December 1st: Presentations

Every group who attended gave a short presentation describing the work they were doing, or plan on doing, with regards to phenotype. These presentations were collected and are available on the wiki at: http://bioontology.org/wiki/index.php/PATO_Meeting#Powerpoint_Slides.

December 2nd: Status Reports and Discussions

Aim of the meeting:
To discuss the key difficulties that are faced and approaches we can take to address these challenges.

Presentation on current status of representation (CJM):
We began with a synopsis describing the current status of the work, presented by Chris Mungall.

1. Principle of Compositionality for phenotypic descriptions
   a. complex descriptions can be made from simpler ones
   b. These can be pre-made or can be made at time of use
2. Pre-composed vs. post-composed
   a. Genus-differentia definitions provide a formal means of composing more complex terms
      i. e.g. plasma membrane of spermatocyte =plasma membrane which is part_of a spermatocyte
3. The building blocks of phenotype descriptions: Entity (E) and Quality (Q)
   a. Entity (bearer) such as spermatocyte, wing
   b. Quality (property, attribute), it is a kind of dependent continuant
   c. Formally, an EQ description defines a Quality which inheres_in a bearer entity
4. The kinds of entities which can be bearers of [biological] qualities
   a. Continuants (3D)
      i. Cell parts (from GO)
      ii. Cells from (Cell ontology)
      iii. Anatomical entities (from FMA, CARO etc)
      iv. Occurants (4D)
      v. Processes such as cell death
      vi. Description of bearer can be composite: e.g. "Cell death in eye", which is composed using an occurant tern and a continuant term
5. Tour of PATO
   a. top level designed using formal ontology principles, see slides for hierarchy
   b. Divisions are also created based on granularity
   c. Qualities are divided into Monadic, Relational, and Relative
      i. Monadic = A quality of a continuant that inheres solely in the bearer and does not require another entity
      ii. Relational qualities are directed towards some other kind of entity (e.g. sensitivity TO red light). Relational = A quality of a continuant that requires another entity apart form its bearer to exist. An example of a relational quality: “Sensitivity of an eye to red light”. The quality inheres_in the eye, with respect to (towards) red light. This can be expressed in Pheno-syntax as: `E=eye Q=sensitivity E2 = red light.`
      iii. Relative qualities have magnitudes (e.g. sensitivity high COMPARED_TO normal sensitivity in wild type). Relational qualities have a implied comparison to the "normal" so temperature-hot, is implied to mean hot as compared to the wild-type. This implied comparison gets tricky for cases where there is no declaration of a "wild type".
   d. Managing "Absence"

Annotation patterns for absence are under discussion, the proposal is to use the lacks relationship from OBO-Rel, and hide the awkwardness of this structure from users with by having the GUI translate it into something more friendly in the user's parlance. e.g. "spermatocyte devoid of asters": `E=CL:spermatocyte`, Inheres in the spermatocyte; `Q=PATO:lacks_part`, the quality/relation of missing some part or parts; `E2 = GO-CC:aster,` the quality is with respect to the type "aster". See also PATO wiki - http://www.bioontology.org/wiki/index.php/PATO:Absent

i.

Discussion of assays:
Maryann Martone and Bill Bug: the observation of a "phenotype" is also tied to the mechanism (or method) by which it is measured and is not easy to disentangle.
Chris M: It might be hard but is definitely worth trying.
Q: Is the purpose of PATO to express experimental result devoid of the experimental procedure/design?
A: Yes
Q: As expression of the result is intimately tied to the experimental procedure then isn't PATO doomed to fail?
A: No, because the experimental procedure will be described using the appropriate portion of the OBI ontology (obi.sourceforge.net) and associated with the phenotypic description in an independent, but correlated, effort.

Discussion of pre and post-composition
Examples of both provided can be found in the presentation. The pre and post-composed phenotype descriptions are equivalent. Pre-composed types can be decomposed to the foundational types. This exercise is much easier to automate if the terms used in the pre-composed annotation have a regular syntax that can be reliably interpreted.

