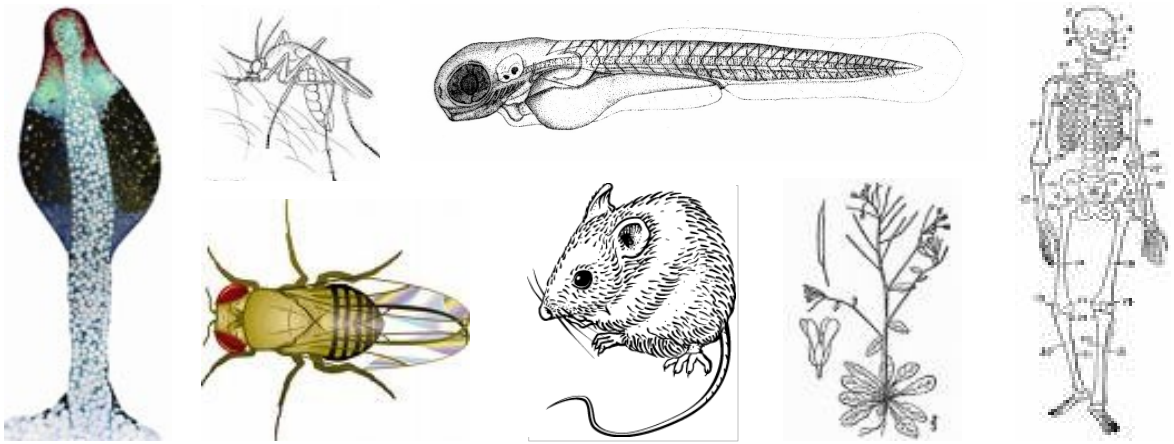
Introduction to ontologies



Melissa Haendel

Contributors: Melissa Haendel, Chris Mungall, David Osumi-Sutherland

Common controlled vocabularies indicate the same meaning under different annotation circumstances

MouseEcotope

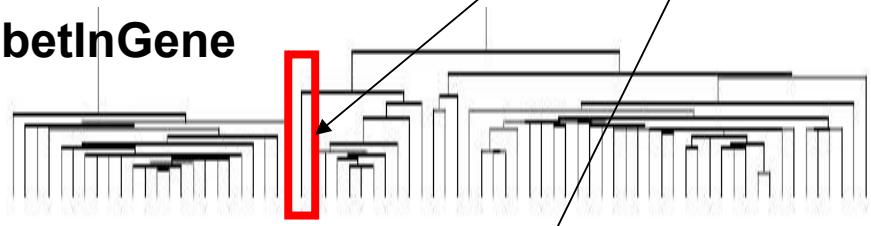
Tool	Statistical model	Correction for multiple experiments	GO Visualization	Microarrays supported	Time to process 200 genes (s)
Onto-Express	χ^2, binomial, hypergeometric, Fisher's exact test	Šidák, Holm, Bonferroni, FDR	Flat, Tree	172 commercial arrays (Affymetrix, SuperArray, Sigma-Genonsys, ClonTech, PerkinElmer, Operon, Takara, NIA); can also upload a user-defined list	7, 8, 16, 28
GoMiner	Fisher's exact test	Relative enrichment	Tree, DAG	uploads from user	77, 123, 223, 340
DAVID	None	None	Not available	Not applicable	15, 17, 27, 54
EASEonline	Fisher's exact test	Bonferroni	Not available	27 arrays (Affymetrix only); can also upload a user-defined list	15, 19, 34, 74
GeneMerge	Hypergeometric	Bonferroni	Flat, no hierarchical structure	Uploads from user	6, 6, 6, 8
FuncAssociate	Fisher's exact test		Not available	Uploads from user	22, 27, 29, 50
GOTM	Hypergeometric	None	Tree	37 arrays (Affymetrix only); uploads from user	59, 60, 157,
FatiGO	Percentage	Step-down minP, FDR (Benjamini and Hochberg, 1995), FDR (Benjamini and Yekutieli, 2001)	Flat, Tree	Uploads from user	15, 49, 69, 105
CLENCH	Hypergeometric, χ^2, binomial	None	DAG	Uploads from user	NA
GOstat	χ^2, Fisher's exact test	FDR, Holm	Not available	Uploads from user	12, 20, 46, 80
GOToolBox	Hypergeometric, binomial, Fisher's exact test	Bonferroni, Holm, Hochberg, Hommel, FDR	Not available	Uploads from user	22, 81, 145, 2[illegible]
GoSurfer	χ^2	q-value	DAG	22 arrays (Affymetrix only); uploads from user	2, 2, 2, 3
Ontology Traverser	Hypergeometric	FDR	Not available	5 arrays (Affymetrix); uploads from user	NA
eGOn	Binomial	None	Tree	Uploads from user	20, 45, 80, 95

