a listing of OBO ontologies, this site also provides a statement of the OBO Foundry principles, discussion fora, technical infrastructure, and other services to facilitate ontology development. We welcome feedback and encourage participation.

Click any column header to sort the table by that column. The paw is link to the term request trackers for the listed ontologies.

### OBO Foundry ontologies

<table>
<thead>
<tr>
<th>Title</th>
<th>Domain</th>
<th>Prefix</th>
<th>File</th>
<th>Last changed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biological process</td>
<td>biological process</td>
<td>GO</td>
<td>gene_ontology_edit.obo</td>
<td>2012/02/05</td>
</tr>
<tr>
<td>Cellular component</td>
<td>anatomy</td>
<td>GO</td>
<td>gene_ontology_edit.obo</td>
<td>2012/02/05</td>
</tr>
<tr>
<td>Chemical entities of biological interest</td>
<td>biochemistry</td>
<td>CHEBI</td>
<td>chebi.obo</td>
<td>2012/01/05</td>
</tr>
<tr>
<td>Molecular function</td>
<td>biological function</td>
<td>GO</td>
<td>gene_ontology_edit.obo</td>
<td>2012/02/10</td>
</tr>
<tr>
<td>Phenotypic quality</td>
<td>phenotype</td>
<td>PATO</td>
<td>quality.obo</td>
<td>2012/02/05</td>
</tr>
<tr>
<td>Protein Ontology (PRO)</td>
<td>proteins</td>
<td>PR</td>
<td>pro.obo</td>
<td></td>
</tr>
<tr>
<td>Xenopus anatomy and development</td>
<td>anatomy</td>
<td>XAO</td>
<td>xenopus_anatomy.obo</td>
<td>2011/10/07</td>
</tr>
<tr>
<td>Zebrafish anatomy and development</td>
<td>anatomy</td>
<td>ZFA</td>
<td>zebrafish_anatomy.obo</td>
<td>2011/12/15</td>
</tr>
</tbody>
</table>

### OBO Foundry candidate ontologies and other ontologies of interest

<table>
<thead>
<tr>
<th>Title</th>
<th>Domain</th>
<th>Prefix</th>
<th>File</th>
<th>Last changed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adverse Event Reporting Ontology</td>
<td>health</td>
<td>AERO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amphibian gross anatomy</td>
<td>anatomy</td>
<td>AAO</td>
<td>AAO_v2_edit.obo</td>
<td></td>
</tr>
<tr>
<td>Amphibian taxonomy</td>
<td>anatomy</td>
<td>ATO</td>
<td>amphibian_taxonomy.obo</td>
<td></td>
</tr>
<tr>
<td>Anatomical Entity Ontology</td>
<td>anatomy</td>
<td>AEO</td>
<td>aeo.obo</td>
<td>2011/12/15</td>
</tr>
<tr>
<td>Ascomycete phenotype ontology</td>
<td>phenotype</td>
<td>APO</td>
<td>ascomycete_phenotype.obo</td>
<td>2011/09/30</td>
</tr>
<tr>
<td>Basic Formal Ontology</td>
<td>upper</td>
<td>BFO</td>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td>Bilateria anatomy</td>
<td>anatomy</td>
<td>BILA</td>
<td>bilateria_mrca.obo</td>
<td></td>
</tr>
</tbody>
</table>
Current OBO Foundry Ontologies

- Biological process (GO)
- Cellular component (GO)
- Chemical entities of biological interest
- Molecular function (GO)
- Phenotypic quality
- PRotein Ontology (PRO)
- Xenopus Anatomy and Development
- Zebrafish Anatomy and Development
Foundry ontologies under review

Cell Ontology (CL)
Infectious Disease Ontology (IDO)
Ontology for Biomedical Investigations (OBI)
Plant Ontology (PO)
Ontologies under construction

