**Phenoscape Advisory Board Meeting**

**NC Botanic Garden**

**Friday, April 27, 2012**

Welcome, introductions. Thanks to Todd Vision for organizing, NESCent, and NC Botanic Garden;

Advisory board: Paul Schofield, Brian Hall, Peter Vize, Alan Ruttenberg present. John Day Richter not present.

**1. Paula Mabee (overview)**

* Summarized beginnings of Phenoscape/first round of funding to develop "proof of concept"
* Development of resources (ontologies, software, Jim, Chris etc)
* Phenoscape II: Expand scope to all vertebrates, but focus on fin/limb - had to narrow scope for NSF. A lot of data in fin/limb transition.
* Another goal to reduce bottlenecks in curation of free text
* Current work includes Homology
* Phenoblast development - fast phenotype search.
* Duration of project 48 months - now is month 10
* minimize risks-high risk /high reward
* Trello tool - used for communication, keeping the project united

**2. David Blackburn (‘capstone’)**

* Presented examples of candidate projects
* Chose classic example of evolutionary transition - complex phenotype.
* Use multiple evolutionary studies - generate a 'phenotype profile' along a set of branches on the Tree of Life; multi step transitions; dissect the transition, changes per branch each "stage" along the transition
* Use model systems to test
* phenoblast
* phenotypes might not exist in modern world; use the present to find out what happened in deep Time.
* testing if we can "get back what we already know" and then generate new knowledge
* "big picture question" about phenotypes (pleiotropy, genetic architecture, morphoscape)

Questions/discussion following David:

* Brian Hall: Problem with applying mutant genes to extinct taxa - need to take into account deep time
* Brian Hall: is fin/limb to restricted - maybe scope should be expanded.
* Brian Hall-"are mutants really reflecting the potential transitions in morphology?"
	+ Why just the limbs? why not the rest of the skeleton?
* Paula: we are getting data from the entire body, this is a relatively simply transitions but enbedded in a large datasource so we can easily expand off it
* what else happened in conjunction
* Brian Hall-so the value is it is historical
* Paula: Look at sets of correlated transitions - what happened together with these changes
* David: Important to work on a well studied subject - allows verification of our results
* Paul S.: Where are the gaps?
	+ control of phenotypic data, gaps in the fossil record are beyond our control, what species are we missing sequences of that we need?
* David-predictive framework for who we should go after that are non-model
* Paul-inferring historical relationships, more genomes less risky the inference
* No literature dealing with sequence data, control sequence, timing etc
* what tools are you using to make the comparisons
* Monte:only working on model orgs
* Brian:regulatory elements seem to be very important in the fin to limb transition
* can you use extant taxa to represent something extinct for 400+MY
* Todd: differences in pathways (mouse/zebrafish)
* Paula-new collaboration with Vertebrate tree of life (Naylor), new Phd working on shark data to be intern with Phenoscape
* Aaron-capturing changes in gene expression pattern, associated with phenotype
* changes that result in modifying the gene. First pass
* Paul-interpreting will be hard
* Monte: what is the use of this system - our hope it is it will be a hypothesis generating project.
* Paul S: great value in having a hypothesis generating tool
* Monte: testing the hypothesis beyond the scope of this project

**3. Monte (Model organism ontologies and curation)**

* 2 types of data - morphological, genetic
* getting data and curation for free - through model organism databases, funding by NIH
* rat is not up there because they do not produce the same type of genetic data as mouse
* milestones - curation of phenotypes & gene expression
* Incorporate phenotype and gene expression into KB
* develop pipelines for uploading xenbase and MGI into Phenoscape
* Developed such a pipeline during Phenoscape 1 - still ongoing
* 61,000 phenotypes for 4000 genes, 800-fin, ~2000 publications + direct submission
* MGI-6100 limb phenotypes for 2500 genotypes
* not as EQ so must be mapped into EQ (done by PATO), keep up to date (Terry by hand), no automated pipeline
* Many EQ's required for 1 MP
* Curators trained at NESCent, using phenote tool
* Xenbase, all is\_a and devolps \_from are complete, defined
* plans: import gene expression data into Phenoscape (ZFIN)
* continue curation (MGI)
* maintain mappings, import EMAPA into Phenoscape (immediate plan)
* develop pipeline to import mappings into Phenoscape

Questions/discussion following Monte:

