SOFG Workshop on the integration of anatomy ontologies

Held at the MRC Human Genetics Unit, Edinburgh 7- 8 April 2004

Attendees:

Stuart Aitken – University of Edinburgh
Richard Baldock – MRC-HGU
Jonathan Bard - University of Edinburgh
Albert Burger – MRC-HGU / Heriot-Watt University
Duncan Davidson – MRC-HGU
Terry Hayamizu – The Jackson Lab, Maine
Helen Parkinson – EBI
Alan Rector – University of Manchester
Martin Ringwald – The Jackson Lab, Maine
Jeremy Rogers – University of Manchester
Cornelius Rosse – University of Washington
Chris Stoeckert – CBIL, University of Pennsylvania
The workshop was organised by Albert Burger, Jonathan Bard, Carmel Corbett and Duncan Davidson[1]. We thank Nick Hastie, Director of the MRC Human Genetics Unit for help with funding.

The workshop was part of a wider initiative to integrate the different anatomy ontologies for human and mouse, please see http//:www.sofg.org/integration/.

Workshop Goals

The initial goals of the Workshop were as follows:

1. to reach agreement on the top-most layers of a common anatomy ontology for mouse and human, covering both: embryo development and adult;
2. within this ontology, to define anatomical concepts at a level that will result in anatomical definitions that are semantically equivalent in mouse and human;
3. to determine to what depth such a common anatomy ontology can usefully be developed;

The resulting ontology should be sufficiently comprehensive to support applications in the area of microarray research. Microarray research was chosen as the initial target application, because:

- the requirements are simple enough to provide an achievable goal within the workshop, and
- there is an immediate need for such an ontology in this area;

As discussion proceeded on the first day it became clear that we may be able to move forward, but with a simple list of anatomical terms rather than an ontology. Though unstated during the Workshop, the following modified goals capture the actual aims of our discussions during the Workshop

1. To explore whether a list of anatomical terms can be produced that can be used in sample-based functional genomics applications, specifically microarrays.
2. If it possible to make such a list what are the use cases, limitations and build criteria?
3. Production of a draft list.

Discussions, led by Cornelius Rosse, following the Workshop also tackled the definition of anatomical concepts (see initial goal 2)

[1] For correspondence please contact Duncan.Davidson@hgu.mrc.ac.uk

Day 1

Session 1. Presentations and discussion:

Cornelius Rosse – Foundational Model of Anatomy
Alan Rector - Lessons from Galen
Jonathan Bard - The Ontology of Human Developmental Anatomy
Albert Burger – Introduction to the Edinburgh Mouse Atlas Anatomy Ontology
Terry Hayamizu/Martin Ringwald – Developing an Anatomical Dictionary for the Adult Mouse
Helen Parkinson/Chris Stoeckert – Functional Genomics Needs for a Core Anatomy Ontology
Albert Burger – The XSPAN project

Discussion items:

What is the level of annotation that is useful for microarray experiments? Is a core ontology possible?

Conclusion from Session 1:

Sample-based functional genomics experiments are usually limited to what is obtainable by conventional dissection, therefore high level terms are useful, though this should not preclude the use of detailed ontologies where experimenters need these, for example when laser capturing samples.

The SOFG Anatomy Entry List (SAEL)

The workshop suggested creating a short list of anatomical terms that could be mapped and linked to each of the existing anatomical ontologies. We will call this list SAEL. It will be designed to provide a starting point for people without anatomical expertise to find and enter the appropriate anatomical ontologies.

On its own, SAEL will also provide a simple controlled vocabulary for low-resolution descriptions of sample origin. Thus we envisage that the list would be used by researchers and curators wishing to annotate the micro-array assays with the origin of the sample and by researchers wishing to query databases that use these annotations. Those who require more anatomical detail or who need more than a simple list of terms – for example where annotations made at different levels of resolution need to be integrated or where knowledge representation is required for anatomy teaching - would use the appropriate target ontologies.

