

## AN ANATOMICAL ONTOLOGY FOR AMPHIBIANS\*

ANNE M. MAGLIA

*Department of Biological Sciences, University of Missouri-Rolla, 105 Schrenk Hall  
Rolla, MO 65409, USA*

JENNIFER L. LEOPOLD

*Department of Computer Science, University of Missouri-Rolla, 317 Computer Science  
Rolla, MO 65409, USA*

L. ANALÍA PUGENER

*Department of Biological Sciences, University of Missouri-Rolla, 105 Schrenk Hall  
Rolla, MO 65409, USA*

SUSAN GAUCH

*Department of Electrical Engineering & Computer Science, The University of Kansas  
Lawrence, KS 66045, USA*

Herein, we describe our ongoing efforts to develop a robust ontology for amphibian anatomy that accommodates the diversity of anatomical structures present in the group. We discuss the design and implementation of the project, current resolutions to issues we have encountered, and future enhancements to the ontology. We also comment on future efforts to integrate other data sets via this amphibian anatomical ontology.

### 1. Introduction

#### 1.1. *The Need for an Amphibian Anatomical Ontology*

Studies of gene expression, molecular markers, and developmental biology are advancing our knowledge of the morphogenetic and evolutionary processes that lead to disease, physiological responses, adaptation, and phylogenetic diversity. Results from these studies promise both to enhance our quality of life and reveal the complex connection between genotype and phenotype. But to understand fully the results, we must have a detailed understanding of the anatomy of

---

\* This work is partially supported by NSF grant DBI-0445752

organisms. Unfortunately, the lack of terminological standardization for the anatomy of most organisms limits our ability to compare results across taxa, and thus has restricted the applicability of many embryological and gene expression experiments.

The scientific community is well aware of this problem. In the hopes of facilitating the integration of genetic, embryological, and morphological studies, several groups are developing anatomical ontologies for certain model species (e.g., mouse, zebrafish). Further demonstrating the importance of anatomical ontologies was the recent National Center for Biomedical Ontology-sponsored workshop<sup>†</sup> to bring researchers together to discuss issues associated with developing anatomical ontologies.

The need for terminological standardization of anatomy is particularly pressing in amphibian morphological research. Amphibians are commonly used for gene expression and embryological studies, yet the three amphibian orders—Salientia (frogs and toads), Caudata (salamanders and newts), and Gymnophiona (caecilians)—are so morphologically distinct that studies of one order are rarely applied to another. As a consequence, morphological and developmental studies of frogs, salamanders, and caecilians are conducted by disassociated research groups, resulting in three different amphibian anatomical lexicons. Language inconsistencies confuse our understanding of homology, and thus, our ability to use morphology to understand the phylogeny and biodiversity among the orders. In addition, disparate anatomical lexicons limit our abilities to conduct comparative anatomical research, while hindering the integration of morphological, genomic, and embryological data.

There are several challenges to developing an ontology for amphibian anatomy. First, the separate anatomical lexicons must be reconciled. Second, there are over 6,000 species of amphibians for which the anatomical terminology must be resolved. Although much of the terminology is similar across species, among-species variation will lead to a much larger ontology than those developed for a single model species. Third, because of anatomical diversity among amphibian orders, homologies of some structures are unknown; therefore, assigning terminological standards to them may be problematic. These challenges can be overcome by forging a partnership between the amphibian morphological community and the power of information extraction technology.

Herein, we describe our ongoing efforts to develop a robust ontology for amphibian anatomy. We discuss the design and implementation of the project,

---

<sup>†</sup> [http://www.bioontology.org/wiki/index.php/Anatomy\\_Ontology\\_Workshop](http://www.bioontology.org/wiki/index.php/Anatomy_Ontology_Workshop)

resolutions to date for issues that we have encountered, and future enhancements and modifications to the ontology. In addition, we comment on future plans to integrate other data sets via the amphibian anatomical ontology.