sphingolipid transporter activity

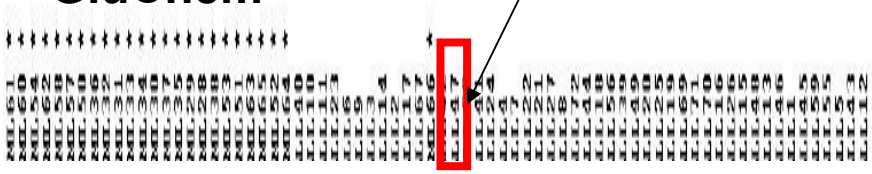
GlyProt

GLUL Glutamate-ammonia ligase
HMOX2 Heme oxygenase (decycling) 2
Ahnak-A Nucleoprotein Ahnak-A
Erythrocyte membrane 50kd glycoprotein
CA2 Carbonic anhydrase II
NFIX Nuclear factor I/X
HOMEOBOX PROTEIN HOX-A5
Homeotic Protein C5, Class I
KIAA0246 gene, partial cds
ELA2 Elastatse 2, neutrophil
MPO Myeloperoxidase
CST3 Cystatin C
ARH9 Ras homolog, member G (rho G)
ANX1 Annexin I (lipocortin I)
CSF3R
SPI1 integration oncogene
D component of complement
NF-IL6-beta
MAN2A2 Alpha mannosidase II isozyme
Homolog suppressor-of-white-apricot
CD2 antigen (p50)
Myosin VIIA (USH1B)
SPRR1B Small proline-rich protein 1B
ACTN2 Actinin alpha 2
KIAA0080 gene, partial cds
DTYMK Deoxythymidylate kinase
Carcinoembryonic antigen precursor
LTB Lymphotoxin-beta
GATA3 GATA-binding protein 3
GB DEF = Escherichia coli unknown
GUANYLATE CYCLASE, BETA-1
Spinal Muscular Atrophy 4
KIAA0239 gene, partial cds
Butyrophilin (BTF4)
MHC-encoded proteasome LAMP7-E1
RPA1 Replication protein A1
Clone 22 mRNA
CTPS CTP synthetase
UBIQUITIN-LIKE PROTEIN GDX
(AF1q) mRNA
S100 calcium-binding protein A13
BLK Protein-tyrosine kinase blk
IGB (B29)
OBF-1
TCL1
CD19 antigen
Skeletal muscle abundant protein

DiabetInGene



GluChem



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:

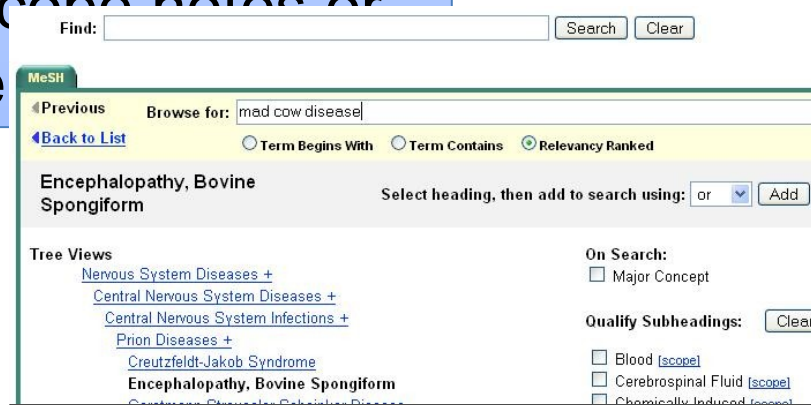
Wine
- **Red**
  - **merlot**
  - **zinfandel**
  - **cabernet**
  - **pinot noir**
- **White**
  - **chardonnay**
  - **pinot gris**
  - **Riesling**

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



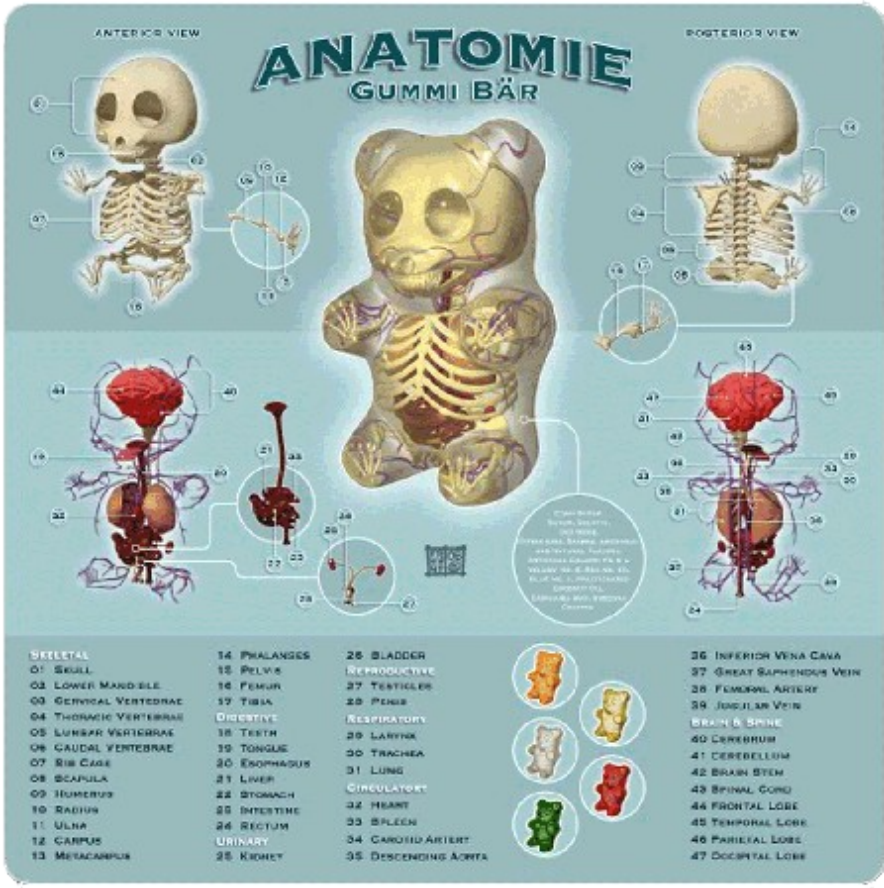
Some well known thesauri are:
WordNet, NCI cancer thesaurus, MeSH

What is an ontology?