* Alan: are these knock outs? wild type? during development
* in the case of unusual anatomy
* when do the part relationships hold?
* Terry: relatively straightforward in limb
* Alan: stage version terms valid but are the relationships still valid?
* Judy: this is a side discussion because we have resolved this.
* Xenbase plans: continue curation; develop pipeline to import data into KB
* Judy: we will import aspects of the data
* Judy: one thing we were discussing, not feasible to import all data, but we will bring in aspects
* Deliverables
* Have a complete phenotype and gene expression data for 3 model organisms
* paper: methods & challenges of integrating such data
* Paul S: tell us about method problems
* Monte: staging is different in each group, gene orthology "the same gene" and best way to approach that
* resolvable but thorny
* Judy: comparative phenotypes - challenging and interesting - doing it independently of the underlying genotype
* expression data is huge, we want to relate it to phenotypes, not just comparative phenotypes, new view with context of evolutionary history, new view of model organisms
* interesting what one can and cannot do with this data
* Paul S: how many are we looking at? how many phenotypes? what kind of granularity?
* Terry - hundreds - has the number someplace. Relevant terms much less than that.
* 1:1 MP anatomical term to quality
* Paul S: Phenotypes description is limiting factor, if not enough detail in description than you are unable to advance. How well do you have to describe it to have any power
* Could generate test case - known relation - then reduce granularity if description - when does it no longer work?
* Jim: like Degrees of freedom
* Judy-relationship between the phenotypes
* Paul S: are you also covering vasculature/skin? Vasculature very important in developing limb.
* Paula – curation creep; it will drift to any other characters that will be used in the papers.
* Peter V: rarely covered though
* Paul S: how much integument?
* Judy: haven't brought in the GO data,
* Paula: integument data well covered in papers.
* Brian: Vascularity is critical, just came from a conference which discussed this. needs to be addressed.
* gene duplication, do you see results of this?
* Monte: for 30% of the genes the duplicates are retained. Fornthe others, every example, the expression pattern/function of the duplicate is less than the sum of the two.
* Brian: Zebrafish better at neofunctionality than other taxa.
* Judy: abstract level of gene, not talking about isoforms etc, need to look at power of data e.g. gene sets
* Brian: MicroRNAs, need to bring this into database, firsthand signal of regulation
* Judy: model organisms are good - this science continues independently of out project, but data flows continuously into our project
* Aaron: clarification, gene expression captures it in model organisms
* David: rare opportunity to tease duplication apart - what happens to genes? great opportunity

**4. Paula (Evolutionary data ontology development and data curation)**

* provide phenotypes for the transition (focus on fin/ limb skeleton)
* develop vertebrate ontology; discussing how to integrate
* retention preserves the duplication of terms in TAO to VAO
* Alan: when you say you have agreed on terminology, is it names or definitions? (ontology)
* Paul S: is it adults only?
* Paula: Adults but working on embryos, across MODs and evol groups
* Alan: not templates, but structures. One subtypes but not copy. So no separate id for different ontologies?
* Paula: Import into other ontologies, share logical reference
* Alan: where do humans fit in?
* ALL: we have some solutions for this - maybe
	+ worked closely with model org ontologies
	+ overlap between ontologies is an issue
* need proper attribution of ontology contribution
* Alan: importance of attribution, serious omission as is now
* Any time you see a fact, you need to know who contributed it and potentially project or group link
* know that FMA has problems on this
* Paula: agrees, good idea
* Paul S: concepts on existing ontology are inconsistent - groupings, higher level stuff etc really don’t relate, and FMA will not be rewritten.
* Paul S: what do you mean by logically consistent

* Alan: but we do talk to Melissa and try to re-work if needed, push to change. Need flexibility, FMA will change; can be a bridge to FMA
* Melissa: 2 things - first - modular concept of anatomy new concept. have an upper level ontology to coordinate the others. Didn’t know how to build multi species ontologies. second - in uberon we have reconciled some of these issues and have quite a bit of consistency. have some mechanisms to reconcile.
* Paul S: that is the power of uberon, impose equivalency assertion, but poor granularity
* Alan: not looking for agreement, but where we are in process, get sense of convergent thinking
* Paula: we are not intending to be a naming resource for taxa
* Alan: issue, altid or more than 1 id for entity hurts. general users of taxonomy desire to go to a taxonomy resource, best of what is around. A patchwork is needed and desired.
* David: a single authority?
* Melissa and Judy: never happen - never just one ontology
* Alan: not we have to deal with, but keep in mind this is a problem the larger community has
* Melissa: we can help guide the community, but agree Obo is a reference
* David: very circular - need to talk about this. Coevolving taxonomy.
* Judy: breakout group because this is huge
* Paula: move to Protege 4 ontology editor
* Alan: talking about practices that will push down to individual ontologies or stay above that? Each group wants a single name space, problem with re-use.
* Jim - most of this applies to evolutionary annotation ontologies and not zebrafish etc.
* Chris: speaking from experience, you find inconsistency, but we find them and correct
* Melissa: goal is one of first attempt
* Model orgs use equivalence statement. But in the future we can create things in OWL.
* Alan: what do you mean by sustain? short term, 10 years?
* Paula: thinking beyond the project. who has control over ontology?
* UBERON has different hierarchical structures. Attribution is key - ownership is tricky.
* Peter V.: more worried about quality or credit?
* Alan: play of contribution and quality - transparency

**5. Hong's presentation**

* bioportal web services ready to use, waiting for Jim to integrate it into Phenoscape
* biocreative workshop results: positive feedback on usability and reliability
* but needs improvement
* issues in translation, unmatched phrases, pre-composed terms in ontology
* All the problems appear to be fixable - granularity a bit problematic however
* immediate future plans: improve charaparser results (by end of summer)
* integrate charaparser and term broker to phenex (~6 months)
* integrate new system into data curation workflow

