

# Ontologies for Evolutionary Biology



July 29 to August 2, 2013

National Evolutionary Synthesis Center, Durham, NC

## Contents

[Class materials](#)

[Monday, July 29](#)

[Introductions and biographies](#)

[Matt Y - The Lifespan of an Ontology](#)

[Students - Breakout exercise](#)

[Melissa H - Introduction to Ontologies](#)

[Students - SVN story exercise](#)

[Melissa H - Ontology Communities](#)

[Students - Mapping exercise](#)

[Jim B - VUE functionality](#)

[Melissa H - Best Practices](#)

[Monday summary](#)

[Tuesday, July 30](#)

[Students - Finishing VUE exercise](#)

[Jim B - Evolutionary Phenotype Data Annotation](#)

[Karen Cranston \(guest talk\) - Open Tree of Life](#)

[Students - "What's out there" exercise](#)

[Jim B - Introduction to OWL](#)

[Tuesday summary](#)

[Wednesday, July 31](#)

[Students - The big Protégé tutorial](#)

[Jim B - Limits to Ontologies](#)

[Thursday, August 1](#)

[Melissa H - Ontology Interoperability](#)

[Class discussion - Exploring ontologies](#)

[Students - Plugins exercise](#)

[Melissa H - Imports and MIREOT](#)

[Students - Import and Ontofox tutorials](#)

[Class discussion - Ontology evaluation](#)

[Wednesday & Thursday summary](#)

[Friday, August 2](#)

[Student presentations](#)

[Friday summary](#)

[Class photo](#)

Course syllabus	<a href="http://bit.ly/14o9itW">http://bit.ly/14o9itW</a>
Google Project	<a href="http://nescent-anatomy-course.googlecode.com/">http://nescent-anatomy-course.googlecode.com/</a>
Mendeley Group	<a href="http://www.mendeley.com/groups/3545281/nescent-ontology-course-2013/">http://www.mendeley.com/groups/3545281/nescent-ontology-course-2013/</a>

## Class materials

Available on the [Google Project](#) site, either check out from Subversion or download as zip.

Throughout these notes, relevant documents are highlighted and can be found in:

2013\_course/material\_for\_course/

## Monday, July 29

### Introductions and biographies

**Melissa Haendel (instructor)** - Currently lead ontologist for the eagle-i project, Oregon Health & Science University, designed to collect and disseminate information about scientific resources. Dr. Haendel has a Ph.D. in neuroscience, where she studied early development in mouse, chick and zebrafish. Dr. Haendel has co-lead several anatomy ontology workshops and has participated in development of development of standards for anatomy ontologies. She has participated in development of the Common Anatomy Reference Ontology (CARO) and taxonspecific anatomy ontologies such as the zebrafish anatomy ontology (ZFA), the the teleost ontology (TAO), and the vertebrate skeletal ontology (VAO). Most recently, she has been working to facilitate cross-species anatomy interoperability (Uberon). Her interest is in using anatomy ontologies as a nexus for linking many types of data - genomic, biomedical, evolutionary - using Semantic Web standards.

**Matt Yoder (instructor)** - A lead researcher on the Hymenoptera Anatomy Ontology and a biodiversity informatician at the Illinois Natural History Survey. Matt contributes to the curation of the HAO and seeks to develop user-friendly tools that enable domain experts, who are not usually ontology experts as well, to employ phenotype ontologies as part of their descriptive work. Trained in Hymenoptera systematics and biodiversity informatics Matt has broad experience in phylogenetic reconstruction, with specific background in rRNA multiple sequence alignment analysis. He has worked on implementing informatics solutions across a wide range of taxa.

**Jim Balhoff (guest instructor)** - Working on the Phenoscope and Hymenoptera Anatomy ontologies, Jim is an expert in semantic phenotype and ontology meta-data modelling, inference, and ontology-based application development.

**Erik Segerdell (course assistant)** - Graduate student in bioinformatics at Oregon Health & Science Univeristy. Ontology Coordinator for Phenotype RCN. Experience in model organism database curation and developing anatomy (zebrafish and Xenopus) and phenotype ontologieis.