A formal conceptualization of a specified domain of interest

Key Features:

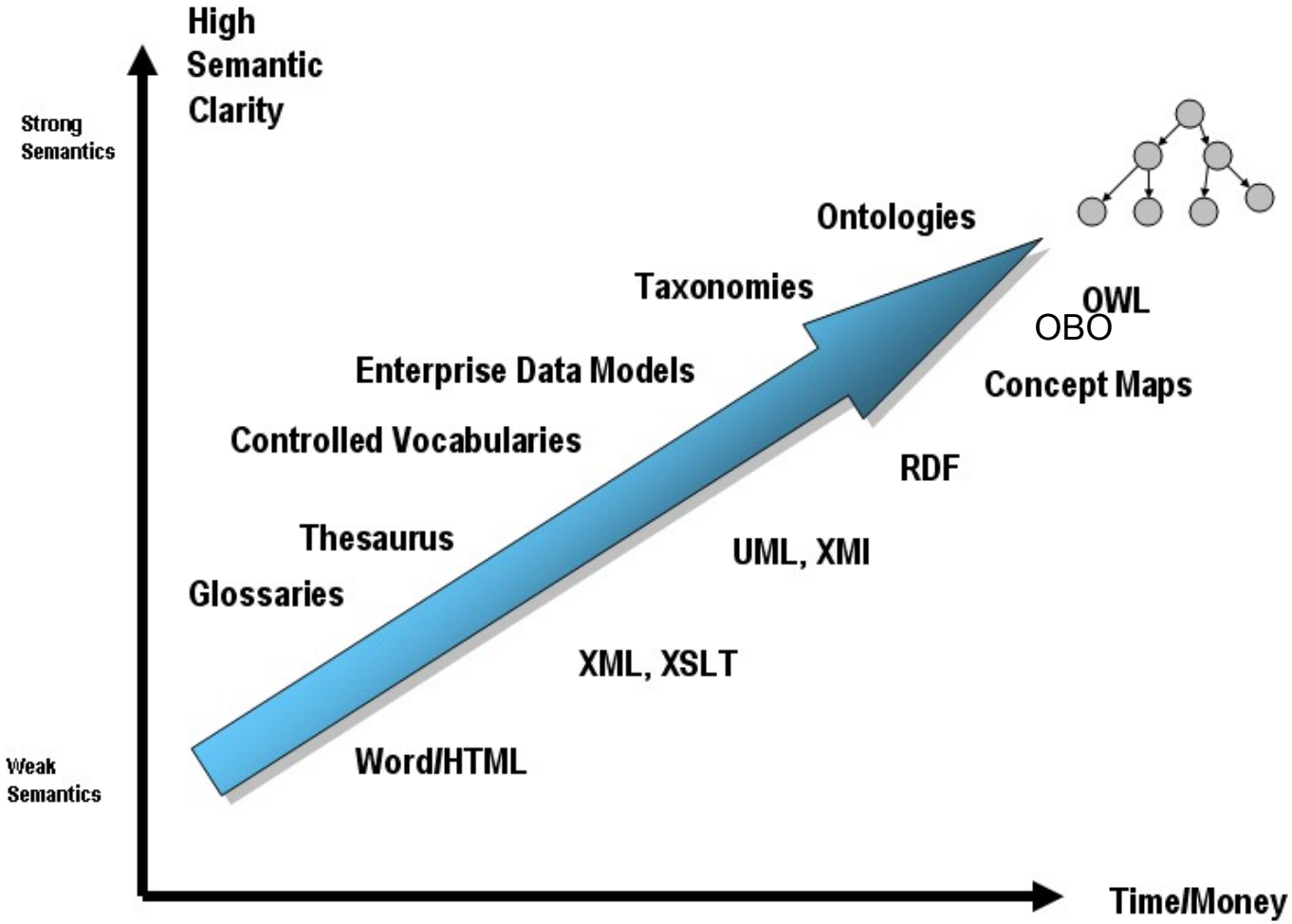
- Terms are defined.
- Relationships between the terms are defined, allowing logical inference.
- Terms are arranged in a hierarchy.
- Expressed in a knowledge representation language such as RDFS, OBO or OWL.



Reproduced with permission, Jason Freeny
http://web.mac.com/moistproduction/flash/index.htm

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

The Ontology spectrum:



http://www.mkbergman.com/?m=20070516

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* <u>*means*</u>?**
- **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 <u>not</u> *ontological questions*. 1 is a purely *linguistic dispute*; 2 is primarily a *practical question*.