Questions/discussion following Hong:

* Alan: what is being generated by charaparser that sis being reviewed?
* Hong:3 categories, words part of E, words part of Q and terms it does not know; when you look at any example you see context so the editor can decide
* Alan: Since some terms can be both a structure and descriptor, if once a structure always a structure?
* Hong: no
* Alan: lexical information is not often in ontologies, we should release the processor once it is ready
* Hong: that is part of the plan - intend to make ontology terms public
* Alan: what is the process?
* Jim: workflow is not exactly worked out yet -
* Alan: seems easy to do - designed an ontology broker that does all of this
* Jim: reprocess from annotation work -
* Alan: in my opinion the wrong path has been chosen, I can point to the documentation after that will correct the potential limitations in the term broker design
* Hong: we certainly did not do everything right
* Hong: good strategies that prevent automated system errors - any suggestions?
* Alan: companies that do this use a dual review system for any automated finding, if both then it is accepted, if not it is discussed. It is often too easy to make mistake
* Hong: right.

**6. Hilmar's talk (Homology)**

* the exact way of determining homology we are using is not as important as the framework
* construct logical explicit model for homology across groups including factoring in which way do we determine if they are "the same" and on what evidence
* we need it to be correct and complete (i.e. not miss true homologies)
* the relationship is ternary, uses related to an ancestral entity
* phylogenetic homology
* treat as macro, because you can't have class level homology assertions
* or use a relation chain
* Alan: don't you always need to specify E3 to ensure you don't get errors form hidden context
* Hilmar: often it is not necessary - last common ancestor logic -
* Alan: assumes taxonomy is static, if it evolves it no longer matches original statement
* Hilmar: good point
* Hilmar: various stages of progress
* Chicago meeting outcomes - should just have one meeting - homonomy terms to have only a single class - defining 'sameness' and necessary user-control for discovering phenotypes is challenging
* Immediate future plans: homology reasoning use cases, reasoning rules, model for iterative homology etc...
* papers, presentations:
* Logic treatment of assertions of evolutionary sameness for organismal parts
* expect well defined relationships in OWl 2, expert curated ontology of vertebrate fins and limb homology
* Challenges: what explanations and controls do users need for homology-based inferences
* how do you asses correctness and completeness
* how much control does the user need?

Discussion following Hilmar’s talk:

* Brian: how do you treat a character homology across tetrapods, how far back do go back to claim they are homologous? back to common ancestor?, origin of tetrapod?
* Chris: can choose level you want
* David: we may not need an actual homology statement - applying same concept (no risk something else is a 'humerus'
* Hilmar: if the bone we consider the same is named differently then we need to apply the evaluation and classes
* David: our way of dealing with homology is practical
* Alan: that relation, developed \_by, actually goes back to first organism, how founded does it need to be
* Chris: that relation pertains to anatomical structure, not organism
* Alan: the relationship as stated is between two entities; the challenge is between two entities does it hold? I can't figure out if you need an extra qualifier to bound that
* Chris: not completely formalized yet -
* Alan do you need 4th character?
* Jim: even if you think that all structures are derived from primordial structures...
* Alan: do you want inputs bounded or not.
* Brian: fin limb good examples - humerus homologous in all vertebrates. Tiktaliik - different views when you start working on these taxa.
* Is the most distal element in transitional form is a carpal or not? which authority do you take and you need to flesh it out
* David: digit morphology - non consensus - need a lot of flexibility, unfortunately we have a 4 fingered model and when losses occur homologies are difficult
* Monte: raises important question - these statements really need evidence associated with them. attribution and evidence codes
* Hilmar: all have evidence
* Paula: in the first Phenoscape - there is a separate ontology for evidence types - didn't have all criteria for homology
* Brian: Shared ancestry defines homology, but we work with it positionally
* Melissa: another thing we did - defined what evidence was used.
* Alan: plan for this group to review those definitions?
* Melissa - could review this evidence
* Paula: in compiling the homology assertions, we may find very few direct statements about homology
* Judy: same in GO; this has lead GO and other groups to state the source of assertion
* Hilmar: no explicit provenience for homology
* Brian: so can you assume if you use that name is it homology by default
* Brian: structures not homology based in ontology
* David: we should talk about that, since Brian and Chris are here because we are treating it that way
* Brian: so you are going to assert something is homologous not assume that
* David: should talk about that later
* Hilmar: how to deal with default homologies - still needs some figuring out

**7. Jim (Knowledgebase development)**

* progress to date: Phenoscape I
* In progress: complete re-working of data loading and resoning process
* move to OWL reasoning (off the shelf)
* reasoners don't scale, so must take "slices"
* future plans: immediate: Phenoscape I offical release
* complete OWl version of existing KB
* documentation
* evaluate back end system
* import new MODs
* presentations and capstone
* Mungall et al The Ontology-Based Database
* Phenoscape I capstone paper

Questions for Jim:

* Are we enabling users to use the KB?
* How would we know?
* How best to develop features useful to audience broader than this project?
* Alan: any tool you build that work on sparkle back end you need to include a link that lets them pop that query in and let it run
* Todd: can we get a tour of KB?
* Paul S: Use of OWL 2, using subset ?
* Jim: parts of reasoning whenever possible, but we have some absence phenotypes
* Paul S: but using negation may cause problems
* Jim: we have work arounds for negation semantics
* Alan: stay in EL but not go outside DL?
* Jim: yes

**8. Todd (Semantic similarity)**

* phenoblast
* Find needles in EQ haystack
* profile: set of phenotypes associated with branch on the tree or genotype
* use KB to find similar profiles between or within organisms
* terms need not match lexically
* no need for 1:1 match for biological or methodological reasons
* similarity in common phenotypes not informative
* match in quality alone is not meaningful
* objective: make it fast and statistically justifiable
* Progress to date:
	+ Have been gaining experience with existing semantic similarity measures on the Phenoscape I dataset
	+ Have collected (some) of the relevant recent literature
	+ Have been built into KB

Discussion following Todd’s talk:

* Alan: mods: change the term from abnormal
* Todd: problem worse for evo data, because these are in relation to another organism, but on MOD side it is a wild type
* Alan: interesting, the thing we flag as problematic is the thing with most info, so double problematic
* Brian: Why are you making the comparison? Why chose these?
* Todd: just a pair to illustrate
* Alan: but should the reason the match be clear? Will user be able to tell why matches are there
* Todd: good point
* most of the existing semantic methods are pairwise
* Alan: do you have a feedback method to reply to the match
* Paul S: some matches will look completely random
* Paul S: need to present the evidence to user and explain what that means, if they don't understand the results then they will ignore it
* Paula: how do you show that to the user?
* Paul S: transparency is key, if not it becomes a toy
* Alan: have a "soft why" so they can tell why they got what they got
* Paul S: these are all hypotheses, so they need to see how this has been generated
* Paula: Provide feedback?
* Paul S: They won't do that
* Alan: are you explicitly thinking about parallelism? You should
* Todd: want to have evaluation of different methods, do higher level analysis - look across data, in a way we currently cannot do.
* How will users take advantage of this capability?
* Paul S: sets of phenotypes and complex description, how will you relate the EQ statements together in Owl, to produce phenoset
* Jim: What set of EQ are you working with, the complete set may be broader than you need for a profile. If you focus down to say what changes on a branch
* Description saying the organisms has a part that is an instance of an anatomical class and quality that is part of another class
* Paul S: this is entity centric
* Jim: yes
* Alan: what does that mean
* Chris: pattern to represent expression
* Alan: search is from facing way in, do you have a way of saying it working
* Todd: is it producing correct results, is it scalable and then the design, all need to be met
* Alan: you need your users to think it is biologically relevant very easy to miss what’s happening
* Alan: make sure it reaches people

**9. Nizar's talk (Outreach-project exploration) and subsequent discussion re: general outreach for Phenoscape**

Paul S: what outreach operations are used for interest groups (i.e. for scientist therefore workshops, planning meetings etc.)?

Todd: rethinking, because in Phenoscape I we gave talks to our potential audience, but we may rethink it because it did not attract a lot of attention

Paula: perhaps a workshop in Sept, with systematic biology, devo-evo communities, but we had little to show them

Paul S: you need something to show

David: the user community is much different than we originally assume, we need to organize them

Paula: maybe the system biologists would be good

Paul S: molecular evolutionary biologists

Todd: functional genomicists

Alan: why can't we see the 'whole story' (entire text description) on the results page of the KB

Can't understand what is meant by 'shape' -

Hong explained it's because there is no term 'Y shaped' - so use higher level term 'shape' for coarse annotation.

Paula: In evolutionary biology the character is denoted as presence/ absence of the Y-shape character, assuming that it is the only shape variation

Hilmar - Alan makes a good point - we understand the character state translation, but it is not obvious to others.

Alan: easiest thing to do is to label on quality, as "changed quality"

Jim: completely true and is confusing, but these EQ characterize the character statements

Hilmar: one further comment - not fundamentally different from model org annotation.

Alan: There isn't a change in 'present'

Paul S: we came across this in human studies, the problem is we are using relations as characters, with respect to something else you see X, but you should annotate that it is either X or Y, not present of absence of Y

* Formally its a problem, because its canonical, not instances of a taxon, the question is what is your relation against, so what is normal? So transform it into a shape data, i.e. straight versus Y-shaped, not jsut presence or absence of a Y shape

Jim: problem with this quality/this particular example

Hilmar: this is a bit of an artifact of Phenoscape I due to us not being able to completely annotate due to lack of terms.

Jim: also about the way the interface represents it

David: problem is the nature of the data

Paula: this author didn't specify the shape of the element - so curator can't rewrite the character description

Alan: we should flag this

Aaron: would an improved quality ontology solve this?

Jim oh yes, it would help – doesn't handle absence. Is being fixed in the new representation.

Hilmar: if you gave me the OWL expression I would know what you mean, should we provide this?