**Mariangeles Arce** - Academy of Natural Sciences of Drexel University. Georeferencing technician for the FishNet 2 Collaborative Georeferencing Project. Phylogenetics and taxonomy, fish in South America, anatomy (muscle), interested in Teleost Anatomy Ontology and evo/devo.

**Erik Chen** - Graduate Student at University of Ottawa. Gene retentions, interest in Gene Ontology.

**Peter Uetz** - Center for the Study of Biological Complexity. Interest in vertebrates/reptiles, limb development, microbiology, bacterial phenotype ontology. An Ontology for Bacterial Phenotypes (joint project with Jim Hu, TAMU, and Michelle Giglio, UMD), the Reptile Database.

**Ashleigh Smythe** - Phylo/systematics of nematodes, extend ontologies from *C. elegans* to other nematodes. I'm a nematologist/parasitologist, mostly working on free-living marine nematode systematics these days. I did my Ph.D. at UC Davis and then a postdoc at the Smithsonian NMNH. I've been a visiting assistant prof. at Hamilton College for six years but am currently a Scholar-in-Residence and looking for a tenure track position. I have no prior experience with ontologies but I'd like to extend what's been done for *C. elegans* to include other nematode taxa and eventually connect the morphology of marine and parasitic nematodes to the wealth of genetic data available for *C. elegans* and other model nematodes.

**Claus Weiland** - Scientific Programmer; LOEWE Biodiversity and Climate Research Centre (BIK-F; part of Senckenberg Society). Natural history collections, hardware infratructure, scientific repos, environmental ontologies.

**Maria Christina Diaz** - Nova South Eastern University. Sponge taxonomy and evolution, Porifera ToL, collaborator on PORO, phylogenetics and morphology, mining of functions. I am a naturalist, specialized Poriferologist, with 20-something years as a taxonomist, and always researching on life histories, microbial symbioses, functional ecology, and evolution, and now plunging with genetic googles into phyllogentic histories, with the Porifera tree of Life Project.

**Chris Sheil** - John Carroll University. Associate Professor of Biology. Herpetology, developmental morphology, interest in homology, turtle anatomy, chondrocranium.

**John Wieczorek** - Information Architect, Museum of Vertebrate Zoology, UC, Berkeley. Lead Programmer, VertNet. Convenor, Darwin Core Task Group of Biodiversity Information Standards. Ontology interests: Biological Collections Ontology (BCO).

**Elise Larsen** - Ecologist, interest in Lepidoptera trait database, applying ontology to it. Currently I'm a PhD Candidate at University of Maryland in Biology (focus in community ecology). Previous MSc in Biology (2007). This fall I will be joining a research project which includes the development of a trait database for lepidopterans. I have done some database development in the past for ecological and toxicological research, but have not worked with ontologies in the past.

**Jing Liu** - Post doc at University of Florida. Biological literature describing species, extracting character matrix.

**Steve Elliott** - Mathematical models in evolutionary biology, interest in semantic extraction from encyclopedia, how people are using ontologies. I'm a PhD candidate at Arizona State. My dissertation research is about how to use information about gene regulation with models from population genetics/quantitative genetics. But my ontology-related work is for the Embryo Project, which has long term goals in post-secondary education, science outreach, and research into how science changes over time.

**Laura Jackson** - Graduate Student at Southeastern Louisiana University. Interest in fish, evolutionary development/ontology using Phenoscope.

**Corrine Blank** - University of Montana. Archaean domain of life, trait and habitat evolution, matching with transitions in evolutionary record.

**Anne Thessen** - My background is in biological oceanography, plankton ecology and protist taxonomy. I'm interested in applying ontologies to the federation of data, particularly for microscopic species. I'm working as a "free lance" biologist right now. I have appointments at Arizona State University, University of Maryland and Marine Biological Laboratory. I am working on several different projects. My PhD is from University of Maryland Horn Point Laboratory and I did my Post-doc at MBL with the Encyclopedia of Life and the International Census of Marine Microbes.

## **Matt Y - The Lifespan of an Ontology**

mon/ontologyLifespan.pptx

Glossary assignment

Our expectations

MH: you are an investment in our future!

