

Phenotype RCN 2012 Summit Meeting Working Group Summaries

1. Vertebrates (Melissa)

Progress on developing new vertebrate anatomy ontologies and interoperability between existing ontologies has been steady. At the 2012 Phenotype RCN summit meeting (February 2012) the vertebrate working group did several things towards this end.

1. We held a neural crest workshop to work on this vertebrate feature across all of the ontologies. The workshop was a great success. Changes to Uberon, CL, and GO will be made first, and then these will be mirrored/mirrored in existing taxon or species specific ontologies. Bringing in the expertise from domain experts and the perspectives from the different vertebrate ontologies was key.
2. We recognized that for newcomers (namely amniote and aves/chicken) that there was insufficient documentation about existing ontology resources (eg. What is CARO? How does it affect the construction of a new ontology?) and tools (SVN, googleCode, Protégé, Oort). We therefore held an ontology 101 mini session to go over how all the ontologies are meant to work together and elected to put up a page describing this. Similarly, we should have a set of links describing available tools.
3. More advanced ontology developers met to discuss the need for a meeting similar to the neural crest one on mesoderm and/or muscles, both of which are also similarly disorganized/variably modeled across existing ontologies. There are many parallel issues.
4. We discussed the uberon/GO strategy for using taxon links in multispecies ontologies. This seems to be working well but needs expert review.
5. We discussed the need to further integrate system specific ontologies better. For example, what is the best integration strategy between VAO and uberon? For the neural crest work? The new muscle work being done? This is orthogonal to the taxon specific approach we've taken in the past.

Feb 2012 attendees: Melissa Haendel, Terry Hayamizu, Chris Wall, Paula Mabee, Erik Segerdell, Monte Westerfield, Yvonne Bradford, Fiona McCarthy, Chris Mungall, Nizar Ibrahim, Radha Pulikanti, Wasila Dahdul, T. Alex Dececchi, Robert Druzinsky, Christina Zorn, Peter Vize, Mary Shimoyama, Cynthia Smith.

2. Plants (Eva)

Discussions at this meeting continued a thread from previous plant working group meetings on ways to demonstrate to the plant community how a common plant ontology for annotating plant phenotypes could be used to address interesting research questions. At this meeting we developed a concrete plan for a pilot project to gather a test set of gene-phenotype annotations for several plant species and explore the predictive power of this dataset for identifying candidate genes from phenotypes. Specific tasks we agreed to complete by the fall:

1. Assemble a set of annotations from existing datasets that are not too far off in format (i.e. annotations made to species-specific ontologies that can be mapped to PO/PATO)
2. Convert these to PO/PATO format (George Gkoutos has kindly agreed to help)
3. Set up a data store for the annotations and associated protein sequences
4. Calculate ortholog sets for the species included in the annotation set

We hope to hold a working group meeting in the fall to work with this dataset as a group and explore what kinds of analyses can be done on it.

Feb 2012 attendees: Eva Huala, David Meinke, Shannon Bell, Rex Nelson, Sue Rhee, Fabio Fiorani, Bruce Kirchoff, Ramona Walls, Sean Walsh, Greg Riccardi

3. Arthropods (Andy)

Progress on developing common arthropod and insect anatomy ontologies has been steady but relatively slow. At the 2012 Phenotype RCN summit meeting (February 2012) the arthropod working group identified several steps that should result in increased efficiency and speedier development:

1. Combine the target anatomy ontologies under one namespace (Arthropod - referred to below as CAAO).
2. Partition the arthropod body by system, and identify expertise for each (neuro, circulatory, etc.); a list was started.
3. Establish regular (monthly) conference calls through Skype. Initial invitees will be prior RCN attendees and others who have been involved in arthropod anatomy ontology development. Summaries of discussions will be posted as blog entries by Andy Deans.
4. Identify and publicly list current use cases for a CAAO (to be posted on Phenotype RCN website). Candidate uses were discussed during this working group meeting, with each participant describing how/he would like to use a CAAO. These examples, which spanned from homology discovery to associating behavior with anatomy to many other kinds of uses, will be summarized in a blog post by 8 March 2012.
5. Identify and publicly list tools to be used in CAAO development (to be posted on Phenotype RCN website). Initial ideas span from sharing flat lists of definitions to direct ontology development in Protégé. A tracker has already been established for the project: <http://code.google.com/p/arthropod-anatomy-ontology/>
6. David Osumi-Sutherland offered to explore the idea of establishing a Web application (based on Virtual Fly Brain tools) that would enable browsing and simple queries.

Feb 2012 attendees: Andy Deans (North Carolina State University; Hymenoptera Anatomy Ontology project), Cassandra Extavour (Harvard University), Peter Midford (NESCent/Phenoscape), David Osumi-Sutherland (Cambridge University; FlyBase), Stefan Richter (Rostock), Aaron Smith (American Museum of Natural History; Coleoptera Anatomy Ontology), Jo Wolfe (Yale University), Matt Yoder (Hymenoptera Anatomy Ontology project).

David Osumi-Sutherland's [notes](#) are also available (has more details).

4. Informatics (Suzi)

- Terry: organism specific needs - keeping ontologies in sync. sharing across mice and other vertebrates
- Harry: reuse across organisms - search - browsing - annotation (for better data reuse)
 - Link datasets in multiple modalities and across multiple organisms
- Andy: diversity across lineages of species (i.e. phenotypes), and reuse of model insects - acceptance easy, expressiveness
- Paddy: systematics, terms used to describe organisms being brought together in an computational form. Data 'conservancies' - an ontological framework to enable reuse of data. Large lumps of data, that needs to be atomized (making ontologies critical).

Ramona: community annotations are often incorrect identifications, how to deal with that?

- some automatic methods

- galaxy zoo approach (based on # of matching annotations)

How to mine current phenotype data?

1. answering scientific questions;
2. Problems to make the technical structures (e.g. the ontologies) transparent to the scientists
 - a. both in usage;
 - b. and in querying.

Challenges: What ontologies are available?

Browsability. Biodiversity heritage project.

Annotation of sequence data.

Ontologies as wire frames.

Can the databases be used to create hypotheses?

Can an ontology be created in an unbiased species neutral manner?

Need to address multiple stages in the process

1. Annotation
2. Searching/Browsing

Graph models are not sufficient.

RGD has some useful functionality.

Sue Rhee: Want ontology to work like Google.

Cassandra had some interesting points. What were they?

Ballhoff - faceted browsing may not work for tall ontologies

How to do legacy data?

Narrow search space

Bug on the ship

Challenges - scaling up reasoners, defining key use cases should be something the RCN does - e.g.

1. narrowing search space given observable characters and local environment
2. rapid identification using something other than DNA (faster) - Maybe using XXX identifications keys.
3. How to use phenotype information from other organisms. re-purposing of data, what kind of questions might someone else do? (e.g. correlations) (e.g. medical implications of environmental sampling data). Enabling the discovery process.
4. RNA binding proteins present in any organism that result in sterility (limit to class of ovary)
5. Extraction of information from digitized library of biodiversity