Alan: most won't do it, but for quality review I would include it in an non-intrusive way

Source is not a source, it has both the source and the phenotype, so not a source, it’s the meat, if you saw the owl you would get that but you need to look hard for it

Jim - when we started development we thought EQ can serve as a translation of the statements - is not really the case

Alan: should of asked me, I could have told you that

Jim - ZFIN similarly coarse

Brian: Implication that any gene mentioned is a zebrafish gene?

Jim - Yes

David: eventually should annotate the model

Jim - yes, would be a major change.

Suzi: I love this display, but table off to the side causes some back and forth, we could improve it and get more information

Alan: put a paraphraser at the bottom of the page would also be useful, i.e. Clupeomorpha has 3 out of 20 that is present in the sister

Peter: species names have to link to something, so users can see what they are.

Paula: will become a problem when more than just fish are included

Jim: this is a general problem, people want to see higher-level groups to see where it fits, but these groups are nested in other groupings and most people don't know every group

Suzi - best thing to do is to find some arrangement amongst the people dealing with the display

Paula: toggle between common name and scientific name

David: this is going to be a huge problem, we need to do something like adding a "fish" label

Alan: a kinda of, then a kind or a kind of a

Melissa: maybe a little fish icon

Brian: looking at cell tissue hierarchy would be good

Alan: asked about a related character

**Advisory board report out:**

Alan: will go to main issues

Paul S: first, doing a great job – terrific; some of these things have already been mentioned

1. Are there other large collections of useful structured data that might be incorporated? -already discussed micro data, do you know any other datasets because the more you have the better it will be. Suggestion to add additional tools to end of outputs; analyze results via APIs. Simple and little investment to add but may have large utility (and appeal).
	1. Can we find change sin gene expression in two lineages, what will you get, 1 gene, 10, 1000? since most of the time its not 1 gene 1 phenotype,
	2. Might be good to have additional tools at end of output.
	3. E.g., gene list in terms of pathway or go function
	4. Good ones available with IPI's bolt onto your interface
	5. Gives a lot of added value.

**2. Outreach,** at this stage, focus on who the users are and strategies for it.

Who are the users? Targeting the right kind of people.

Have a group of people go out and sit down in a key department and do a small talk and move on "flying squad" (like Ensemble); a group of people (or even 1) to go out and work with other people; gypsy bioinformatician;

Have a domain expert user workshop - targeting a dozen people or so -

Have them bring specific questions. when the database is better populated,

sit down for 2 days and try to answer those questions. Good to generate converts.

How are they using it?

**3. Anatomy Ontology**

VAO/Uberon follow-up. Avoid overlap, redundancy. Sense that it is still unresolved (getting UBERON and VAO working together) Uberon and VAO working together, Alan can help lay that out, both sides need to modify

Avoid a split - should be unified

**4. Quality of phenotype annotations**: can you develop metric to test this? one way might be to take this annotation and then try and regenerate the phylo relationship between species, I.e. can you recreate the pattern.

* Brian: issue of how much variation in anatomical detail can be captured? Documenting changes in phenotype is one of the main goals.
1. **Numerical data requirement for understanding phenotypic/gene expression**
* Are you capturing the necessary info, is it important? I.e., if the trait is 5-7 fin rays, and you have a mutant does that, can you find that. Count data
* Alan: variation within taxa.
* Paula: can't reason across it yet.
* Alan: if the normal is something, say 5-8 fins, and then you see a mutant that only gets 6 (so are we capturing normal distribution or not)
* Do you capture normal variation within one species
* Judy: very important to capture it.
* Peter V: even in basic set up you are not capturing the necessary info
* Alan: but that is two things are getting enough info
* Paula: we do capture variation - just can't reason yet
* Judy: we need to get more info than simply giving a ref to go too
* Todd: data very incomplete - have to keep that in mind
* David: we are capturing variation within a matrix - but there is a lot of variation people don't include in a matrix. should add morphology from non-matrix publications.
	+ and its not" phylogenetic informative"
* Brian: case where you have numerical when there are more than 1, becasue if that is people looking for genetic data need to know that. Is it high, low, present across the whole arch etc
* Terry - info is captured in some way = especially pattern
* David: but in a matrix you get "1 or more than 1", and more is useful
* Limitations of matrices

**6. More representation of groups (i.e. birds and wings) for more data, specifically chick**

HPO good source of hands and feet for human data

**7. Effort allocation** - is there enough focus on development? Jim is overloaded.

* Judy: useful what are the priorities
* Peter V.: Would more database support help the project? more data and more sources?
* Todd: not just EQ data, other sources
* Monte: needs a driver to be entered
* Aaron: Jim is overworked..
* Alan: we don't feel like one is getting too much attention - but just want to know how resources are allocated.

**General discussion (whole group) re: issues above and more:**

number 2, limited resources, in balance in things that are done and not done, and goals that need to be done. Need to get data in in a robust and tractable, and getting it out. If not value to users, then what’s the point. New kind of specialist or someone who tends to it, the more you get out is what is important. Data out issue is so important

Alan: outreach efforts crucial here - ask people what they need/want to know

get in gear for tight connects to get things out

which are feasible, advised and so you can make a plan

sooner you get a tight loop that wants to ask a question, tries and can't and tells you is very important.