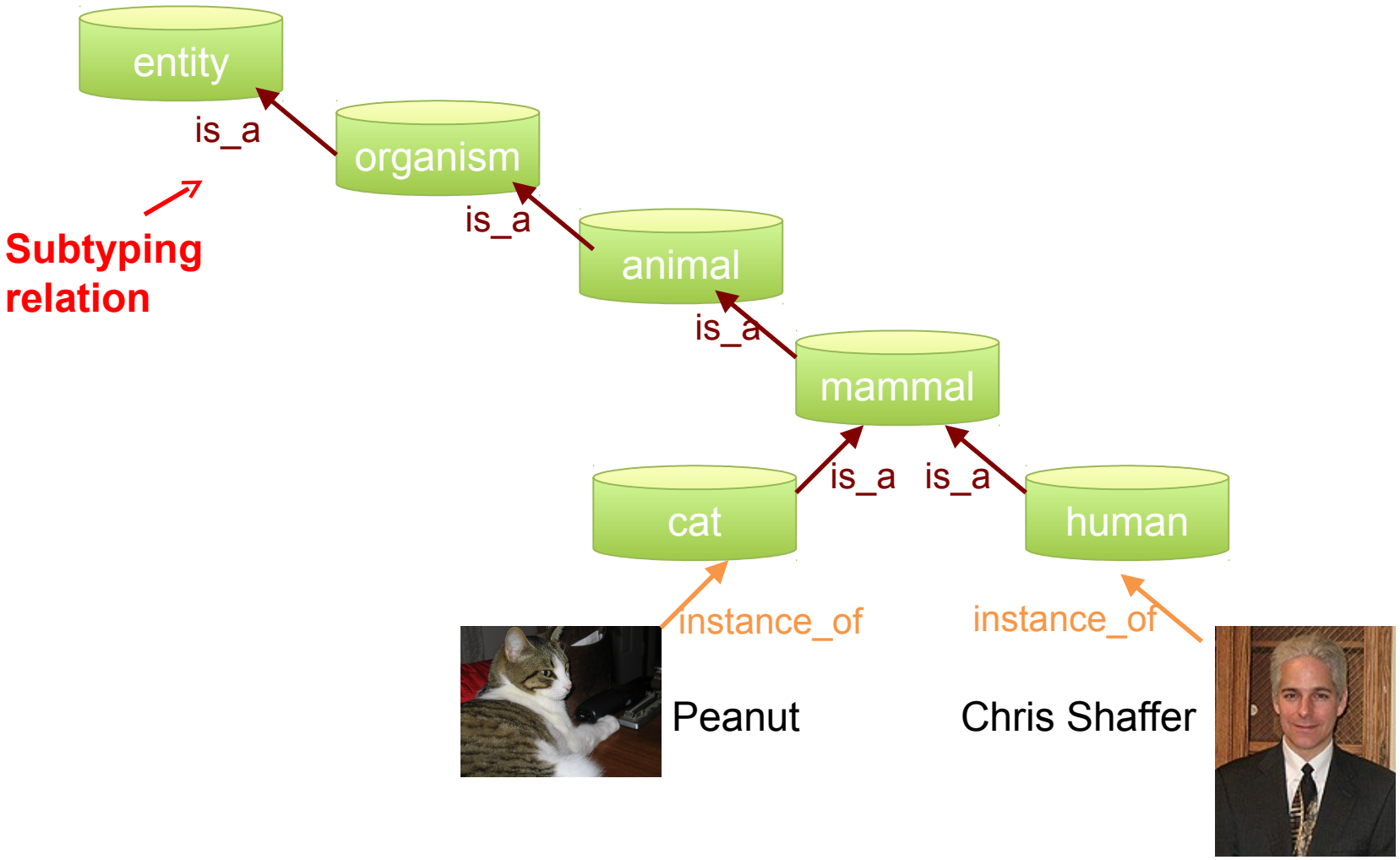
The ontological questions are:

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

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

Adapted from Gary Merrill

Types, subtypes, and instances



is_a = SubClassOf

How do you tell if it is an instance or a class?

there more than one in existence?

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

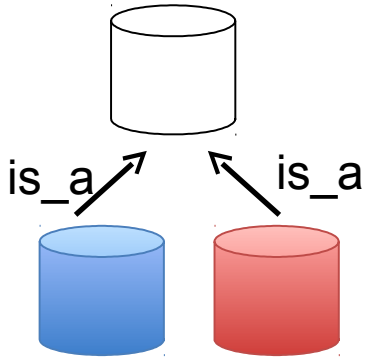
There is a class of things labeled “Jeffrey Pine Cones”

General Principle for Logical Definitions

Definitions are of the following Genus-Differentia form:

X = a Y which has one or more differentiating characteristics.

where X is the is_a parent of Y.



Definition: cylinder = Surface formed by the set of lines perpendicular to a plane, which pass through a given circle in that plane.

Definition: Blue cylinder = Cylinder that has color blue.

Definition: Red cylinder = Cylinder that has color red.

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?

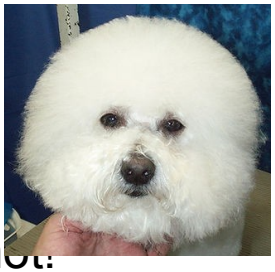
Transitivity.

<u>Some relations are transitive, and apply across all levels of the hierarchy.</u>

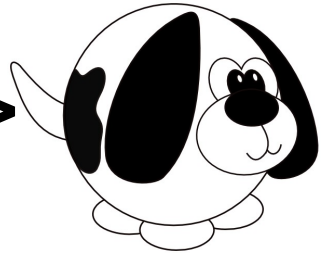
For example, a cat *is_a* mammal, and a mammal *is_a* vertebrate
SO
a cat *is_a* vertebrate
=> This is the true path rule and is because the *is_a* relation is transitive.