The creation of SAEL would be a first step towards considering the relationships between different ontologies, for example by focusing attention on the similarities and differences in meaning of the same high level terms in the different ontologies and by comparing the relationships of these common terms to others in the ontologies.

Organisational Principles for SAEL:

Organising principles from the FMA were considered:
Organ, organ part, organ system and *body substance* was added to include urine, blood etc.

To populate these organisational classes the top level (non abstract) nodes in the respective ontologies were considered.

SAEL Content

Top level concepts were built into a draft list based on the organising principles above and the sampling requirements for functional genomics. This is a work in progress and participants agreed that it needs review, extension and mapping to the respective anatomy resources after the workshop. Note that SAEL is a flat list – it has been organised under headings for the purpose of presentation. These headings are variable and do not represent any logical relationships (e.g. is-a or part_of) in the list.

Attributes of Objects in the Target Ontologies:

We considered the set of attributes that might be returned by a web services query when using one of the terms in SAEL list to query a SAEL-mapped ontology resource.

The list developed at the workshop was tested as follows: queries for '*lung*' were tested against the following ontologies and attributes returned were used to refine the initial attribute list:

Table 1. Results of query for '*lung*' from three anatomy ontologies

Attribute List	FMA	Adult Mouse	Developmental mouse
has_id	UWDAID:7195	MA:0000415	EMAP:16728
maps to (mouse/human)	human	mouse	mouse
has_dev_stage	No	TS28	TS15-TS26
Is_tissue	No	no	Nil
is_cell_type	No	no	Nil
is_organ	Yes	yes	Nil
has_superclass	lobular organ	thoracic cavity organ	Nil
has _subclass	left lung, right lung	left lung, right lung	Nil
part	(too many to enumerate)	alveolar system, bronchiole, lung connective tissue, lung epithelium	left lung rudiment, right lung rudiment
part_of	chest, lower respiratory tract	respiratory system	respiratory system
Is_system	No	no	Nil
has_uri	http:// FMA cgi for lung	http:// MGI cgi for lung	Nil
has_definition	lobular organ the parenchyma of which consists of air-filled alveoli which communicate with the tracheobronchial tree. Examples: There are only two instances, right lung and left lung.	nil	Nil
has_authority	Cornelius Rosse	nil	Kauffman (for all terms)
has_history	(last date modified)	(last modified date)	(last modified date)
has_name	Lung	lung	Lung
has_synonym	Nil	nil	Nil

These attributes were refined and defined by Chris Stoeckert into the following list after the workshop:

Table 2. Definition of example attributes returned from target ontologies by queries using terms on the SOFG Anatomy Entry List (SAEL)

Attribute	Allowed Values	Definition
dev_stage	n/a	The developmental stage(s) that is annotated to the anatomical entity. Note that general associations are not sufficient such as those implied from the anatomical resource being developed for the "adult" of the species.
is_tissue	yes, no, nil	The anatomical entity is explicitly categorized as a tissue. For example, the anatomical entity is within a strict is-a hierarchy for tissue
is_cell_type	yes, no, nil	The anatomical entity is explicitly categorized as a type of cell. For example, the anatomical entity is within a strict is-a hierarchy for cell type.
is_organ	yes, no, nil	The anatomical entity is explicitly categorized as an organ. For example, the anatomical entity is within a strict is-a hierarchy for organ
is_system	yes, no, nil	The anatomical entity is explicitly categorized as a system. For example, the anatomical entity is within a strict is-a hierarchy for system.
superclass	n/a	The immediate class(es) above the anatomical entity in an is-a relationship.
subclass	n/a	The immediate class(es) below the anatomical entity in an is-a relationship.
part	n/a	The immediate class(es) below the anatomical entity in a part-of relationship .
part_of	n/a	The immediate class(es) above the anatomical entity in a part-of relationship
uri	n/a	The uniform resource identifier for the anatomical entity in the anatomical resource. A URL is a common type of URI. The URI should provide a pointer directly to information on the anatomical entity in the anatomical resource. This may be in the form of a cgi command
definition	n/a	The definition provided by the targeted resource for the anatomical entity.
authority	n/a	The source of the information provided at the target resource for the anatomical entity. The source may be a person, an organization, or a literature citation
history	n/a	The data provenance of the anatomical entity. Usually, this is the last modification time and date at the anatomical resource for information directly associated with the anatomical entity.
name:	n/a	The name given to the anatomical entity by the anatomical resource.
synonym	n/a	The alternate name(s) given to the anatomical entity by the anatomical resource