Need to get users

Monte - strategy is part of that - who are our potential user communities - we can figure the details out ourselves. who are the potential users?

we are a little ingrown, need new use cases

Brian: a little in term, a obvious end users. Go to them (say SVP) and talk this up

Or if you want molecular genetist, is it going to meeting?, is it a short article in a target journal? and list contacts

Should have had target audience in mind at the beginning

Alan: you need to list this and then we can flesh this out

Big syst biology audience -

\* people who perform gene set enrichment

\* we are providing gene sets

go off into GO, get a hint into enrichment, the go find out. If you can provide them with something new (like phenotypes) you can get them interested, so seek them out, tell what you can do for them and get them as clients

Paul S.: Molecular phylogeny - exactly the target audience

David: elaborate please

Paul S: establishing phylogenetic relations based on say hox, and body pattern evolution because it has a phenotypic pattern

Paula: is this a tool within a specific tool? targeted evo/devo, but not much luck so far

target DevoEvo, but small group, and even if we get them in we need more

Brian: evo devo meetings - not one single one that all go to

People who are doing it have a handle on the gene side, but not on the anatomy side so

Tie genes and anatomy together

Paula: did we maybe miss a big target audience?

Todd: Molecular evolutionists

Paul S.: they are basically developmental biologists

Todd: but using species variation too look back in time

Alan: right now you are speaking of classes of researchers - alternatively, look out for particular groups that are doing something you think is interesting and that this could help and invite yourself in

- word of mouth is important - don't have to go to groups.

Target who you think can benefit

David: we had some discussions with individuals who might be part of the target community

Todd: another way to look is the current is open ended but if we need to reach out we need to look at application we serve in a simple way that can be grasped

Alan: but maybe not present an interface, you go in, figure out what is needed, write up the script, run it and give them the results

Judy: when we go to a developmental biology meeting, it’s a very interactive environment.

more generally this project is plowing new ground, Comparative phenotypes

new way, project is evolutionary based, but there is many other applies to many areas

start looking for correlations etc

setting the paradigm of what is required is needed, what are the tools that present comparative phenotypes - Big question

the more this grows as the model if we give them this

Brian: Society for Integrative Biology - has several divisions; 1200-1500 people, multiple divisions so you can get could get associations with one of their divisions - cross divisional symposium can be organized; all overlaps so it is an obvious target

Peter V.: big community, what is being captured - people want to see fine detail

Brian: was just at a anatomy meeting with a whole morning on microRna

Peter V.: quality terms don’t capture all the variation - if you don't capture the richness that would approve the usefulness

Paula: this is due to Pato, which is why we have these Q states like shape

Peter V.: E is very well worked, maybe more focus on the Q

Melissa: PATO ontology has been source of many issues. In the comparative phenotype plos paper, we didn’t find many advantages using pato.

Complexity and time it takes to create something use post composition is enormous, might not be worth it at this time

Balance between presenting to the user and retrieving data

PATO is a tool to retrieve data

Alan: but it is in computation, and you retrieval

Judy: where is the useful computational level of research?

Alan: potential target audience in clinical, bone regeneration and healing (Army), and congenital deformity, do exploratory calls to see if they would use this; clubfoot

Paul: American College of Surgeons has a series of papers on morphology looking at terminology of skeletal malformations

Paul S: can present a name if you ask after

Paula: do you have specific names?

Paul S.: yes

Paula: should we add more tools?

Paul S.: all ties into your users

Paul S: pathway membership analysis, overrepresentation analysis,

Jim: no one is giving us ideas of what they want

More speculative now

- data services, data download, but no one is saying " it is missing x data" we need someone to complaining " i need this"

Alan: no if you are saying that if you are looking at the genetics you need to ask yourself how can you make it meaningful to them, so need to hook into tools that into this info is it going to be meaningful?

**Discussion of pathway analysis, network analysis**,

- customize to individual users; if we are saying that we are trying to vie info and hope it is meaningful, - is it and how to make it? If we think we are trying to be in the business of… Gene set enrichment, pathway analysis; if we are trying to do this – should we being doing interface right now; devote energy to people; let Jim write specific queries;

tying genes to cells extra level ;of analysis; pathway analysis very grass roots effort; bringing in new kind of data and applying to problem;

basic question - should focus on interface now, or not?

Aaron: get Jim doing specific requests, and when you go to conferences turn interest into users

translating interest into users is key

bring in new data types and pipeline; do one-off with what we have now. What is the question? Do you have the data; get data if required.

Brian: tying genes into the cell and then to Phenoscape.

- include info in where gene is expressed, so if gene is in fin and metapterygia it gives a whole new level of analysis

Alan: build tool for singular uses, did it for them, got more and finally enough asked their boss to request a tool, then we built it

Paula: we haven't thought of a grass roots approach

Aaron: don't abandon this strategy, but do it in conjunction

Monte: didn't have an interface

Jim: as far as capstone, no need for interface

Paula: does Jim have time for these efforts? talk about a way to divert effort at this point in the project?

