Introduction to ontologies

Melissa Haendel

Contributors: Melissa Haendel, Chris Mungall, David Osumi-Sutherland
Common controlled vocabularies indicate the same meaning under different annotation circumstances.

### MouseEcotope

<table>
<thead>
<tr>
<th>Tool</th>
<th>Functional Model</th>
<th>Correction for multiple experiments</th>
<th>GO Visualization</th>
<th>Microarray support</th>
<th>Time to process 200 genes</th>
</tr>
</thead>
<tbody>
<tr>
<td>GeneExpress</td>
<td><strong>c</strong> maternal, non-membrane proteins</td>
<td>Flat, Tree</td>
<td>17/38</td>
<td><strong>c</strong> maternal, non-membrane proteins</td>
<td>7.8, 10, 28</td>
</tr>
<tr>
<td><strong>GlyProt</strong></td>
<td>Alpha-2-macroglobulin</td>
<td>Flat, Tree</td>
<td>17/38</td>
<td><strong>c</strong> maternal, non-membrane proteins</td>
<td>7.8, 10, 28</td>
</tr>
<tr>
<td>EAEMouse</td>
<td><strong>c</strong> maternal, non-membrane proteins</td>
<td>Flat, Tree</td>
<td>17/38</td>
<td><strong>c</strong> maternal, non-membrane proteins</td>
<td>7.8, 10, 28</td>
</tr>
<tr>
<td>NeuGene</td>
<td><strong>c</strong> maternal, non-membrane proteins</td>
<td>Flat, Tree</td>
<td>17/38</td>
<td><strong>c</strong> maternal, non-membrane proteins</td>
<td>7.8, 10, 28</td>
</tr>
<tr>
<td><strong>DiabetInGene</strong></td>
<td><strong>c</strong> maternal, non-membrane proteins</td>
<td>Flat, Tree</td>
<td>17/38</td>
<td><strong>c</strong> maternal, non-membrane proteins</td>
<td>7.8, 10, 28</td>
</tr>
<tr>
<td><strong>GluChem</strong></td>
<td><strong>c</strong> maternal, non-membrane proteins</td>
<td>Flat, Tree</td>
<td>17/38</td>
<td><strong>c</strong> maternal, non-membrane proteins</td>
<td>7.8, 10, 28</td>
</tr>
</tbody>
</table>

### sphingolipid transporter activity

- MouseEcotope: c maternal, non-membrane proteins
- GlyProt: Alpha-2-macroglobulin
- DiabetInGene: c maternal, non-membrane proteins
- GluChem: c maternal, non-membrane proteins
What is a controlled vocabulary?

Any closed, prescribed list of terms used for classifying data.

Key Features:
- Terms are not usually defined.
- Relationships between the terms are not usually defined.
- Can be a list.

Here is a CV of wines:

*Pinot noir, red, chardonnay, Chianti, Bordeaux, Riesling*….

These are all different types - color, location, varietal, and are present in a list.

Another example would be the map locations list at the end of your Gazeteer.
What is a Taxonomy?
Any controlled vocabulary that is arranged in a hierarchy.

Key Features:
• Terms are not usually defined.
• Relationships between the terms are not usually defined.
• Terms are arranged in a hierarchy.

Here is a wine taxonomy:

- Wine
  - Red
    - merlot
    - zinfandel
    - cabernet
    - pinot noir
  - White
    - chardonnay
    - pinot gris
    - Riesling
What is a Thesaurus?

A taxonomy that contains additional information about use of the terms

Key Features:
• Terms are not usually defined.
• Relationships between the terms are not usually defined.
• Terms are arranged in a hierarchy.
• Statements about the terms are included such as scope notes or instructions for use.

Some well known thesauri are:
WordNet, NCI cancer thesaurus, MeSH
What is an ontology?

A formal conceptualization of a specified domain of interest.

Key Features:
• Terms are defined.
• Relationships between the terms are defined, allowing logical inference.
• Terms are arranged in a hierarchy.
• Expressed in a knowledge representation language such as RDFS, OBO or OWL.

Some well known ontologies are:
Foundational Model of Anatomy, Gene Ontology, Linnean Taxonomy of species.

Reproduced with permission, Jason Freeny
The Ontology spectrum:

Bottom line: you get what you pay for.
A common misconception
Are ontologies about terms or things?

When you are arguing about including something in your ontology,

- Are you arguing about what a term means?
- Or are you arguing about what term should be adopted in your ontology language to represent a well-characterized entity or concept?

These are terminological questions, and not ontological questions. 1 is a purely linguistic dispute; 2 is primarily a practical question.

The ontological questions are:

What kind of things should we recognize in our ontology? (Never mind, for the moment, what we might choose to call them.)

What are their relations to one another? (Not: What are the relations of their terms/names to one another?)

Adapted from Gary Merrill
Types, subtypes, and instances

Subtyping relation

entity \( \text{is}_a \) organism

\( \text{is}_a \) animal

\( \text{is}_a \) mammal

\( \text{is}_a \) human

\( \text{instance}_\text{of} \) Peanut

\( \text{instance}_\text{of} \) Chris Shaffer

\( \text{is}_a = \text{SubClassOf} \)
How do you tell if it is an instance or a class?