Evolution of ontology development

OBO → OWL/Protege

OBO now considered a subset of OWL

Workshop flow

Mental landscape

Anatomy concepts

Wasp paramere example - it has been used to mean many different things

Goal of ontology - clarifying concepts

Large number of descriptions for what we are talking about

Example of concepts and terms from 1916

We are not doing anything new, just with more formalization

What is an ontology?

Concepts, and relationships among concepts

Are ontologies the answer?

MH: There is a role for Wikipedia (will talk later about this)

Ontology is a tool, not necessarily aiming to represent all of biology

What do I actually need my ontology to do for me?

Roadblocks

Example of representation of antenna in Manchester syntax

Examples of structured datastores

Biological ontologies

Driving a user interface

Landmark paper for identifying human disease

Eye phenotypes annotated with ontologies

3-D reconstructions, annotations

Linkages using ontologies

Breakout exercise - Will ontologies work for you?

Ontology management

Lifespan

- Inception
- Planning
- Data gathering
- Understanding
- Formalization
- Integration
- Distribution
- Refinement
- Reasoning
- Applications
- Research
- Obsolescence

**Students - Breakout exercise**

## Melissa H - Introduction to Ontologies

mon/IntroToOntologies.pptx

Overview of controlled vocabularies, taxonomies, and thesauri

What is an ontology?

Computers can use the taxonomies in a way humans can't do

The ontology spectrum

Ontology languages

A common misconception

Are ontologies about *terms* or *things*? (Very important slide)

What matters are things

Numeric identifiers for concepts, so we can focus concepts, not labels

Why build an ontology? A simple example

Types, subtypes, and instances

Instances vs. classes

Chris: It is often context-specific

General principle for logical definitions

Genus-differentia

The True Path Rule

Transitivity

Relationships and definitions

Universality and directionality

Symmetric relations

Disjoint classes

About reasoners

Different kinds of definitions

Anatomy ontologies: exemplar use case

Ask these questions for your own domain

Many useful ways to classify parts of organisms

Not all classification is useful

Don't try to represent everything! Be practical

Relationships record classifications too

Equivalent classes

Why create equivalent classes and class restrictions?

Compositionality and avoiding asserted multiple inheritance

Let the reasoner do the work!

Post vs. pre-composition

Landscape of anatomy ontologies

Many ontologies overlap in content

Organization according to upper ontologies

Continuants and occurrents

Basic Formal Ontology high level classes

Common Anatomy Reference Ontology

Using CARO as a template  
Example of complexity arising from multiple species contexts  
Using reasoners to detect errors  
Representing different levels of granularity  
Ontology considerations  
Chris: can't you overuse disjoint classes?  
MH: Yes. It's better practice to have fewer high level ones.  
MY: Use your use cases as a guide

## Students - SVN story exercise

### Melissa H - Ontology Communities

mon/ontologyCommunities.pptx

OBO Foundry  
    Concept of orthogonality  
    Listserve  
BioPortal  
Ontology Lookup Service  
Ontobee  
QuickGO at EBI  
Ontology editor developer and user communities  
    OBO-Edit, Protege  
Extracting knowledge from domain experts  
Phenote - <http://www.phenote.org/>  
Trackers

## Students - Mapping exercise

mon/domainModeling.docx

mon/domainModeling.vue

What were some of the issues/problems encountered?  
Chris: Trying to do it all at once, instead of choosing one attribute to start with  
Claus: Difficult to determine the line between is\_a and part\_of  
MH: CARO helps provide some strategy for dealing with these things  
    Use of colors, shapes useful for separating things conceptually  
Lack of definitions make it difficult to know where to place some things (only labels were given)  
Mariangeles: Identifying which relationship to use  
Claus: Does it make sense to give everything an is\_a relationship?

MY: In Protege world, it is required. Necessary for reasoning.  
MH: Design patterns exist, but there are different approaches to modeling the same concepts

## **Jim B - VUE functionality**

[mon/vue2owl.pdf](#)

Overview of rules for drawing VUE diagrams that can be translated into OWL files

## **Melissa H - Best Practices**

[mon/BestPractices.pptx](#)

OBO Foundry and OBO Library

Overview of principles

Foundry take-home messages

When to obsolete

- deprecate = obsolete

- destroy  $\neq$  delete

Intrinsic ontology evaluation

- Design documents

Metadata standards

Ontology documentation

Internal documentation

Quality textual documentation and definitions (concept-first approach)

- Example: molecular labels

Evaluating text definitions

Testing definitions against instances

Example of a difficult-to-define entity: "reagent"

Is there complete metadata?