Alan: if you are going to do targeted science - good idea to have Jim doing this

if it is going to take away from improving the interface and that will target users, than it’s not a good idea

Hilmar - likes this idea

Paula: suspend interface, target individuals, get some cool science

We now have post-docs

Aaron: In know budget it tight, but should we look at budget re-allocations at this point?

Hilmar - we have something to show now - we have problems turning visitors into users. take a different approach.

Monte: possible to do both, bring in new data types, pipelines, expand data we have but also target some people who work on fish and can use the data we have, do a one off with what we have now while we work on the Phenoscape

Judy: you can approach John Lundberg

- people come in with a dataset, and you have to ensure all the knowledge needed is there before they start

**SD September workshop discussion:**

* Paula: idea for September workshop
* Todd: time set aside for workshop; what to do with workshop?
* Vertnet, variety of bioinformatic resources, that do not link to us, but we could reach out to them
* Systems biology community maybe
* phenotype enrichment
* help us understand the machines to abroad and incorporate into their system, informatic user base
* Todd: ideas not fully fleshed out

**Linking to Museum data discussion:**

* David: manages museum data - you can drill down to a museum collection number with a huge amount of data associated with this, and this userbase would love that
	+ museum community would be good
* Todd: all these users have their own userbase, who would click a link
* Paula: Paris museum happy to jump in and link to Phenoscape
* David: California Academy of Sciences could do it now
* Paula: should we do this?
* Judy: Do you have enough data for this?
* Alan: don't answer now, call them and see if they think you have enough data

Monte: really have to look at what our time investment would be

Brian: do you want to spend 10 months targeting bioinformatics or spend that time to select ZFIN

Monte: we have talked about the space based data already

Brian: target 4-5 labs; think about combining lab projects; integrate for broad range of applicability.

Judy: have to build up over time - easier now; need a publication or a strong user community

Monte: each one-off may generate a new publication

Alan: that’s why you are doing this

Judy: make sure choose the examples

And each one builds up the

Brian: look at last couple of papers from this lab and think of how you can ask a couple of interesting questions a broad range of applicability

Paula: that's a good idea.

Brian: fish bone bio meeting in Portugal, 80-100 fish bio people, next spring

David: if we could get some publications for ICVM - need them by then

Question: whether to make it a priority to enable community to add data? Chorus of ‘no’ – communities do not come through and add data voluntarily.

Judy: repositories: make it standard in Phenoscape that attribution be recorded; if we have already created an easy method for curators to put in data and point to it if ever a call for such an institutional db; repository.

* TJV: why has community taken off with hao contributions?

**Homology problem**

* Brian: one topic we did not put on the list was homology, surprised that it was not based on it, and I would not put any more time to thinking about it, let homology simmer in the background; ‘default’ homology ok.

**Anatomy ontologies:**

Alan: issue is some question about, trying to avoid two efforts building same content (Uberon and VAO), would like to have one resource,

- some of the issues are about control, sense of responsibility for the community and a part has to do with quick changes, who can do that couple of things about why do you need control

- Less autonomous, if you have to go through Chris

- but from Chris, "we have an approach" it seems to work,

- couple of ways to handle this

Suzi: attribution versus maintenance separate issues

- in what ways can VAO be part of UBERON?

Uberon does not have to be maintained as a single unit, here is a part of this, and the role of maintaining that is by the people within it. Does not have to be permanent; For this period of time, e.g., 6 months, you give away a chunk; see if it works and change model as needed

Chris: 2 models, one is granting editing rights for all Uberon, or for a single unit

Alan: I think that maybe if you have a single part, and it’s all you can see, then you can work on that

Chris: here is the problem, anything you pull out, e.g., the skeletal system, drags in the cell, and germ layer sections.

Alan: but if there was a simple solution, you would have, you have to think do you want to take on the responsibility for all of Uberon, or do you split it?

2) Go to an attribution model: Granular attribution - independent, satisfy the demands of your project when needed. make sure you attribute each term (should be done in any case). extra disadvantage - credit less concentrated.

different than if Chris puts it in 5 min later, than when you look over Uberon, you can see who contributed, and see who has done what effort,

little more complex on our side, life simpler for Chris,

Chris - not commit rights - make edits and send a diff to Chris

Check out Uberon, make edits, send diff, and we vet it before it is added,

Alan: third model, in-between step

good model if reviewers are prompt and responsible, and get another layer of eyes looking it over

Jim - keep in mind that need unified ontology for all mods, if our group is making something that is not always going in, we will get frustrated,

Alan: need to establish a contract in all of these cases. That’s what you review at the end.

- model if another group needs to be added, and you can pass it on to them

Alan: Hate name Uberon and think it’s harmful; fighting words

Who accepts responsibility for maintenance on behalf of a community? Who receives control over quickly editing the ontology?

Model A: VAO as a part of UBERON

UBERON does not have to be maintained as a single unit. Responsibilities for certain areas can be designated for a (potentially renewable) periods of time (for example 6 months) to certain stakeholder communities or their representatives.