- Is there more than one in existence?
- Is the entity referencing a group of things with common properties?

Class or instance?
- There is only one Snoopy
- There is a class of things labeled “Snoopy toys”

Class or instance?
- There is only one Alaska
- There is a class of things labeled “States”

Class or instance?
- There is only one Jeffrey pine cone in my specimen collection
- There is a class of things labeled “Jeffrey Pine Cones”
General Principle for Logical Definitions

Definitions are of the following Genus-Differentia form:

\[ X = \text{a } Y \text{ which has one or more differentiating characteristics.} \]

where \( X \) is the is\_a parent of \( Y \).

---

**Definition:** Blue cylinder = Cylinder that has color blue.

**Definition:** Red cylinder = Cylinder that has color red.

**Definition:** cylinder = Surface formed by the set of lines perpendicular to a plane, which pass through a given circle in that plane.
The True Path Rule
The pathway from a subClass all the way up to its top level parent(s) must be universally true.

**GO Before:**
- cuticle synthesis
  - [i] chitin metabolism
- cell wall biosynthesis
  - [i] chitin metabolism
    - [i] chitin biosynthesis
    - [i] chitin catabolism

**BUT:** A fly chitin synthase gene could be annotated to chitin biosynthesis, and appear in a query for genes annotated to cell wall biosynthesis (and its children), which makes no sense because flies don't have cell walls.

**GO After:**
- chitin metabolism
  - [i] chitin biosynthesis
  - [i] chitin catabolism
  - [i] cuticle chitin metabolism
    - [i] cuticle chitin biosynthesis
    - [i] cuticle chitin catabolism
  - [i] cell wall chitin metabolism
    - [i] cell wall chitin biosynthesis
    - [i] cell wall chitin catabolism

**NOW:** all the subClass terms can be followed up to chitin metabolism, but cuticle chitin metabolism terms do not trace back to cell wall terms, so all the paths are true.
Where does the True Path Rule come from?

Transitivity.
Some relations are transitive, and apply across all levels of the hierarchy.

For example, a cat is_a mammal, and a mammal is_a vertebrate
 SO
 a cat is_a vertebrate
=> This is the true path rule and is because the is_a relation is transitive.

Some properties are not transitive.

For example, head has_quality round.
 and,
 head part_of organism.
 So is the organism round? Of course not.

BUT, eyes are part_of head, and head part_of organism, SO eye part_of organism is true, because part_of is a tranistive relation.

Relations are logically defined in a common relation ontology or within each ontology that uses them.
Relationships and definitions

A relationship from one class to another is a formalized part of its definition.

A subtype relation (is_a in OBO, SubClassOf in OWL) specifies **necessary** conditions for membership of a class.

For example, finger *part_of* hand (all finger *part_of* some hand) states that a necessary condition of being in the class finger is to be part of some hand.

So... if a finger exists, it is part of some hand. But...this does not mean that if a hand exists, it has as a part a finger.
About reasoners

A piece of software able to infer logical consequences from a set of asserted facts or axioms.

They are used to check the logical consistency of the ontologies and to extend the ontologies with "inferred" facts or axioms

For example, a reasoner would infer:

Major premise: All mortals die.
Minor premise: Some men are mortals.
Conclusion: Some men die.
Classifying anatomy

Subclasses are like subsets
Relationships record classifications too

‘floral organ’ SubClassOf part_of some ‘flower’
The knowledge in an ontology can make the reasons for classification explicit

- These are necessary and sufficient conditions, also called an equivalent class axiom
Many perspectives, many ontologies – that overlap in content.

- Gross anatomy
- Tissues
- Cells
- Cell anatomy
- Proteins
- Chemical entities
- Vascular system
- Evolutionary characters
- Pathologies
- Physiological processes
- Development
- Cellular processes
- Reactions
- Processes
Using multiple ontologies

- Identify key points of integration between ontologies

- Invest energy in understanding what is out there, i.e. seek to import and reuse, rather than “aligning” later

- Modularize based on domain or taxon

- Let the reasoner help do the work

- Work together to distribute work
Example of a post-composed anatomical entity

*Plasma membrane of spermatocyte*

- *Plasma membrane* [GO CC]
- *Spermatocyte* [Cell Ontology]

Genus

- *a plasma membrane which is part_of a spermatocyte*

Differentia

- Gene Ontology
- Basic Formal Ontology
- Cell Ontology
Entity-quality model of phenotype representation

Entity

- Sepal

Quality

- Curved

Instance_of

Inheres_in
Pre vs. post-composition of phenotypes

RO/BFO
(awn and 'is bearer of' some White) and (part_of some lemma)

OR
(awn color) and (part_of some lemma)

TO
PO

PO
Lists, trackers, ontologies, annotation, oh my!

