

**Phenoscape Advisory Board Agenda
12 December 2013; 1:00 pm Eastern
Minutes**

Project team attending: Paula Mabee, Todd Vision, Hilmar Lapp, Jim Balhoff, Prashanti Manda, Wasila Dahdul, Alex Dececchi, Laura Jackson, Monte Westerfield, Nizar Ibrahim, Judith Blake, Aaron Zorn, David Blackburn

Advisory board members: Peter Vize, Alan Ruttenberg, Brian Hall, Cynthia Parr, Paul Schofield, John Day-Richter, Todd Vision

1. Introduction (Paula, 5 min)

2. Short updates of accomplishments since May

2a. Model organisms (Monte, 10 min)

- Vize: what are 'sources'? A: Annotations may be to published figures from model system databases (i.e. Zfin, MGI), but may also be to bulk downloads from large-scale projects, so more generic term than 'publication'

2b. Knowledgebase (Jim, 10 min)

- Schofield: did something change that required inference of absence and development of SCOWL? A: became newly important with EQ character matrices.
- Ruttenberg: Does the axiom for inferring presence from developmental lineage have unintended or inconsistent entailments? A: No, although we may not have considered all cases.
- Parr: Are you bringing these axioms just for Phenoscape or are they published as part of the ontology? A: Just in build of PhenoscapeKB. Would be wary of incorporating some of these into the anatomy ontology because of broader reasoning implications.
- Rutterberg: Are you communicating, consulting, coordinating with Bijan Parsia and other experts in that community, regarding consequences of axiom generation used for presence and absence inference? A: Hasn't really happened yet. However, have consulted with Chris Mungall on some modeling, but should indeed describe our approach more publicly to get feedback.

2c. Anatomy glossary & outreach (Todd, 10 min)

- Hall: Will the images always be in standard orientation and perspective (proximal up, distal down, lateral view, for example). A: In principle

yes, that is the plan. However, rules aren't written down (yet). Essentially we are using the rules that would be used for an osteological paper.

- Hall: If there is sexual dimorphism, do you have males and females? A: Don't think sexual dimorphism will be an issue. Also have quite a lot of images. Hornbill comes to mind where this might be relevant.
- Schofield: Need to have a way to look at multiple images to compare structures across taxa. Consider a cart-like system to be able to grab multiple images and drag into cart to compare in the same view.
- Parr: Flickr collection images have All Rights Reserved. Also no machine tags. A: We were already planning to change license. Will need to look into machine tags.
- Vize: Is the taxonomic coverage of Phenoscape covered in these images? A: To a considerable extent, will need to revisit.
- Schofield and others: There are a number of anatomy literature sources, such as Hallgrímsson, that could be leveraged and reused rather than recreating. Also researchers with rich micro-CT collections that could be used. A: Indeed. Need to look into this further. Suggestions of resources greatly welcome.

3. Understanding user needs through collaboration (Paula, 25 min)

- Hall: Are the inferences robust to laterality? A: It is possible to do inferences based on an entire dataset, such that smaller inferences of parts of the dataset are not necessary
- Schofield: Problem is there is no gold-standard dataset against which to validate inferences. Could, however, try to do computational validation. For example, can you remove assertions and see whether they can be inferred? Sort of a cross-validation type process. A: Good idea. So far have only spot-checked by hand. Also have a tool now that allows query of which character states entail presence or absence of structure X in taxon Y.
- Day-Richter: What are the conclusions from the observation that there have been so few users? A: Have seen a lot of desire to get access to these same tools. Clearly there is a user base for what we can provide, but need to get better at communicating this. Q: Does it seem like every collaboration of this kind is going to require new custom tools? A: Not necessarily. There are some foundational tools that are going to be needed.
- Parr: Are there people asking to collaborate or jump on board. This would be a test. A: Have to make a decision as project as to how mature we want or need these tools coming out of these collaborations. This also pertains to directions for future funding.

4. Experiment to improve curation workflow (Wasila, Prashanti & Hong, 25 min)

- Day-Richter: Is common ancestry along any relationship? A: Only along `is_a` and `part_of`.
- Prashanti's slide character example, incorrect annotation by both curators...(not 'separated from'); find another example next time...
- How to evaluate 'intercurator similarity - EQ vs E' numbers?
 - Ruttenberg: Need to see examples to get a feel for them.
 - Hall: There may be more similarity than the stats (or proximity in ontology) show. For example, "humerus diaphysis" and "diaphysis" may look more dissimilar than the curators' intents were.
 - Todd: it occurs to me that calculating stats for random character pairs would provide a good baseline (yes)
 - Paula: and look at 'EA' vs. EQ and E
 - Ruttenberg: Should stop worrying about recall and focus on precision. One of the problems with text mining is that it's not clear what the targeted result is. High precision is always a win, but high recall is not necessarily so (such as when it comes at the price of low precision). It's like high likelihood of a low prize versus low likelihood of a high prize. I suggest being conservative. A: For this study we focused in fact on recall, because the driving use case was to make a proposal to the curator. In practice, our hope is to develop a reasonably good score for the expected accuracy of a computationally-generated phenotype; the score would accompany each proposal so a curator could tell how much to (dis)trust it.

5. Strategies for future directions

- Prospective funding ideas (5 min, Todd)

- Knowledgebase tool development for human and machine accessibility of comparative phenotype data, in collaboration w/ driving projects (e.g. like the collaboration w/ K. Sears). Target: NSF Software Infrastructure for Sustained Innovation, complementary to Monarch
- Accelerating curation tools for phenotypic data in the taxonomic & natural history literature. (Target: ABI? In collaboration w/ EOL, Fishbase, other biodiversity resources?)
- Use of comparative phenotype and genome data for discovery of enhancers. Could involve data collection, including targeted genome sequences. Collaborative w/ G. Bejerano & others. (Target: NIH)
- Application to evo-devo, linking development, phenotype, and genetics (w/ external collaborators incl. Monarch, Target: NSF, NIH?)

- Planning Spring 2014 workshops (10 min, Todd)

- Ruttenberg: Have a workshop that invites scientific collaborators.

- Parr: For grant proposals, want to think of workshop to document demand for and impact of chosen direction.
- Schofield: consumers of phenotype data would be focus of interoperability / API-oriented workshop. Particularly excited about regulatory enhancer research.
- Hall: Evo-devo meeting at NESCent may result in a national meeting and a society. May be an opportunity to utilize for workshop. Getting a community together could turn out as more profitable than focusing on a single lab.

6. Dates for face to face advisory board meeting (Todd & Paula, 5 min)

- Schofield: Early June will be difficult for UK. Lots of exams.
- Day-Richter: Google IO conference (first two weeks of June) won't work.
- Ruttenberg: 18/19 June will not work.
- Hall: Consider earlier time for meeting given grant writing in summer.
- Vize: May is out.