Allergy Ontology
Environment Ontology (EnvO)
Immunology Ontology (IDO)
Mental Functioning Ontology (MFO)
  Emotion Ontology (MFO-EM)
  Pain Ontology
Mental Disease Ontology (MDO)
Neurological Disease Ontology (ND)
Vaccine Ontology
1. Who am I?
2. How to find your data
3. How to do biology across the genome
4. How to extend the GO methodology to clinical and translational medicine
5. An OBO Foundry success story
6. The Infectious Disease Ontology
7. The Environment Ontology
Anatomy Ontologies

Fish Multi-Species Anatomy Ontology (NSF funding received)
Ixodidae and Argasidae (Tick) Anatomy Ontology
Mosquito Anatomy Ontology (MAO)
Spider Anatomy Ontology (SPD)
Xenopus Anatomy Ontology (XAO)

*undergoing reform:* Drosophila and Zebrafish Anatomy Ontologies
Ontologies facilitate grouping of annotations

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>brain</strong></td>
<td>20</td>
</tr>
<tr>
<td><strong>hindbrain</strong></td>
<td>15</td>
</tr>
<tr>
<td><strong>rhombomere</strong></td>
<td>10</td>
</tr>
</tbody>
</table>

Query brain without ontology 20
Query brain with ontology 45
Basic Formal Ontology (Top Level)

- Continuant
  - Independent Continuant
  - Dependent Continuant
- Occurrent
  - Process
  - Stage

Anatomical Structure

Quality

http://www.ifomis.org/bfo/
OBO Foundry organized in terms of Basic Formal Ontology

through the methodology of downward population
Each Foundry ontology can be seen as an extension of a single upper level ontology (BFO)
Example: The Cell Ontology

SUBCLASS EXPLORER

For Project: DC_CL

Asserted Hierarchy

owl:Thing

- Entity
  - Continuant
    - DependentContinuant
    - IndependentContinuant
      - FiatObjectPart
  - Object
    - Biological_Macromolecule
  - Cell
    - CD11c_Low__Plasmacytoid_Dendritic_Cell
    - CD11c_Negative_Plasmacytoid_Dendritic_Cell
    - Conventional_Dendritic_Cell
      - CD8_alpha_Neg_CD11b_Neg_Dendritic_Cell
        - Immature_CD8_alpha_Neg_CD11b_Neg_Dendritic_Cell
        - Mature_CD8_alpha_Neg_CD11b_Neg_Dendritic_Cell
Continuant

Independent Continuant

Dependent Continuant

Quality

Disposition
depends_on

Continuant

Independent Continuant
thing

Dependent Continuant
quality

Occurrent
process, event

temperature depends on bearer
the universal red

this particular case of redness (of a particular fly eye)

instantiates

the universal eye

instantiates

depends_on

an instance of eye (in a particular fly)

Phenotype Ontology (PATO)
the particular case of redness (of a particular fly eye) depends on an instance of an eye (in a particular fly)
This portion of H₂O instantiates at t₁, portion of ice.

This portion of H₂O instantiates at t₂, portion of liquid water.

This portion of H₂O instantiates at t₃, portion of gas.

Phase transitions
plant

zygote

embryo

seed

Phase transitions

instantiates at $t_1$

instantiates at $t_2$

instantiates at $t_3$

this plant
In nature, no sharp boundaries here.

John (exists continuously)
John’s temperature (exists continuously)

In nature, no sharp boundaries here.

37°C instantiates at t₁
37.1°C instantiates at t₂
37.2°C instantiates at t₃
37.3°C instantiates at t₄
37.4°C instantiates at t₅
37.5°C instantiates at t₆
John’s coronary heart disease (exists continuously)
1. Who am I?
2. How to find your data
3. How to do biology across the genome
4. How to extend the GO methodology to clinical and translational medicine
5. Anatomy Ontologies: An OBO Foundry success story
6. IDO: The Infectious Disease Ontology
7. The Environment Ontology
We have data

TBDB: Tuberculosis Database, including Microarray data
VFDB: Virulence Factor DB
TropNetEurop Dengue Case Data
ISD: Influenza Sequence Database at LANL
PathPort: Pathogen Portal Project
...

