

## Minutes from Phenoscope Advisory Board Meeting

11 November 2011  
10:30-12:30pm Eastern

### Attending:

#### Advisory board:

- John Day-Richter (Google)-yes
- Brian Hall (Dalhousie University) -yes
- Alan Ruttenberg (University of Buffalo)-yes
- Paul Schofield (University of Cambridge)-yes
- Peter Vize (University of Calgary) -yes

#### Phenoscape

- Paula Mabee (University of South Dakota)-yes
- Todd Vision (University of North Carolina at Chapel Hill)-yes
- David Blackburn (California Academy of Sciences) -yes
- Judith Blake (Mouse Genome Informatics, Jackson Laboratories) - no
- Hong Cui (University of Arizona)-yes
- Hilmar Lapp (National Evolutionary Synthesis Center)-yes
- Paul Sereno (University of Chicago) - no
- Monte Westerfield (ZFIN, University of Oregon)-yes
- Aaron Zorn (Xenbase, Cincinnati Children's Hospital Medical Center)-yes
- Jim Balhoff (Software developer)-yes
- Wasila Dahdul (Senior Curator) -yes
- Peter Midford (Taxonomy Ontology Curator)-yes
- Nizar Ibrahim (postdoctoral fellow, Sereno) - no
- Terry Hayamizu (MGI curator) -yes

### Meeting Agenda

#### I. Introductions (Paula)

#### II. Roles and responsibilities of the Advisory Board (Paula)

1. Evaluate goals, plans and progress; share knowledge; and provide guidance to the project team in prioritizing activities, assessing opportunities, and ameliorating risks
2. Attend semi-annual virtual meeting and annual project meeting

#### III. Overview of the project (Paula)

- Goals and integration
- Phenoscope II award (4 years)
- Project meeting in June 2011 with milestones
- Aims of current award (see agenda)
  - manual curation bottleneck, develop scalable workflow
  - Fin/limb for living and extant vertebrates requires expanded anatomy and taxonomy ontologies
  - integrate evolutionary phenotypes with model organism phenotypes
  - anticipate KB with 2.5 billion assertions from fin/limb curation
  - homology framework for fin/limb
  - KB expansion

- Capstone

#### **IV. Project components (about 5 minutes each)**

##### **1. Curation workflow development**

###### **1.1 NLP (Hong Cui)**

- Phenex workflow curation bottleneck - slow, inconsistency among multiple curators
- Objective 1: extract terms from characters automatically, batch process; identify new terms for review by curators
- software in place; tested; have feedback; improve in 2-3 months
- Objective 2: help curator find matching terms in ontologies; return matches based on semantic similarity (Year 1)
- Objective 3: generate EQ statements for review by curators
- Objective 4: integrate NLP components with Phenex
- [Advisory board comments:](#)
- [JD Richter: How does human curation validation work; check every candidate?](#)
  - [Hong: System will give certainty scores; always check below a threshold](#)
- [Schofield: difficulty with complex phenotypes; using formal patterns for EQ construction? started validating EQ statements by reasoning over;](#)
  - [yes formal patterns known](#)
  - [validation aiming to do - so far has been manual; helps with consistency](#)
  - [Paul Schofield can provide examples and feedback](#)

###### **1.2 Ontology term broker (Jim Balhoff)**

- in curation, bottleneck of adding needed terms to ontology
- backend implemented at NCBO Bioportal
- NLP will produce lists of terms to add to ontologies; script will make requests to ORB and return temp IDs; curation work can then proceed
- also adding term request interface to Phenex
- overview slide of process
- Milestones -
  - bulk term request for NLP output
  - NeXML file provision ID update tool
- [Advisory board comments?](#)

##### **2. Ontologies**

###### **2.1 Anatomy ontologies (Wasila Dahdul)**

- coordinate development and alignment of anatomy ontologies
- major aspect is maintaining synchronization among model organism and multi-species ontologies
- have already created effective communication practices across group
- have created draft Amniote Anatomy Ontology
- focus of year 2 is Amphibian Anatomy Ontology development, handled by David's postdoc to be hired
- [Advisory board comments:](#)
- [Alan Ruttenberg - concerns about xref method for ontology synchronization](#)