Quality logical expressions

Ontology reuse

Iterative data-ontology evaluation

Ongoing intrinsic issues for development of biological ontologies

## **Monday summary**

The first day of the NESCent and Phenotype RCN sponsored Ontologies Course started with introductions. What a diverse group we have this year, with expertise in: phylogenetics and muscle anatomy of South American fishes, gene retention in plants, reptile limb development, microbiology, systematics of nematodes, natural history collections, sponge taxonomy and



evolution, herpetology and turtle anatomy, biodiversity standards and Darwin Core, ecology and Lepidoptera traits, text mining species and character matrices, mathematical models in evolutionary biology, developmental biology of fishes and evolution of habitat and physiological traits in Cyanobacteria and Archaea. WOW!

The day was packed with a lot of lecturing about logic and how it can be your friend, true path violations, ontology best practices, and the community that we are convincing our students that they are now part of :). However, we promised that on Tuesday we'd get them using what they had been taught and it would all make more sense once they got their hands dirty. The students had homework last night – they started working on modeling their own ontology project for the course in VUE, and we plan to convert these files to OWL with Jim's new script so that they can continue their work in OWL following Wednesday's Protege tutorial. We also had a very interesting discussion about the differences between specimens and samples, intent to collect, and whether or not populations or tissues can be target populations for sample collection.

**Tuesday, July 30**

### **Students - Finishing VUE exercise**

tues/modelling.txt

### **Jim B - Evolutionary Phenotype Data Annotation**

tues/phenotype\_annotation.ppt

- Phenotypic data
- Phenotype annotation
  - Ontology of phenotypes
  - Compositional approach
- Entity-Quality model
- EQ model for evolutionary phenotypes
- Creating annotations
- Using annotations
- ZFIN annotation example ("head")
- Phenoscape KB
- Semantic similarity
- Phenex exercise

### **Karen Cranston (guest talk) - Open Tree of Life**

Using phylogenetic metadata in large-scale phylogenetic synthesis  
Minimal Information About a Phylogenetic Analysis (MIBBI)  
NeXML - Rich phyloinformatic data  
CDAO - Comparative Data Analysis Ontology  
PhyLoTA Browser - <http://phylota.net/>  
Tree of Life web project - <http://tolweb.org/>  
Taxonomic Name Resolution Service - <http://tnrs.iplantcollaborative.org>

#### **Data standards sites:**

BioSharing - Developing standards for describing and sharing biosciences experiments - <http://www.biosharing.org/>  
BioCoreDB - Information specification for biological databases

MH: You need an ontology for data inferencing, but for data linking maybe you do not

### **Students - “What’s out there” exercise**

tues/out\_there\_exercise.txt

### **Jim B - Introduction to OWL**

tues/OWLintro.pptx

RDF - the data standard for the semantic web  
    Triples  
RDF data integration - example of combined datasets  
OWL - added semantics  
    OWL 2 Full  
    OWL 2 DL  
OWL primer - <http://www.w3.org/TR/owl2-primer/>  
    Excellent resource!

MY: An ontology only knows what you tell it  
Everything is possible unless you state otherwise (Open World Assumption)  
Satisfiability - <http://ontogenesis.knowledgeblog.org/1329>

### **Tuesday summary**

Day two started with a “speed-dating” approach with instructors pairing off for short periods with participants to strategize and work on individual projects. VUE files representing participants’ projects continued to be formalized, some now contain many nodes and some are even very pretty. These visual representations will be translated into OWL files shortly, and further refined

in Protege. The morning progressed into a presentation on annotations, where tools like [Phenex](#) and [Phenote](#) were outlined.