We need to annotate this data to allow retrieval and integration of
- sequence and protein data for pathogens
- case report data for patients
- clinical trial data for drugs, vaccines
- epidemiological data for surveillance, prevention
- ...

Goal: to make data deriving from different sources comparable and computable
IDO needs to work with

Disease Ontology (DO) + SNOMED CT
Gene Ontology Immunology Branch
Phenotypic Quality Ontology (PATO)
Protein Ontology (PRO)
Sequence Ontology (SO)
...

We need common controlled vocabularies to describe these data in ways that will assure comparability and cumulation.

What content is needed to adequately cover the infectious domain?

- Host-related terms (e.g. carrier, susceptibility)
- Pathogen-related terms (e.g. virulence)
- Vector-related terms (e.g. reservoir,)
- Terms for the biology of disease pathogenesis (e.g. evasion of host defense)
- Population-level terms (e.g. epidemic, endemic, pandemic,)
IDO Qualities

- quality of host
  - case of infectious disease
  - co-infected
  - immunity
  - immunocompromised
  - immunosuppressed
  - infected
  - refractoriness
  - resistance
  - susceptibility
  - swollen
  - temperature

- quality of host population
  - herd immunity
  - infectious disease
  - prevalence

- quality of infectious disease
  - endemic
  - incidence

- quality of pathogen
  - contagious
  - dormancy
  - drug resistance
  - endogenous
  - exogenous
  - invasive
  - latency
  - pathogenicity
  - tropism
  - virulence
  - zoonotic
IDO Roles

- **role**
  - colonizer
  - factor
    - adhesion factor
    - colonization factor
  - virulence factor
    - invasin
    - toxin
      - endotoxin
      - exotoxin
    - fusion protein
  - host
    - carrier
    - mobile genetic element
    - mode of transmission
  - pathogen
  - pathogenicity island
  - reservoir
  - route of entry
  - sign
  - symptom
  - treatment
  - vector
IDO provides a common template

IDO contains terms (like ‘pathogen’, ‘vector’, ‘host’) which apply to organisms of all species involved in infectious disease and its transmission.

Disease- and organism-specific ontologies built as refinements of the IDO core.
Disease-specific IDO test projects

MITRE, Mount Sinai, UT Southwestern – Influenza
  – Stuart Sealfon, Joanne Luciano,

IMBB/VectorBase – Vector borne diseases (A. gambiae, A. aegypti, I. scapularis, C. pipiens, P. humanus)
  – Kristos Louis

Colorado State University – Dengue Fever
  – Saul Lozano-Fuentes

Duke – Tuberculosis
  – Carol Dukes-Hamilton

Cleveland Clinic – Infective Endocarditis
  – Sivaram Arabandi

University of Michigan – Brucellosis
  – Yongqun He
1. Who am I?
2. How to find your data
3. How to do biology across the genome
4. How to extend the GO methodology to clinical and translational medicine
5. Anatomy Ontologies: An OBO Foundry success story
6. The Infectious Disease Ontology
7. The Environment Ontology
<table>
<thead>
<tr>
<th>Complex of Organisms</th>
<th>Continuant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biome / biotope, territory, habitat, neighborhood, ...</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Organ and Organism</th>
<th>Environment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Work environment, home environment; host/symbiont environment; ...</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cell and Cellular Component</th>
<th>Environment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extracellular matrix; chemokine gradient; ...</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Molecule</th>
<th>Environment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hydrophobic surface; virus localized to cellular substructure; active site on protein; pharmacophore ...</td>
<td></td>
</tr>
</tbody>
</table>

http://obofoundry.org
The Environment Ontology

OBO Foundry
Genomic Standards Consortium
National Environment Research Council (UK)
USDA, Gramene, J. Craig Venter Institute ...
Applications of EnvO in biology

- Support the annotation of meta-data related to:
  - Data about biological samples produced from various technologies
    - Metagenomics, Metabolomics, Proteomics, Transcriptomics, Genomics...
  - Data produced from remote sensing equipment
  - Images
    - Web 2.0, tagging
  - Physical holdings
    - Museum artifacts, (preserved) biological samples / organisms
  - ...anything that has an *environment*
How EnvO currently works for information retrieval