### 1.2. *Prior Work in Biological Ontologies*

As stated in [1], “ontologies are becoming popular largely due to what they promise: a shared and common understanding of a domain that can be communicated between people and application systems.” The importance of ontologies has not been lost in the biological community—a research domain that is notorious for its complex form and semantics, and one that will benefit tremendously from data integration and analysis [2]. Perhaps the best known of the biological ontologies is the Gene Ontology (GO)<sup>‡</sup>, which began in the late 1990’s as a collaboration among three model-organism databases (FlyBase<sup>§</sup>, the *Saccharomyces* Genome Database<sup>\*\*</sup>, and the Mouse Genome Database<sup>††</sup>), but has grown to include many other genomic databases.

The biomedical research community has made significant strides in developing medical and clinical ontologies. One of the most extensive projects is the U.S. National Library of Medicine’s Unified Medical Language System (UMLS)<sup>‡‡</sup>, a comprehensive knowledge-representation system that includes data sources and software tools (e.g., the Metathesaurus, the Semantic Network, and the Specialist Lexicon) that facilitate information retrieval, natural language processing, and other vocabulary services for biomedical research data. As an extension to the UMLS, the Digital Anatomist Foundational Model (FMA), an ontology of human anatomical relationships, was developed as part of the Digital Anatomist project [3]. Both GO and UMLS have proved to be extremely valuable for several widely-used applications (e.g., PubMed<sup>§§</sup>, Swiss-Prot<sup>\*\*\*</sup>).

Some bio-ontology projects have begun integrating genomic and anatomical information for model species (e.g., the Zebrafish Information Network (ZFIN)<sup>†††</sup>, The Jackson Laboratory’s Mouse Anatomical Dictionary project<sup>†††</sup>, and the FlyBase list of Anatomy and Development terms<sup>§§§</sup>).

---

<sup>‡</sup> <http://www.geneontology.org>

<sup>§</sup> <http://flybase.bio.indiana.edu>

<sup>\*\*</sup> <http://www.yeastgenome.org>

<sup>††</sup> <http://www.informatics.jax.org>

<sup>‡‡</sup> <http://www.nlm.nih.gov/research/umls>

<sup>§§</sup> <http://www.pubmed.gov>

<sup>\*\*\*</sup> <http://www.ebi.ac.uk/swissprot>

<sup>†††</sup> <http://zfin.org>

<sup>§§§</sup> [http://www.informatics.jax.org/searches/anatdict\\_form.shtml](http://www.informatics.jax.org/searches/anatdict_form.shtml)

Unfortunately, some of these anatomical ontologies have restrictions that prevent their application to other organisms. For example, often there is a narrow set of relations, such as *is-part-of* and *develops-from*—terms that limit the options for describing the inter- and intra-relationships of anatomical parts. This limitation of concepts and properties also limits their use for phylogenetic and comparative anatomical analyses.

## 2. Methodological Considerations and Ontology Construction

The architecture of an ontology typically is sufficiently complex to require a considerable amount of manual effort. As such, the development of an ontology usually is carried out by experts in the knowledge domain. Based on [4], the process of constructing an ontology can be represented by the following steps:

1. Determine the boundaries of the ontology.
2. Consider reusing (parts of) existing ontologies.
3. Enumerate all the concepts to include.
4. Define an appropriate taxonomy to describe concepts, properties and relationships.
5. Define properties of the concepts.
6. Define facets of the concepts such as cardinality, required values, etc.
7. Define instances.
8. Check the consistency of the ontology.

Using the Protégé-OWL editor [4], we developed an ontology in OWL DL for amphibian morphology that was consistent with the recommendations outlined in the Suggested Upper Merged Ontology (SUMO) [5].