<u>Some properties are not transitive.</u>

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



$\not\Rightarrow$



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

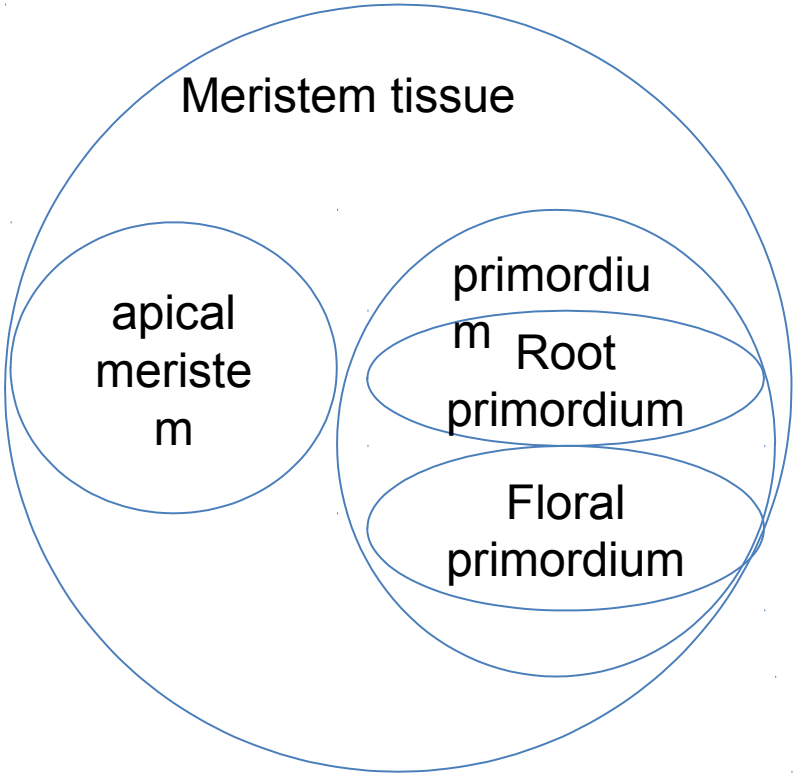
A piece of software able to infer logical consequences from a set of asserted facts or axioms.

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.