Case Studies:
Chris has carried out some pilot studies to evaluate how well this process of decomposition of pre-composed terms can work.
Case study #1: Defining plant traits with PATO
Results: 252/784 terms provided with manually checked genus-differentia definitions so far
Resulted in some new terms being added to PATO
Case study #2: Animal phenotypes
Done via an automated OBOL run on all 784 terms. 252 were manually validated, the rest were skimmed. He found some systematic problem in decomposing these terms that can be readily corrected.
Case study #3: Bacterial phenotypes
Similar to plant study
26 terms added to PATO
See this page for more information:
http://www.bioontology.org/wiki/index.php/PATO:Pre_vs_Post_Coordinating

Discussion of comparing phenotypes
We want to compare and query both within and across species. CARO is the first project aimed at demonstrating proof of concept of the above.

Emerging requirement we will need terms in OBO for representing aggregates of organisms
Q: Is there a tool that other phenotype annotation creators can use to perform these decompositions of their pre-composed terms for their own phenotype annotations?

A: There is an online OBOL, which may or may not work for all organisms. Contact Chris M with your specific issue.

Measurements
Ontologies provide qualitative partitions on the kinds of entities we fid in nature. We may also want to record quantitative information that comes from "measuring" these qualities. The measurement is not the phenotype. There was lots of discussion on what variables are actually measured [using some units] in order to determine the phenotype.

Phenotype exchange formats
Phenotypes (and to a degree, genotypes): pheno-syntax, pheno-XML
General purpose: OWL (using EQ encoding) [or its equivalent in OBO syntax]
GO annotation files: works with pre-coordinated terms only

OBD-Phenotype
It is a database for associations, including phenotype. It is tuned for inference, reasoning and graph traversal. Currently it contains annotations from OMIM, ZFIN and FlyBase. The current dataset is too small dataset to analyze (2 genes).

Phenote: Phenotype Annotation Tool demo: MG
Mark Gibson gave a demo of Phenote for creating phenotype annotations. There was lots of discussion on what constraints Phenote enforces and does not enforce, mostly centered around 'absence' and 'normal'

Requested features for Phenote *(during discussion):*
1. Ability to retrieve phenotype annotations from a database (as in the dictybase version)
2. Incorporate obo-edit's full featured term-composer
3. Enable reading of OWL ontology files: NOTE: by Jan 1st 2008 OBO-edit will read OWL ontology files or John will quit and go work for industry :-) (further note December 10, 2006, this is now working!)
4. Ability to annotate images by defining region of interest (ROI). ROI need not necessarily square

Linking Animal Models to Human Diseases (using Phenotype annotations) (MW)
ZFIN phenotype annotation is done as Phenotype = entity + quality. Entities come from a lot of different ontologies. Qualities come from PATO. Example of annotating a mutant eye phenotype: `E=optic vesicle, Q=apoptotic, T=during (Prim-15), Tag = abnormal.` Showed the (beta) interface for querying the stored phenotype annotations. This will be released on zfin.org very soon. NCBO project (annotate mutant phenotypes). Participating in annotating a target set of genes (about 200) with detailed phenotype annotations (along with Flybase and OMIM). The OMIM annotations will be done by 3 people (Nicole, Michael and Eric) to allow for comparisons between annotators in the results. Just one mutant (EYA, or eyes absent) results in hundreds of annotation assertions in EQ format.