Day 2

The work of day 1 was reviewed and use cases for the list were considered. Participants considered their requirements and use cases for a core list and each expressed this as a single sentence:

1. Richard Baldock: Interoperability so that people can annotate data and search other databases

2. Jonathan Bard: A means for the computationally naïve user to get to the correct tissue in a species specific way

3. Cornelius Rosse: A basis for considering similarities and differences for anatomical structures that have a common name across species

4. Stuart Aitken: Principles for representation. Explicit representation of stage/sex/species indexed to anatomy

5. Alan Rector: A community resource, extendable, usable, future mapping possible

6. Martin Ringwald: A limited list of terms that are clearly same in human/mouse to provide easy mapping to species-specific ontologies

7. Duncan Davidson: A mechanism by which a core term in the list can be used to query and join data annotated using that term in different ontologies, e.g. which genes are expressed in the pancreas at a particular developmental stage?

8. Terry Hayamizu: A usable resource that allows easy access to different information sources linked by anatomy

9. Albert Burger: A basis for automated analysis across organisms

10. Chris Stoeckert: Enabling web services

11. Helen Parkinson: A portal to other resources and a usable annotation resource

These statements were then organised into two categories. Those that relate to content and annotation and those relating to software, including end user interfaces and web services:

1. Content/Organisation:

Usable annotation resource
Core terms in list and integrating data annotated using that term, e.g. pancreas which genes are expressed by developmental stage
Limited list terms clearly same human/mouse provide easy mapping to species specific ontologies
Community resource, extendable, usable, future mapping
Principles for representation. Explicit rep stage/sex/species indexed to anatomy
Basis for sim/diff for anatomical structures in species that have a common name

2 Software:

a. **End user interface**

Portal to other resources
Usable resource that allows easy access to diff info sources, linked by anatomy
Naive (comp) user can get to the correct tissue, in a species specific way

b. **Web services/applications**

Enabling web services
Basis for automated analysis across organisms
Interoperable so that people can annotate data and search other databases
Start with the the XSPAN interface
need to add attributes
work out web service specifications

Conclusions regarding SAEL, Day 2

Definition: A freely available selection of cross species anatomical terms relevant to functional genomics. It provides an entry point to anatomical resources via mapping to concepts in those resources. A resource is defined as a controlled vocabulary or ontology consisting of identified anatomical terms which are freely available. It was generated by EMAP, FMA, GALEN, GXD, MGED, EHDA under the auspices of SOFG and initial workshop funding by the MRC-HGU Edinburgh.

Purpose: The SAEL is an entry mechanism for computational access to anatomy resources, to facilitate automated information retrieval.

It is a resource for curators, biologists, informaticians and developers of software supporting functional genomics.

SAEL Software Architecture:

The SAEL list and the mappings to the individual anatomy ontologies will be made publicly accessible through a programmatic interface and a user interface. Figure 1 shows the preliminary software architecture in support of SAEL. The SAEL Anatomy and Mapping database will hold representations of the SAEL list of anatomical entities, parts of the target ontologies and mappings between SAEL entities and the ontologies. The COBrA tool will be used to capture the mappings and make them available to the SAEL database.