In accordance with the list above, we first determined that the boundary for the ontology should include all anatomical *physical, self-connected* objects<sup>\*\*\*\*</sup> for all amphibians (i.e., frogs, toads, salamanders, newts, and caecilians). We evaluated a number of existing sources for reuse, including the SUMO mapping of WordNet [6], the Unified Medical Language System (UMLS), and several species-specific anatomical ontologies (e.g., the Jackson Laboratory's Mouse Anatomical Dictionary, the Anatomical Dictionary<sup>†††</sup>, and the ZFIN

---

<sup>§§§</sup> <http://flybase.bio.indiana.edu/cgi-bin/fbcvq.html?start>

<sup>\*\*\*\*</sup> No *abstract* concepts were defined in the amphibian morphology ontology. Furthermore, each concept in the ontology is considered a *self-connected* object whose parts are all mediately or immediately connected with one another, and no *collection* concepts have been defined at this time. No *process* concepts are currently included in the ontology; however, such an extension may be added in the future to represent functional and physiological knowledge. See [5] for a more detailed discussion of these SUMO top level ontological categories.

<sup>†††</sup> <http://www.dinosauria.com/dml/anatomy.htm>

Anatomical Ontology of the Zebrafish). The SUMO mapping of WordNet provides basic descriptions of terms, and although we were able to identify a few concepts applicable to amphibian morphology, the terminology is too general to be useful for this project. The UMLS is an extensive biomedical ontology containing numerous concepts and relationships. However, our initial attempts to incorporate the UMLS terminology into the amphibian morphological ontology proved to be difficult because: 1) UMLS contains numerous concepts that are not relevant to the amphibian anatomical lexicon and, 2) those concepts that are relevant are not detailed enough for our needs. We also experimented with using an approach similar to the Foundational Model of Anatomy. Interestingly, the top-level organization of this ontology is based on abstract geometric concepts and relationships (e.g., spaces, points, adjacency, direction, etc.). Although such conceptual organization facilitates spatial queries at different levels of complexity, we felt that, for our initial efforts, a top-level organization based on anatomical systems was more consistent with facilitating comparisons among amphibian taxa.

Of the species-specific anatomical ontologies, the ZFIN Zebrafish Anatomical Dictionary is most in line with the goals of our project<sup>\*\*\*</sup>. We adopted relevant concepts, hierarchy, and relationships from ZFIN as an initial framework for the amphibian morphological ontology. Subsequent modifications and enhancements to our knowledge base, including the addition of concepts and properties and the identification of instances, were made by manually mining literature sources [e.g., 7, 8, 9, 10]. Finally, the consistency of the ontology was evaluated through tools provided in the Protégé-OWL ontology builder. End-user evaluations of the usability and usefulness of the ontology are planned (see Section 3.3).

### **3. The Amphibian Anatomical Ontology**

#### **3.1. *The Semantic Network***

The amphibian anatomy semantic network currently consists of 212 semantic concepts and 58 relationships. Each concept is given a textual definition, adopted from ZFIN (where appropriate) or manually mined from the literature. Properties in the ontology are symmetric (e.g., *is-fused-to*), inverse (e.g., *forms*

---

<sup>\*\*\*</sup>It is important to note that at the time of this writing no information was publicly available about the dictionary of embryological anatomy of *Xenopus* (African clawed frog); thus, we could not evaluate the appropriateness of the contents of that knowledge base. When it becomes available, we plan to explore the integration of the dictionary with our amphibian anatomical ontology.

vs. *is-formed-from*), functional and inverse functional (e.g., *is-defined-as* vs. *is-the-definition-of*), or transitive (e.g., *is-part-of*). A partial view of the concept hierarchy and properties for the amphibian anatomical ontology (as displayed in Protégé) is shown in Figure 1.

### 3.2. *Challenges and Current Solutions*

Because of the broad range of organisms and morphologies included in our amphibian anatomical ontology, we faced several challenges in its development. For example, we were required to represent anatomical diversity in a logical and meaningful manner within the terminological and hierarchical framework of the ontology. To do this, we included taxonomic (i.e., Linnaean nomenclature) references as concepts in the ontology. In this way, we were able to designate the range of an instance of a concept as a given taxonomic group. This method also provided us with a way of referencing homologous and partially-homologous structures, while allowing the community to continue to use commonly-accepted terminology (e.g., the orbitosphenoid in salamanders is homologous to the sphenethmoid in frogs).