Talking points and discussion

Everyone agreed that collaborative tool development was highly desirable. The tools were listed on the white board and described by name, purpose, type of technology, and readiness.
PATO + OBI + NLP + Image based annotation

Phenote	Creation of formal phenotype statements	Java	beta
ZFIN Phenote Servlet	web-based creation of formal phenotype statements	Java	beta

DictyBase Phenote Servlet	web-based creation of formal phenotype statements	Perl, and Java	beta
Neuroscholar		Java + mysql full PDF Markup and metadata + RDBMS store	?
NLP tools		Perl + Java + web	alpha
NeuArt		Java + mysql full atlas based segmentation viewer with annotation	
Mosaic	describe phenotypes from pharmacogenomic data and literature	Java	pre-alpha
Smart	Atlas	Java + RDBMS + GIS full image annotation 2D	
Jinx	3-D image annotation	Java	alpha
Biomediator	full identifying articles to curate		
Textpresso	full identifying articles to curate		
PubSearch	full identifying articles to curate		

Now using post-composition: ZFIN, FlyBase
Now using pre-composition (decomposable): MGI, Wormbase, Dictybase, Arabidopsis, Agri, SGD.

Road blocks to (using PATO) for creating and using structured phenotype annotations

Ontology content

1. Gaps in PATO
2. Absence/Gaps in existing entity ontologies:
3. A useful chemical ontology. ChEBI is most likely candidate, but they need to respond faster to requests
4. Need for behaviorial terms

Momentum (legacy phenotype annotations)

1. Tools / documentation or using PATO
2. Tools for managing existing phenotype data
3. Representation / interface for the users for EQ
4. Lot of discussion on explaining the equivalence between pre and post-composed terms and how that would actually work in practice.

TODO CJM: Explicitly define the syntax and semantics of the formalism for creating EQ assertions. Discussion on Phenotype annotation formalism (aka Pheno-syntax and Pheno-XML) See http://www.fruitfly.org/~cjm/obd/formats.html

TODO CJM: Write a paper describing the needs and issues in creating structured phenotype descriptions, describe the resources (ontologies, tools etc) for that activity and describe the formalism (Phenotype Annotation Language?) for actually "writing" the phenotype descriptions.

TODO MG: PaTO, and most of the other ontologies, have trackers for submitting new terms. Phenote should have a module that allows users to generate new terms and submit them to these trackers. This will include the provision of a temporary id for a newly suggested term so that the new term can be used in annotation (and then later the temporary id is replaced with the "official" id when it gets assigned)

Relational qualities and when to use them

Q: How do you represent the difference in phenotype at 1 molar vs. 0.1 molar concentration of a drug

>: The phenotype in this case is represented in relation to a "normal". If this normal is different form the usual interpretation of normal then there should be away to describe that in free-text, which can then be converted to a structured representation later. E.g. the "normal" hematocrit value of people living at high altitude is very abnormal for people living at sea level and this "different normal" has to be declared in a rigorous manner.

>: is this "different normal" part of the phenotype?

>: Lots of discussion of whether this is a phenotype or that the original experiment based on which normal is defined is just a bad experiment.

>: What is the minimal about of information there needs to be declared in order to determine that some phenotype description is "normal" or a "different normal"?

>: point raised by Wormbase that some communities do not consider "normal" as a phenotype. Wormbase does not use "normal" instead they use not-abnormal to describe a phenotype where nothing was found to be wrong for the features/characteristics assayed.

>:Chris did not have time to present his proposal of “compared-to” tag that would address this representational issue.

TODO (CJM & FABIAN): pursue the use of negation in phenotypic description as a mechanism for including an "unaffected" relation.

List of required ontologies

1. PATO (contact George Gkoutos, g.gkoutos@gen.cam.ac.uk), however, use the tracker listed on the PATO wiki to submit requests so that they can be processed correctly
2. Species anatomies (CARO)
3. Environment
4. Unit
5. Chemical (CheBI) including drugs
6. GO-BP, GO-CC, GO-MF
7. Cell
8. OBI (Assays) [Images]
9. Relations
10. Spatial Relationships: (adjacent_to, contains etc), (dorsal, ventral etc)
11. Disease
12. Protein Family
13. Toxicity (adverse reactions)
14. Sequence Ontology (SO)
15. Taxonomies

Follow-up meetings and online discussions to be scheduled

1. Sequence variation
2. Environment
3. Phylogeny/"Homology"
4. Spatial
5. Assays (OBI)