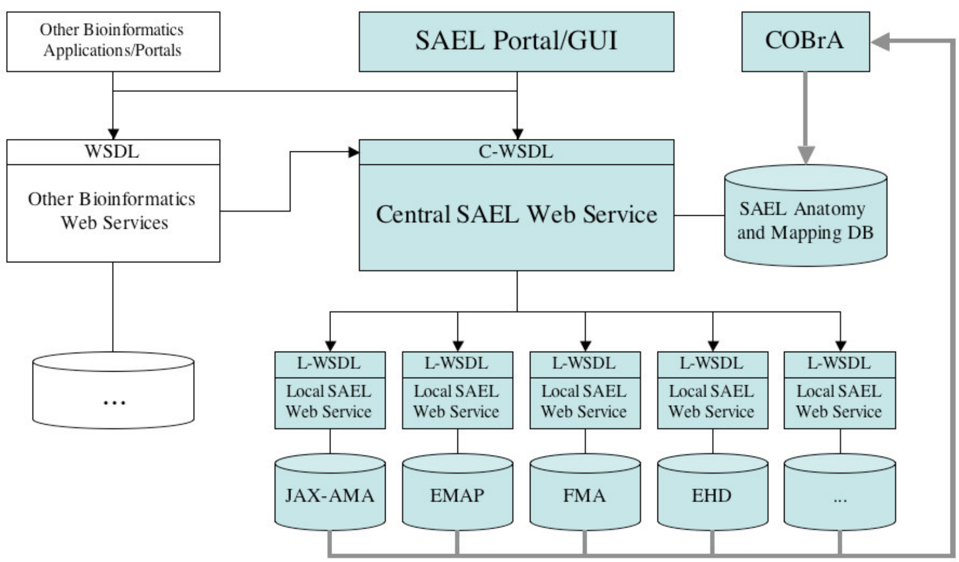

Figure 1: SAEL web service architecture

Each participating anatomy resource will provide a web service interface to its ontology, supporting a simple query mechanism based on the attribute list described in Table 2. Queries, based on these attributes, that involve more than one ontology will be supported by the Central SAEL Web Service. The corresponding WSDL descriptions (C-WSDL and L-WSDL) will define the exact access details to these services. The SAEL Portal provides a graphical user interface for researchers to look up the mappings between the SAEL list of anatomical entities and the target ontologies.

We expect that the SAEL service will become part of a wider network of bioinformatics resources that links anatomy ontologies with other ontologies, databases and computational services.

Most of the computational infrastructure – the Central Web Service, the Anatomy and Mapping DB, the Portal and COBrA – required to deliver SAEL services has already been developed as part of the XSPAN project (www.xspan.org) and will be used for SAEL.

Additional points for discussion:

1. Other applications should be considered - virtual soldier/human, beyond microarrays - ask the community what they need
2. We need to state the principles by which terms are generated
3. Do existing definitions serve our purpose - should we make our own definitions - or use source definitions?
4. How to include cells?
5. How to link back to original ontologies?
6. Identify areas for detailed vertical mapping between ontologies (it was agreed that branchial arches, prostate gland, stomach would provide a good set of entities for this purpose)
7. Process for annotation and assignment of identifiers.
8. Developmental stage resource needs:
stage/day/species mapping e.g. somite/TS/day/mouse CS/day/human
whole embryo and structural mapping between species across dev time e.g. pancreas in human/mouse even when structures/stages are not precisely equivalent
9. Developmental stage should be in the SAEL - do we need entry via stage?
10. What are the structural needs for the list? Multiple lists?
11. Are the list items objects or terms?
12. Should SAEL terms themselves have the attributes listed in Table 1?

Workshop Outcomes

1. The Workshop allowed anatomists, ontology experts and computer scientists to meet and work together. This provides a platform for future work to explore the relationships between existing anatomy ontologies for mouse and human in line with the wider goals of integration (http//:www.sofg.org).
2. For functional genomics applications, a core list of anatomy terms - the SAEL– is possible and a draft is available (appended)
3. A list of attributes for web services have been identified and defined and tested against the FMA, adult and developmental mouse ontologies.
4. Extension, review and mapping the SAEL to external ontologies is ongoing.