Retrieve all experiments on organisms obtained from:
- deep-sea thermal vents
- arctic ice cores
- rainforest canopy
- alpine melt zone

Retrieve all data on organisms sampled from:
- hot and dry environments
- cold and wet environments
- a height above 5,000 meters

Retrieve all the omic data from soil organisms subject to:
- moderate heavy metal contamination
extending EnvO to clinical and translational research

- we have public health, community and population data
- we need to make this data available for search and algorithmic processing
- we create a consensus-based ontology which can interoperate with ontologies for neighboring domains of medicine and basic biology
Environment = totality of circumstances external to a living organism or group of organisms

- pH
- evapotranspiration
- turbidity
- available light
- predominant vegetation
- predatory pressure
- nutrient limitation …
extend EnvO to the clinical domain

– dietary patterns (Food Ontology: FAO, USDA) ... allergies

– neighborhood patterns
  • built environment, living conditions
  • climate
  • social networking
  • crime, transport
  • education, religion, work
  • health, hygiene

– disease patterns
  • bio-environment (bacteriological, ...)
  • patterns of disease transmission (links to IDO)
# Grantees

**Obesity & the Built Environment**

<table>
<thead>
<tr>
<th>Institution</th>
<th>Project Title</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Columbia University of Health Sciences</strong></td>
<td>Obesity, Physical Activity &amp; Built Space in New York City</td>
</tr>
<tr>
<td><strong>Dartmouth College</strong></td>
<td>Environmental and Family Influences on Adolescent Overweight</td>
</tr>
<tr>
<td><strong>Georgia Institute of Technology</strong></td>
<td>Walking on Campus: Correlates &amp; Web Tools</td>
</tr>
<tr>
<td><strong>North Carolina State University</strong></td>
<td>Measuring Physical Activity Affordances in Preschool Outdoor Environments</td>
</tr>
<tr>
<td><strong>Oregon Research Institute</strong></td>
<td>Environmental Predictors of Elderly Obesity</td>
</tr>
<tr>
<td><strong>Pennsylvania State University - University Park</strong></td>
<td>Neighborhood Food Environment, Diet &amp; Health: Quasi-Experimental Study</td>
</tr>
<tr>
<td><strong>RAND Corporation</strong></td>
<td>Impact of Light Rail on Physical Activity &amp; BMI</td>
</tr>
<tr>
<td><strong>University of Massachusetts Medical School Worcester</strong></td>
<td>Obesity &amp; Neighborhood Characteristics</td>
</tr>
<tr>
<td><strong>University of Michigan at Ann Arbor</strong></td>
<td>Healthy Environments Partnership: Lean &amp; Green in Motown</td>
</tr>
<tr>
<td><strong>University of Texas - El Paso</strong></td>
<td>Physical Activity, Nutrition &amp; Built Environment in a Bi-National Border Setting</td>
</tr>
<tr>
<td><strong>University of Washington</strong></td>
<td>Child Weight Status, Physical Activity, &amp; Nutrition</td>
</tr>
</tbody>
</table>

**Funded by other NIH ICs**

- University of Illinois at Chicago
- University of Massachusetts Medical School Worcester
- University of Michigan at Ann Arbor
- University of Texas - El Paso
- University of Washington
- Health Empowerment Zones for People with Disabilities
with thanks to

**BFO**: Fabian Neuhaus (NIST), Melissa Haendel (Oregon), David Sutherland (Flybase)

**EnvO**: Dawn Field, Norman Morrison (NERC)

**FMA**: Cornelius Rosse, J. L. E. Mejino (Seattle)

**IDO**: Lindsay Cowell, Albert Goldfain (Dallas)

**OBO Foundry**: Michael Ashburner, Suzanna Lewis, Chris Mungall (Flybase, GO), Alan Ruttenberg (Buffalo, Neurocommons)

**NCBO**: NIH RFA-RM-04-022

**PRO**: NIH R01 GM080646-01

**PO**: The Plant Ontology Consortium