In the afternoon we had a great overview from Karen Kranston, PI on the [OpenTreeOfLife](#) project, and we discussed how ontologies may or may not be useful for projects like OToL. We continued with a survey of web-based resources related to evolutionary biology, with participants auditing well known websites for their use, or lack thereof of ontologies. The day concluded with the last bit of preparation prior to our big practical exercise on Protege on Wednesday, a nice overview of OWL, with specific reference to the (very nice) [primer](#). We're looking forward to the first real taste of formalization with the Protege tutorial, and the creation of individuals' own ontologies.

## Wednesday, July 31

### Students - The big Protégé tutorial

[wed/protege\\_tutorial.doc](#)

How to download and run the script for converting VUE files to OWL:

[wed/running\\_vue2owl.txt](#)

### Jim B - Limits to Ontologies

[wed/ontology\\_limitations.ppt](#)

Reasoning performance

OWL 2 profiles

EL: good for anatomy ontologies

OWL EL

ELK reasoner

Triplestore reasoning

Queries with Sparql

Expressivity

Examples:

- absence
- counts
- structured objects

Time

Model consistency

Thursday, August 1

## **Melissa H - Ontology Interoperability**

`thurs/OntologyInteroperability.pptx`

Goal of talk: thinking about modularization

The problem: data silos

Ontology alignment

How to synchronize anatomy ontologies

Issues with mappings

Example: In BioPortal, FMA colon mapped to GAZ Colon, Panama

Reconciliation and linking between TAO and ZFA

Xref semantics across ontologies

Using UBERON for alignment facilitates identification of missing classes

Uberon classes connect to other ontologies via a variety of relations

Synchronization by reuse from external ontologies

Modularizing ontologies - what we need

It's a lot easier to reuse than to align

Ontologies working together

Example from HAO and PATO

Ontologies can help reconcile annotations

## **Class discussion - Exploring ontologies**

`thurs/exploring_ontologies.txt`

## **Students - Plugins exercise**

`thurs/plugins.txt`

## **Melissa H - Imports and MIREOT**

`thurs/imports_mireot_ontofox.pdf`

Importing ontologies

OWL imports

Challenges for importing other ontologies

Possible solutions

Full import

Generating your own terms and using xrefs

Generating an import module

Dealing with imports in Protege

See the tutorial document

Idea: import only classes that are needed

Pro and con

MIREOT: Minimal Information to Reference External Ontology Terms

Define the minimal information we need

Additional information

Implementation strategy

OntoFox tour

Example: importing the CARO term “cell”

## Students - Import and Ontofox tutorials

[thurs/imports\\_tutorial.doc](#)

[thurs/ontofox\\_tutorial.doc](#)

## Class discussion - Ontology evaluation

John W: BCO is a representation of Darwin Core

Use case: data about species distribution

Idea behind BCO is to expand to genomic, environmental information

MH: can tap into a community of developers

A lot of class axioms coming from BFO

Two classes that are the same class but different URIs

BFO2 is an experimental file and has some issues

Usage panel in Protege is very useful

Save often when using Pretege

Commit early and often, write good commit messages

MH: Developing ontologies involves partnership between ontologists and domain experts

CB: When you import part of another ontology, won't there be a lot of broken links in your ontology?

MH: Using a tool like Ontofox will clean those up (ensure closure)

“Dangling” classes will be subclasses of OWL “Thing”

EL: Butterfly habitats, a hierarchy of more specific habitat types

References ENVO ontology “tropical”, but with her own definition

MH: Good genus-differentia text definitions are a good place to start wrt class expressions

Ontologies imported from OBO usually are not very expressive

## Wednesday & Thursday summary

We spent the entire Day 3 working on learning Protege and exploring the use of various OWL2 capabilities. The students worked at their own pace, but all made it to the light at the end of the tunnel. They are all now enlightened :-)) (bearers\_of some instance\_of PATO:0001291). We had also converted the students' VUE files into OWL, and some students were able to start work on their OWL files. Day 4 of the course was focused on developing skills around reuse of other ontologies and data interoperability. We learned techniques for performing OWL imports and the use of MIREOT (Minimum Information to Reference and External Ontology Term), which is basically a way to use a subset of another ontology in your own.