An additional challenge arose from the need to include developmental stages in the ontology. Most ontologies that include development information are created specifically for that purpose, and often do not include information about adult anatomy (let alone anatomical diversity among groups). To overcome this challenge, we took an approach similar to the one above and included developmental stages as classes. As such, we could designate the range of a concept as an instance of a particular developmental stage.

### 3.3. *Planned Modifications and Enhancements*

As is the case with most biological ontologies (e.g., Gene Ontology, Plant Ontology<sup>§§§§</sup>), the current ontology of amphibian anatomy can be considered a partonomy, because it uses both *is-a* and *part-of* relationships in the hierarchical foundation. Although the use of *part-of* relationships appears to be a logical representation of biological hierarchy, as shown by [11], the inclusion of *part-of* relationships in the hierarchy of a structural ontology can result in inconsistencies and multiple inheritances that are illogical, and can limit the mapping of an ontology into other such systems.

---

§§§§ [http://www.plantontology.org/docs/otherdocs/poc\\_project.html](http://www.plantontology.org/docs/otherdocs/poc_project.html)

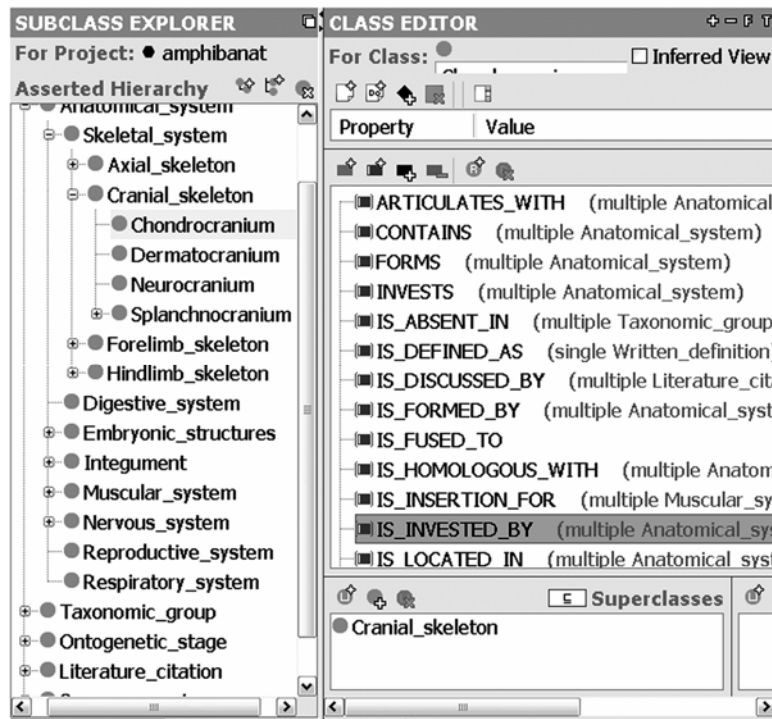


Figure 1. Protégé-OWL editor screen shot of a portion of the class hierarchy and properties associated with the amphibian anatomical ontology.

At the recent NCBO workshop on anatomical ontologies<sup>\*\*\*\*</sup>, it was resolved that a Common Anatomy Reference Ontology (CARO), based on the Foundational Model of Anatomy [3], would be developed to facilitate the integration of anatomical ontologies representing various model organisms. Because the top-down foundational model of CARO is based on sound principles of ontology design, and is explicitly designed to accommodate anatomical diversity, we plan to adopt the CARO model in future implementation of the amphibian anatomical ontology.

In addition, our current practice of including developmental and taxonomic information in the anatomical ontology presents logical inconsistencies. Although the CARO model explicitly excludes developmental and taxonomic information from the ontology, it does include plans to map concepts to other

<sup>\*\*\*\*</sup> [http://www.bioontology.org/wiki/index.php/Anatomy\\_Ontology\\_Workshop](http://www.bioontology.org/wiki/index.php/Anatomy_Ontology_Workshop)

ontologies that do include such data. Therefore, by adopting the CARO model, future implementations of the amphibian anatomical ontology will be logically sound while accommodating biodiversity and developmental information.