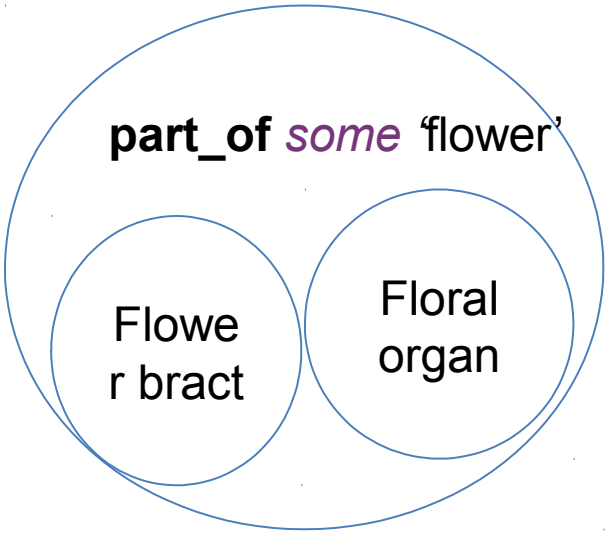
Classifying anatomy

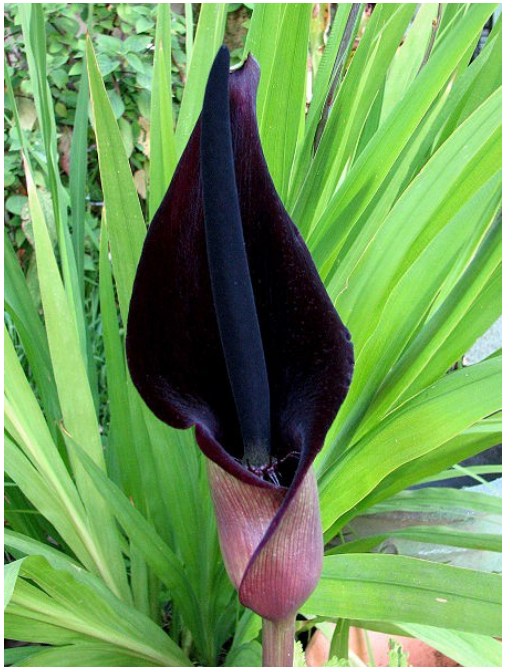


Subclasses are like subsets

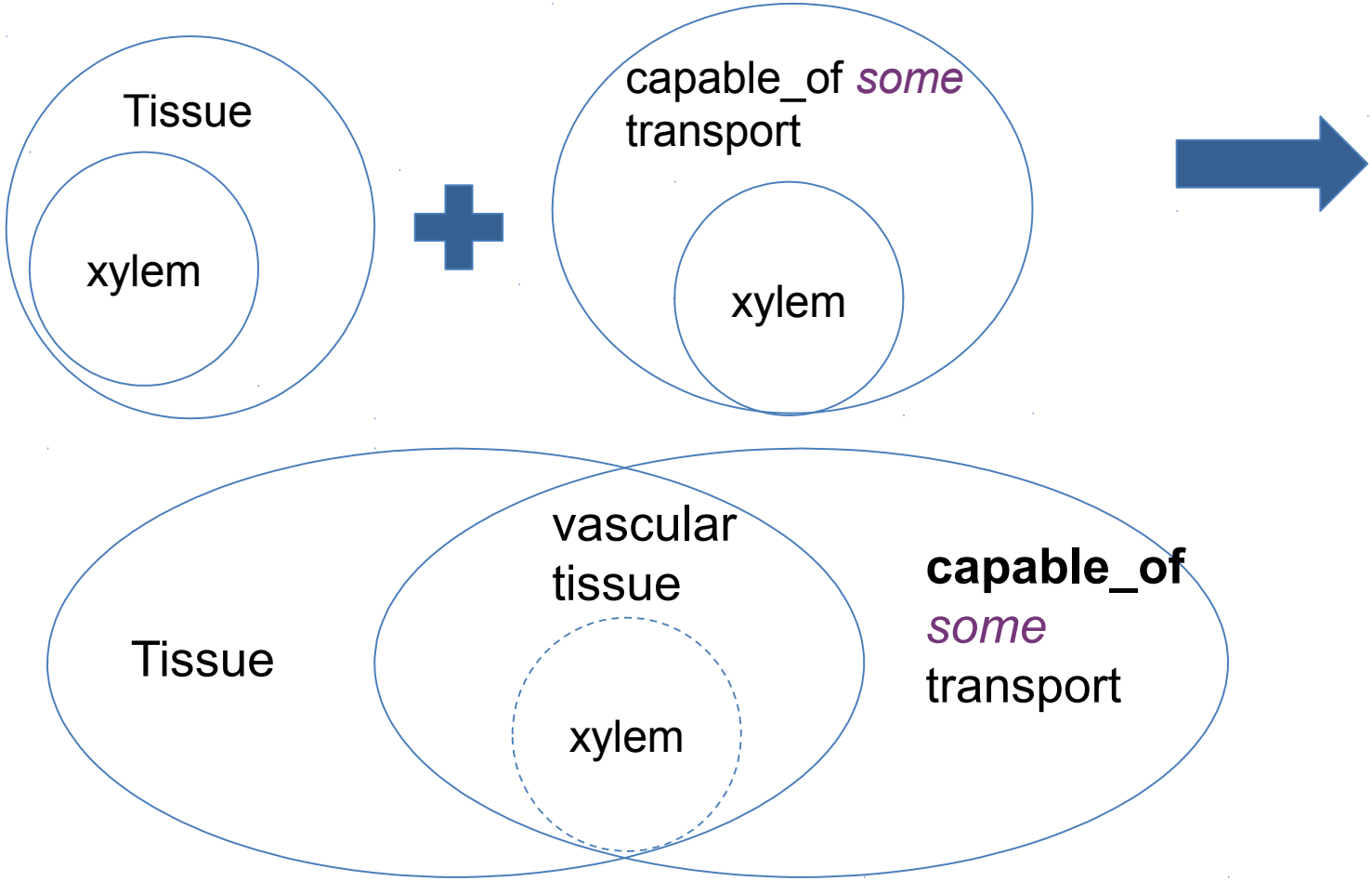
Relationships record classifications too