- Term requested
- Term discussed by community
- Term needed for annotation
- Tracker IDs can be in ontology metadata
- Design and requirements analysis
- Term brokers are being developed to create temp classes during ontology editing or annotation
- Design documents comment on existing ontologies
- Trackers are often autoemailled to integrated listserves

**Ontology Edited**

- Term requested
- Tracker IDs can be in ontology metadata
- Term brokers are being developed to create temp classes during ontology editing or annotation
- Design documents comment on existing ontologies
- Term discussed by community
- Term needed for annotation
- Design and requirements analysis
- Trackers are often autoemailled to integrated listserves
What is a tracker?

- A tracker is a place to put a formal ontology request.

- Trackers have long been used in the software community for keeping track of bugs, feature requests, etc.

- In the ontology community, they are quite valuable because they provide a documented, structured requests for changes or additions.

- Tracker IDs can be referenced in ontology metadata—such as in an editor note or definition annotation.
How do you write a tracker request?

- It is important that when you make a tracker request, you provide as much information as possible, in order to facilitate the change you are requesting and future reference.

- For new terms, or term rearrangements, provide the intended hierarchy – both SubClass as well as any other relations required (such as partonomy).

- Provide text definitions, that make sense in the Genus Differentia context, for all new or edited terms.

- Provide attribution for the definitions.

- Some commentary may occur on the tracker item, but can sometimes lead to long listserve discussions before returning to a decision on the tracker.

- Complex issues requiring decision are best first discussed on the listserves or in design documents, but it’s always better to say something somewhere!
Example tracker request


ossification expansion and restructuring - ID: 3456359

Details: GO_REF:0000034 — Phenoscape Skeletal Anatomy Jamboree

Brian K. Hall (Dalhousie University), Matthew Viskarousky (Ontario Veterinary College, University of Guelph), David Blackburn, University of Kansas; Wasila Dahdul, University of South Dakota and NESCent; Alexander Diehl, Mouse Genome Informatics (MGI); Melissa Haendel, Oregon Health & Science University; John G. Lundberg, Department of Ichthyology, Academy of Natural Sciences, Philadelphia; Paula Mabee, Department of Biology, University of South Dakota; Martin Ringwald, Mouse Genome Informatics (MGI); Erik Segerdell, Oregon Health Sciences University; Geri Van Slyke, Zebrafish Information Network (ZFIN); Monte Westerfield, Zebrafish Information Network (ZFIN) and Institute of Neuroscience, University of Oregon., 2010.

Proposed new hierarchy:

GO:0001503 ossification
   └i-GO: new direct ossification
      └i-GO:0001957 intramembranous ossification — narrow syn dermal ossification
      └i-GO: new perichondral ossification
      └i-GO: new metaplastic ossification — related synonym metaplasia
      └i-GO: new replacement ossification — exact synonym indirect ossification
      └i-GO:0001958 endochondral ossification
     └i-GO: new ligamentous ossification
     └i-GO: new intratendonous ossification
     └i-GO:0043931 ossification involved in bone maturation
     └i-GO:0043932 ossification involved in bone remodeling
    └p-GO:0001649 osteoblast differentiation

Definitions:

GO:0001503 ‘ossification’
Current def: The formation of bone or of a bony substance, or the conversion of fibrous tissue or of cartilage into bone or a bony substance. (No change needed)

GO: new ‘direct ossification’
def: Ossification that does not require the replacement of preexisting tissues.

GO: new ‘replacement ossification’
def: Ossification that requires the replacement of a preexisting tissue prior to bone tissue formation.
   [exact synonym: Indirect ossification]

GO: new ‘ligamentous ossification’
def: Ossification wherein bone tissue forms within ligamentous tissue.
   COMMENT: Ligamentous ossification may occur via replacement ossification or metaplastic ossification or both in any one instance.

GO: new ‘intratendonous ossification’
def: Ossification wherein bone tissue forms within tendonous tissue.
Ontology documentation

Where does it happen?
Potentially too many places and at the same time, not enough!

- Commit messages. Example: https://code.google.com/p/cell-ontology/source/detail?r=44
- Releases and release notes. Example: http://obi-ontology.org/page/Releases/2012-07-01
- Internal documentation: Definitions, Definition source, Comments, etc. – documentation right in the ontology.
When to obsolete

Why are we discussing this in the communities section?

Because deprecating a class and the reasons for doing so are incredibly important to your community of users

- Deprecate = Obsolete
- Destroy = Delete

1. Has your ontology been made available to the public? If yes, consider obsoleting rather than deleting. (Note- keeping your ontology in GoogleCode is public!! Once an ID is out in the wild, it needs to be tracked.)

2. Has the text or logical definition of the class or property changed substantially? Remember, the ID is attached to the logical/non-logical definition, NOT the label. Changing a label does not require obsolescence.

3. Is the term confounded and needs to be split or merged into another class? Consider using the replaced_by or consider annotation properties on obsoleted entities to point users to the right new entities

4. Communicate ontology changes, in particular obsolescence, in release notes and VC commits
The OBO Foundry

http://www.obofoundry.org/

More than just a website, it’s a community of ontology developers dedicated to working together using common principles – they are there to help!