- meaning of xref is variable
- suggests making all possible efforts to use same identifier when equivalence is intended
- when a different relationship is intended, use an explicit relationship
- tools that don't understand Chris's expansion mechanism will have the wrong answer
- Schofield - don't some terms across synchronized ontologies have different types of relationships - is this a problem?
  - Wasila - not all terms are given xrefs
  - how does Uberon factor in?
  - Wasila - development in VAO is picked up by Uberon, and VAO is using non-skeletal terms from Uberon

## 2.2 Vertebrate Taxonomy Ontology (Peter Midford)

- extension of TTO
- taking taxonomies from authoritative sources (CoF, amphibiaweb, amniote, IOU checklist, NCBI as scaffolding (substituting out branches as they become available), extinct taxa from PaleoDB, Catalog of Life for synonyms); curating into PaleoDB
- [Advisory board comments:](#)
- Alan: NCBI - users use NCBI; plan to make VTO replacement of NCBI in OBOFoundry?
  - Peter: hasn't been discussed with OBO Foundry; not sure if could meet its needs
- Milestones/Status: TTOupdate, VTOTool done;
  - Owl-individual based taxonomy planned

## 2.3 Ontology synchronization (Jim Balhoff)

- maintaining xrefs with multispecies and single species ontologies
- works by comparing two ontologies xrefs, missing terms, conflicting data, etc...
- new addition: manage terms not wanted to sync
- [Advisory board comments?](#)

## 3. Phenotype data

### 3.1 Evolutionary phenotypes (Paula Mabee)

- fin/limb phenotypes from extinct and extant vertebrate using EQ formalism
- Milestones
  - priority papers for fish and amphibians; archosaur underway
  - AMAO under dev; VTO
  - Training in EQ curation and ontology dev (Paul, Nizar)
  - Future milestones: annotation of amniote data, annotation of amphib data
  - Aaron: cloned skeletal terms from VAO; clone out AAO in Year 2
- [Advisory board comments:](#)
- Brian: What about genes duplicated in fish but only single copy in mouse, i.e., ones that have changed in function?
  - Monte: 30% of genes formed in duplication in teleosts are preserved; usually 2 copies in teleost have subfunctionalized; rare that have new functions

### 3.2 Model organism phenotypes (Monte Westerfield)

- Xenopus- have developed anatomy ontology
  - no phenotype annotation occurring; curating gene expression
  - start training curators in EQ within 6 mo
- Mouse
  - were using Mammalian Phenotype Ontology; mapping of terms nearly complete
  - annotation can continue using MP terms and will be mapped to KB
- Zfish
  - using EQ syntax
  - since July, have annotated 3000 EQ from 100+ pubs; have gone back to older literature rich in fin phenotypes
- Todd - Funding for MOD curation cut by NSF; Paula - request to advisory board for any leads on funding opportunities for this.

#### 4. Homology reasoning & assertions (Hilmar Lapp)

- multispecies anatomy ontologies label and define terms without taking reference to homology; use structural criteria
- have assumed when apply same term in different species that it's homologous
- Plan to put formal framework which takes homology out of what's implied
- homology table with evidence and references
  - "homologous as" statements
- Milestones
  - have logical model for homology between taxa; need serial homology model
  - have beginnings of homology table
  - future: finalize use-cases; which inferences correct or not
    - dataset of fin/limb homologies for testing
    - implement reasoning test suite
- Open questions:
  - how to handle and deduce default homology in scalable manner?
  - what explanations do users need for homology-based inferences?
  - how to assess correctness and comprehensiveness of models?
- Advisory board comments:
- Alan: subclass and homology statements?
  - H: subclass not between species specific AOs but between species specific to multispecies AO; technically no implication; adding restriction that it's part of a particular species
- Paul: equivalences imply homology? formally implying yes ...
  - Brian: level of homology; for parts not systems;
  - homologous\_to\_part\_of
  - Brian: be specific of level of homology: set of tarsal bones in amphibians homologous to those in other groups (tarsal bone h\_to tarsal bone), regardless of number of bones;
- Alan: we need to follow up on subclass vs. homology; work out defaulting issue; need to get homology vetted by experts