‘floral organ’ *SubClassOf* **part_of** *some* ‘flower



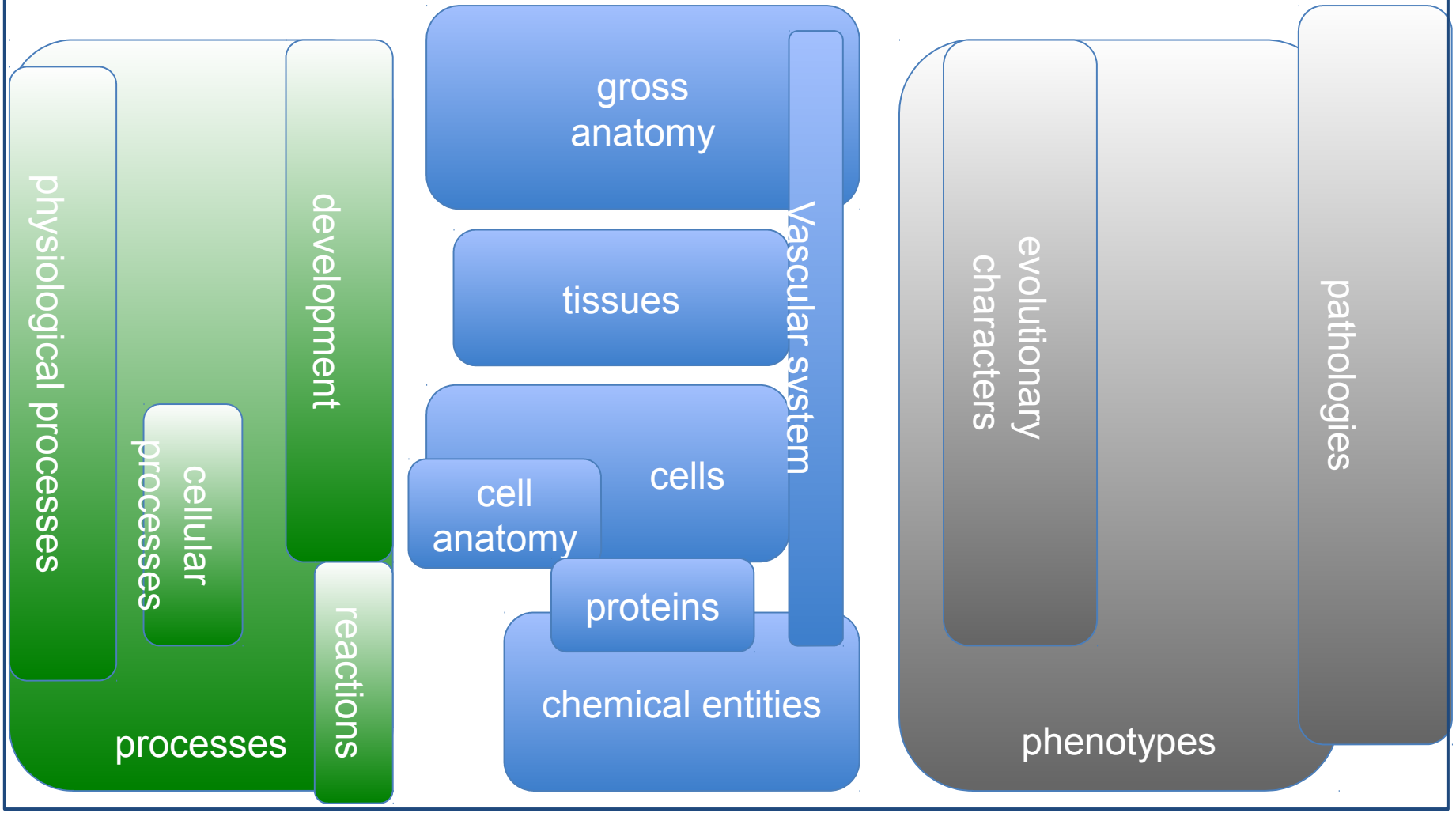


The knowledge in an ontology can make the reasons for classification explicit



- **=> These are necessary and sufficient conditions, also called an equivalent class axiom**

Many perspectives, many ontologies – that overlap in content



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

Differentia

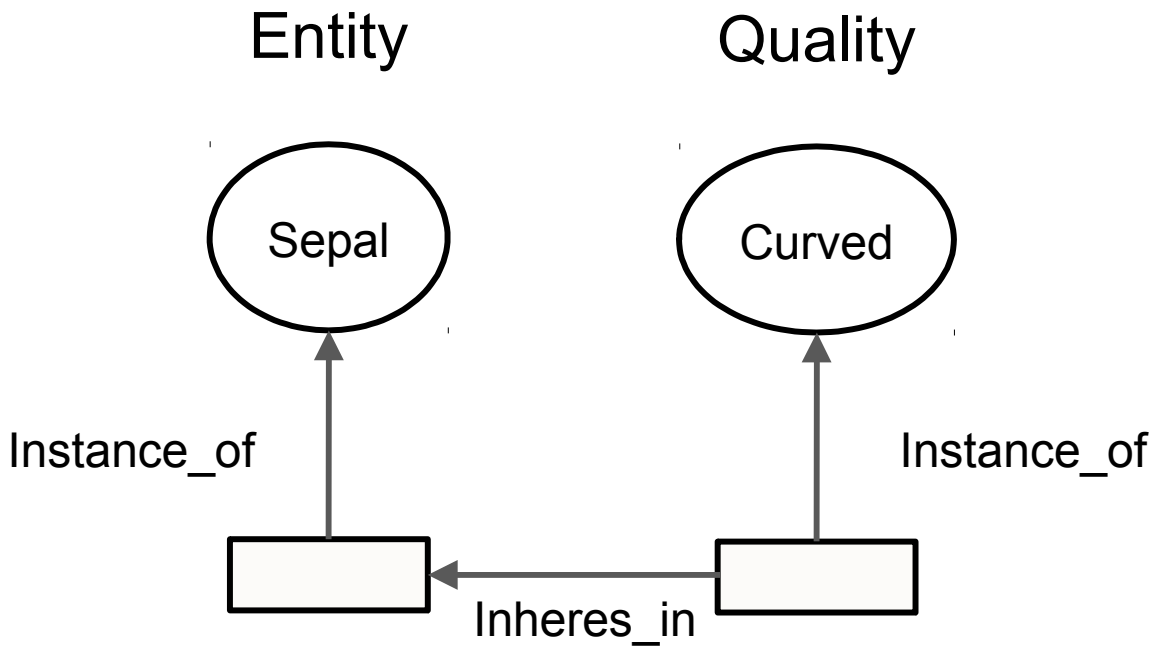
a plasma membrane **which** is *part_of* a spermatocyte

Gene Ontology

Basic Formal Ontology

Cell Ontology

Entity-quality model of phenotype representation



Pre vs. post-composition of phenotypes

RO/BFO

(awn and *'is bearer of'* some lemma)

PO

RO/BFO

) and (*part_of* some

PO

OR

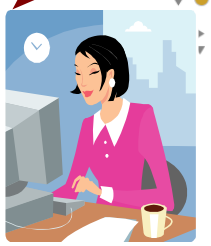
(awn color) and (*part_of* some lemma)

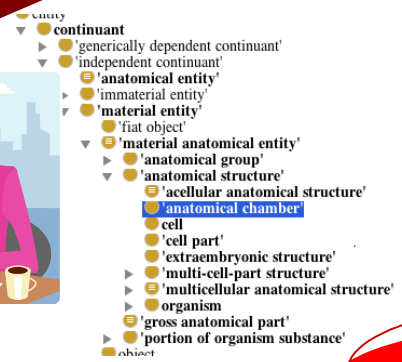
TO

PO

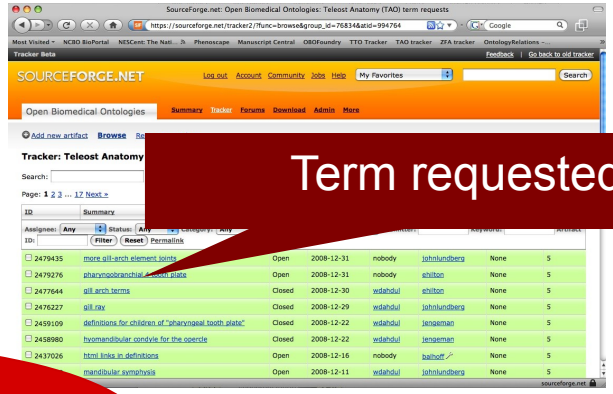
Lists, trackers, ontologies, annotation, oh my!