Issues with modularity: Goals of UBERON are broader than those of VAO.

Model B: VAO maintained separately

Credit is not as concentrated (but need credit model anyway)

Regardless of model and solution, need to establish a contract that gets reviewed on a regular basis.

**Term broker discussion:**

Alan: worry is the most important communication is between requestor and curator

has to work for any form of automated solution.

need to preserve that

tools need an id, if term goes away does this id go away, and if you use NCBI does it go all the way - How does info get to the curators, how do you close the loop -

If they just give you and ID, it is not enough

Hilmar: what is the tool for the curation, editing ontology, pushing the nature of solution back to request that curator decides to act upon, that’s why use web protege - would be integrated into the term request - could actually pull within web protege within bioportal.

get the notes, change them which would trigger alerts that

our data curation tool could interact with

Alan: sounds like that could work

Hilmar - at the time there were a few pieces missing (protégé still not that compatible) -

still missing,

Still can't change ontologies back to bioportal, gaping whole is what takes Owl's place

what is the ontology editor that can interact and integrate, pull the note, push back info to API

that’s why we use protege 4, does not have an interaction with bioprotal, but editor would fill that hole

Alan: you need to follow up, because how do you force them to respond?

Hilmar - we are using community interface, forums, mailing lists - the actual discussion would not happen on a 1 - 1 basis - but rather on mailing list etc

Suzi-still confused, isn't it predicated on everyone who edits an ontology

Jim: solution has to work on ontologies we have no influence on, if we need a new term in GO, we can monitor the request from our users. Responsibility on our ontology manager (Melissa) uses Go mechanism to get the term, the go to bioportal interface adn give the go id that fulfills the requirement

Alan: What does this provide, a bad mailing list

Jim: only part we are using from bioportal is that asking for provisional term ID

have services to update to a permanent id later

Alan: that’s trivial, the way to interact to get a unique id is not worth it and not controlled by this project,

Suzi - but then Melissa would have to make 2 requests

Alan: nothing useful from bioportal,

you know you have half, you have a request that comes, in, it has been answered, you need closer, you have another generated,

you have to do all this work

Todd: are there re-usable components?

Alan: have that problem no matter what, need to put the effort in anyway, but like need it in a way that can be used by everyone, I have the specs, bioportal hasn't

Judy: take Melissa as a use case, need a new term, holding up curation of editor

Hilmar - typically the people adding terms and the ontology curators are not the same people

Melissa explains what happens now when you need a new term

**Discussion of whether we should work on mechanisms to get the community to submit phenotype data.**

Todd: one issue that is a pipe dream, we make this a win for the community so that we can bring in people outside our project to provide data

Alan: forget about it!! Does not happen - lovely dream.

Paul S: getting the community to submit data of their own accord will not happen; only unsolicited ontology contributions I ever got were insane

Suzi - this will only happen when people use the data - and when it becomes good for your CV when your data is being used....

Paul S: is that now, been available for 2-3 years (the 'my contributions system), but no one cares.

Brian: in the use, the primary data needs to be deposited in a repository, if you can get access to this or be the repositor than get immediate access to anything funded by NIH or NSF where it is plonked

Paul S: You need institutional funding first before you can become a real depository. to last a decade or more, legacy database

no one would agree to continue to accumulate data and curate it

the other model, the woods hole model is more feasible, and more chance of joint funding

Judy: Issue of repository, lot of effort to have them,

Maybe momentum building - we can design a system - we are brining in data, attributions should be accessible - so we could take in data in the future - be prepared. if the directive comes, are we prepared to accept that data. right now, unless it is required, not going to happen

if you already have a easy web based methodology and point to it, right now, don't see it happening

don’t see people doing that that now unless they are required.

Melissa: if you hit the author just as they publish or the publisher and say "is this correct?" they feel obliged to provide the rest.

Jim: HAO has some community interest, but they don't get a lot of input to the ontology, talk about labels,

have an idea, you publish a paper, submit the paper to the HAO, get the URI for glossary, stick it in a table at the end of the paper

HAO community seems to be more interested in the ontology

No idea, why more take up in HAO

Brian: involved in embryo project, goes onto a website, each article gets an ISBN number, like a chapter

If you can do it for Phenoscape, it would be useful

David: it’s like a collection. Maybe Dryad model.

Monte: at ZFIN, you get a publication, looks like a journal

Brian: could be done at Phenoscape

Monte: taken advantage of quite a lot now

people submit data directly without going to a journal

Judy: did the same thing when you could no longer publish mapping data, include abstract

Monte - some people even include their methods

Judy: and you can cite it

Terry: we get large-scale screens. People deposit a lot of data - then analysis begins.

Hilmar - do you provide peer review?

Monte: the curators do, but not reviewed scientific

Peter V. - really need a community first.

**Discussion about getting Advisory board report**

* Brian: report will be short -
* Paula: maybe next week?
* Brian: do not need any follow up/data
* Paul S.: we had enough data
* Paul S.: a week will be fine.
* Paula: respond in electronic board meeting?
* Brian: only action needed - no actual report
* Alan: do it