## 5. Semantic similarity (Todd Vision)

- Profile: sets of EQs pertaining to a phenotype in a gene or taxon
- terms don't need to match lexically; one-to-one match not necessary; similarity between common phenotypes is less informative; a match in quality along is not meaningful
- Aims: scalable algorithm needed; enable in KB
- Milestones: see slides...
- Postdoc position 2 available in semantic matching and NLP workflow development
- [Advisory board comments:](#)
- [Brian Hall - many genes affect a given phenotype, genetic background also affects result - is it better to look at a single phenotype rather than a whole syndrome?](#)
  - [Todd - open question whether we will be able to provide biologically meaningful matches across profiles](#)
  - [Schofield - we are finding that these methods do work well to recover models of human diseases - they key issue is working out the metrics](#)
- [Alan - be sure to hire post-doc with experience in parallelization, which seems to be applicable to the semantic similarity search problem](#)

## 6. Phenoscape Knowledgebase development (Hilmar Lapp)

- KB brings together evolutionary and mutant phenotypes and ancillary data (eg., gene expression); integrated in form of logical expressions
- Reasoning: many inferences added by reasoner; brings together evol. and mut. EQs
- Milestones
  - polish KB layout and design (almost complete)
  - user testing of profile and tree view; UI navigation alignment (Q2)
  - Public release of new KB user-interface (soft release on-going)
  - alternatives for reasoner and KB storage backend (Q4)
  - complete definition and validation of new logical data model (Q2-3)
  - support for inferring absence of entities from absence of dev. precursors (Q2-3)
- Open questions - see slides

## 7. Capstone project (Paula)

- test against known pathways against KB

## 8. Outreach and broader impact (Paula Mabee)

- summer internships in bio-ontologies , various locations
- internship through USD to Lakota population
- Junior Biocurator (implemented by Sereno's Project Exploration)

## VI. Questions and Discussion

[Alan: Planning to work with OBO Foundry to be part of the foundry? Interacting with them?](#)  
[PM, WD: Yes, aware of principles and implementing them in development](#)

### Questions for Advisory board:

- How to:
  - create a community of KB users?

- make sure there is a case to stakeholders for sustainability?
  - support curation of additional taxa and genetic models?
  - incentivize contributions of data from other researchers?
  - promote community participation in maintenance of ontologies and softwares?
- Schofield: MGI experience of developing community of users: satellite workshops at key meetings; hands-on guided use of databases to gently introduce users
    - Terry H., by email following the meeting “In the past, MGI has had booths at meetings and hosted workshops, but these are quite costly and it is my impression that funding for this has dwindled. However, there has been some (limited) discussion about the idea of 'webinars' as a more cost-efficient way of familiarizing potential users with various aspects of our resource. This might be something for Phenoscape to consider.”
  - Brian: who are the users of KB? e.g., AIBS outreach to teachers
    - Todd: NSF would like to see these broader impacts
  - Schofield: incentivizing, move towards microattribution of data to databases; people want acknowledgement in contribution; worthwhile to look into what
    - Alan: OBO Foundry principle that must be done for ontologies at least
    - Todd: has been in contact with authors of Nature microattribution folks; lots of promise.
  - Richter: companies as a long term source of funding? who might these users be? if used by companies could be huge incentive to keep up;
    - Brian: government funding for species of concern.

## VI. Planning for next Advisory Board meeting

- Advisory board meets face to face with Phenoscape on Friday, April 27, 2012;
- Phenoscape project meeting will precede this on April 25, 26
- booze fund courtesy Jim's presentation award!

### Action items:

- **Send slides to advisory board** (done, Paula 12 Nov 2011)
- **Send postdoctoral ad(s) to advisory board**
- **Send reminder of advisory board meeting**

## VII. Adjourn