Ontology Edited





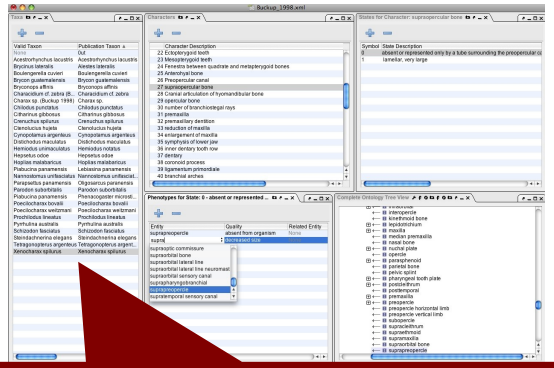
Tracker IDs can be in ontology metadata



Term requested

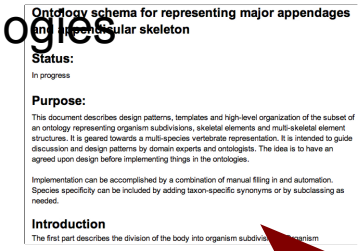
Trackers are often autoemailed to integrated listserves

Term brokers are being developed to create temp classes during ontology editing or annotation

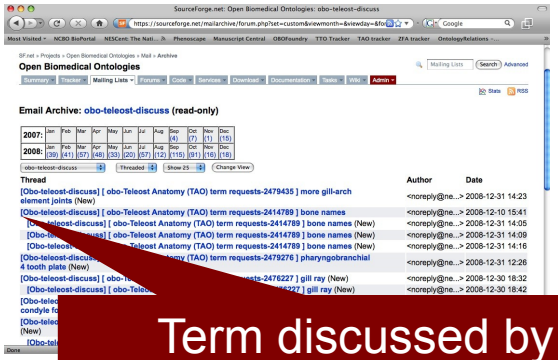


Term needed for annotation

Design documents comment on existing ontologies



Design and requirements analysis



Term discussed by community

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

Example tracker request

https://sourceforge.net/tracker/index.php?func=detail&aid=3456359&group_id=36855&atid=440764

5 ossification expansion and restructuring - ID: 3456359 Last Update: Comment added (rfoulger)

Details: GO_REF:0000034 – Phenoscape Skeletal Anatomy Jamboree

Brian K. Hall (Dalhousie University), Matthew Vickaryous (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; Ceri 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!

- Wikis: a nice public way to describe the overall content. Examples: uberon.org, http://code.google.com/p/eagle-i/wiki/Documentation
- 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**

The OBO Foundry

http://www.obofoundry.org/

www.obofoundry.org

The Open Biological and Biomedical Ontologies

Home | Contact

Ontologies | Resources | Participate | About

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 🐾s link to the term request trackers for the listed ontologies.

OBO Foundry ontologies

Title	Domain	Prefix	File	Last changed
Biological process	biological process	GO	gene_ontology_edit.obo	2012/07/25
Cellular component	anatomy	GO	gene_ontology_edit.obo	2012/07/25
Chemical entities of biological interest	biochemistry	CHEBI	chebi.obo	2012/07/03
Molecular function	biological function	GO	gene_ontology_edit.obo	2012/07/25
Phenotypic quality	phenotype	PATO	quality.obo	
PRotein Ontology (PRO)	proteins	PR	pro.obo	
Xenopus anatomy and development	anatomy	XAO	xenopus_anatomy.obo	2012/02/17
Zebrafish anatomy and development	anatomy	ZFA	zebrafish_anatomy.obo	2012/06/21

OBO Foundry candidate ontologies and other ontologies of interest

Title	Domain	Prefix	File	Last changed
Adverse Event Reporting Ontology	health	AERO	aero.owl	
Amphibian gross anatomy	anatomy	AAO	AAO_v2_edit.obo	
Amphibian taxonomy	anatomy	ATO	amphibian_taxonomy.obo	
Anatomical Entity Ontology	anatomy	AEO	aeo.obo	2012/06/01
Ascomycete phenotype ontology	phenotype	APO	ascomycete_phenotype.obo	2012/03/01
Basic Formal Ontology	upper	BFO	1.1	
Bilateria anatomy	anatomy	BILA	bilateria_mrca.obo	
Biological imaging methods	experiments	FBbi	image.obo	2011/05/24
BRENDA tissue / enzyme source	anatomy	BTO	BrendaTissueOBO	
C. elegans development	anatomy	WBls	worm_development.obo	
C. elegans gross anatomy	anatomy	WBbt	WBbt.obo	
C. elegans phenotype	phenotype	WBPhenotype	worm_phenotype.obo	2012/07/24
Cell type	anatomy	CL	cl.obo	
Chemical Information Ontology	biochemistry	CHEMINF	cheminf.owl	
Common Anatomy Reference Ontology	anatomy	CARO	caro.obo	2011/12/14
Comparative Data Analysis Ontology		CDAO	cdao.owl	
Dendritic cell	anatomy,immunology	DC_CL	DC-CL_deployed.obo	2009/06/30
Dictyostelium discoideum anatomy	anatomy	DDANAT	dictyostelium_anatomy.obo	2010/04/06
Drosophila development	anatomy	FBdv	fly_development.obo	2010/07/01

Quick Links

- Mappings between ontologies
- Download alternate formats
- About the OBO Foundry
- Current events
- How to join
- OBO Foundry paper in *Nature Biotechnology*, November 2007

Other Ontology Lists

- BioPortal (NCBO's ontology repository)
- Ontology Lookup Service (OLS) (OBO Foundry term lookup)

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