## Friday, August 2

### Student presentations

#### Chris S:

Interests: developmental morphology, ontogeny  
Need for ontologies: better understanding of chondrocranium anatomy, controlled vocabulary  
Focus on orbitotemporal region  
Challenges: open spaces that are defined by cartilages surrounding them  
Material vs. immaterial entities  
Dealing with dimension of time

#### Elise L:

Goals: knowledgebase of North American butterflys, analytical approach for monitoring data  
Map of butterfly life histories and ecological traits  
Focus on habitat, ENVO "terrestrial habitat" class  
Challenge: not well defined in ENVO  
Expanded the concept of habitat, connected it to a "biome"  
part\_of or subtype of biome?  
Brainstormed habitat types

#### Ashleigh S:

Interest: nematode anatomy ontologies  
C. elegans has been the baseline subject of study  
Focus: feeding structures in Enoplida, plant parasitic nematodes, amphid structures  
Ontology: captured shapes of amphids, parts of head and nervous system  
Simple organization: part\_of and subclass relationships  
Challenge: where to meet in the middle between C. elegans and UBERON?

**Mariangeles A:**

Interest: cranial and ventral musculature of fish

Diagrammed cranial muscle types, pectoral girdle

Modeled function and action (e.g. adduction)

Challenge: nomenclature coordination between research and ontology communities

Controlled vocabulary can help understand relationships

**Peter U:**

Interest: The Reptile Database, need to represent diversity of species

Diagrammed character ontology for snakes

Extending/connecting AOs, biodiversity ontologies, Darwin Core, ecology

Taxonomy or ontology?

Goal: character tables for reptiles, standardized nomenclature

Long-term goal: from genome to phenome, using ontology for searching database

**Corrine B:**

Interest: Next generation phenomics for Tree of Life

Needs a comprehensive microbe-specific ontology

Natural language processing tools using an ontology (CharaParser)

Mostly focused on plants, need to extend methods to microorganisms

Challenges in making microbial descriptions

Spreadsheet of terms (Archaea, Mycoplasmas, Cyanobacteria) → VUE file

Plans: core structure, merge and extend existing ontologies, populate character matrices

**Jing L:**

Interest: CharaParser and ontology

Generates marked-up output: structure - character - state → character matrix

Existing problems: unreasonable structure-character combos, sparseness of output matrices

Ontology:

Incorporated PO, TO, PATO

Added object properties is\_trait\_of, has\_state

Use cases: retrieving reasonable structure-character-state by exploring is\_a, part\_of rel's

**Maria Christina D:**

Interest: Sponges - interesting skeletons, colors, shapes

Focus on filtering apparatus

PORO - Porifera Ontology based on thesaurus

Need to expand PORO to genus of skeletaless sponge

Diagrammed body layers, cells, subcellular structures

Next steps: adding taxa, chemical compounds, genes

**Eric C:**

Interest: Gene retention after whole genome duplication, triplication

Functional basis in what is retained?  
Gene balance hypothesis  
Focus: genomes of cacao, grape, peach  
Orthologs in multiple genomes  
Ontology: characterizing fractionization data

**John W:**

Interest: Darwin Core  
Model in VUE: simple representation of what is in Darwin Core  
Addition of concepts that are missing or conflated in it:

- organism
- collection
- evidence

Imports form relevant ontologies: BFO, ENVO, etc.

**Laura J:**

Interest: understand relationships in primordial germ cell development  
Modelled fins, genes, developmental processes  
Ontology: emphasis on what processes genes are involved in  
Goal: interpretation of phylogenies via the relationships  
Links to genes in other taxa

**Steve E:**

Works for Embryo Project - online encyclopedia geared to non-specialists  
Content: articles, pictures  
Goals: education, research, outreach  
Job is to edit and clean up the encyclopedia  
RDF triple store and ontology  
Relationships between people, experiments, technology, etc. ~200 relationships  
Classes similar to domain and ranges, e.g. "Literature/People"  
Challenges: figuring out object properties to use, some weird classes like "Ethics"

**Anne T:**

Phenomics of Eukaryotic microbes

**Friday summary**

The final day of the course was presentation day, where we got to hear about everyone's research questions and how they were going to use (or not!) ontologies for their work. We group brainstormed solutions and next steps for everyone's projects. The group was very synergistic and we believe that we'll see some nice contributions and connections being made in upcoming months in the ontology community.



Class photo