#### **3.4. *Software-Based Ontology Enrichment***

Although we have developed the hierarchical class structure for the amphibian anatomy ontology, we have not yet fully instantiated those classes, nor all of the properties associated with the classes. We plan to enrich the amphibian anatomical ontology by using information extraction (IE) technology to mine the amphibian anatomical literature. We currently are developing software to extract elements relevant to anatomy from electronic media, based on previous work by [12]. By combining pattern-based extraction methods with statistical natural language processing algorithms, the software identifies and weights the most important elements. It will be trained using an initial set of values taken from a portion of the current ontology, with focus on extracting highly-weighted, domain-specific terminology (e.g., nouns and noun phrases), important term relationships (e.g., terms related by domain-specific cue words), and inter-concept relationships (most likely indicated by verbs connecting terms specific to two or more concepts).

Our ultimate goal is to provide a software system that can adapt *any* existing ontology automatically by mining concepts from the literature, extend the ontology by adding related concepts to those that are over-represented in the literature, and remove unused concepts. We assume that a concept with many instances in the literature probably is under-represented in the ontology. Through a series of subdivisions of the largest concepts, the ontologies can be supplemented to include new subconcepts with increased specificity, thereby providing a better match to the contents of the representative literature. In addition, machine-learning techniques can be employed on documents that contain few or no instances of concepts in the ontology in order to identify new concepts that might be missing from the ontology. Relationships between these new concepts and the existing concepts can then be inferred using IE techniques.

By experimenting with the size of the initial seed ontology that is adapted, we will be able to evaluate the effect of the amount of information provided and the quality of the automatically generated ontology. As a proof-of-concept, we will seed the learning system with our current amphibian anatomical ontology and allow it to grow. We will also seed it with subsets of the amphibian morphological ontology and evaluate how much the automatically adapted ontology matches the current one, and how well each performs on extrinsic and



intrinsic benchmarks (e.g., similarity to a community-accepted ontology, similarity to concepts represented in a literature subset). We will also investigate which information sources produce the best ontologies. In the best case scenario, an entire ontology could be created from an initial root concept; however, we do not believe this to be probable. The automatic ontology is likely to contain only the simplest inter-concept relationships, e.g., *is-a* or *has-a* or *more-general/more-specific*. The resulting ontology will be evaluated empirically using benchmarks and by evaluation from the user community.

### 3.5. *Community Curation of the Ontology*

Because a knowledge management system can only function satisfactorily if it is integrated into the organization in which it is used [13], it is imperative that the expert user community be highly involved in this project. As discussed in [14], the use of a collaborative ontology builder (COB) environment is vital to properly support the following tasks:

1. Knowledge integration.
2. Concurrence management.
3. Semantic consistency maintenance
4. Privilege management (i.e., to ensure accuracy of the ontology based on a user's expertise, authority, and responsibility for different parts of the ontology)
5. History maintenance.

The collaborative environment for the Gene Ontology is based on a concurrent versions system (CVS), with a request tracking system hosted on SourceForge, and communication between users and curators facilitated via email lists<sup>†††††</sup>. However, as observed in [15], such an environment suffers from the following drawbacks:

1. Absence of a principled mechanism to ensure curator privilege assignments, and organization of the ontology into smaller manageable units.
2. Risk of inconsistency from unintended couplings and over-writing.
3. Lack of support for restricting editing to only part of the ontology (i.e., a curator has to download the entire ontology before editing, and then submit the entire ontology after editing).
4. Expensive history maintenance (i.e., even a minor edit to the ontology causes the entire file to be replicated in its entirety).

---

<sup>†††††</sup> <http://www.geneontology.org/GO.contents.curator.guides.shtml>

5. The inability to grant different levels of privileges to different types of users, subsequently limiting community participation.

For the amphibian anatomical ontology we are currently using the same CVS protocol employed by the Gene Ontology. However, we are in the process of evaluating alternatives such as the use of COB Editor<sup>\*\*\*\*</sup>, a recently developed COB software system that overcomes many of the aforementioned problems, and has been used successfully for the Animal Trait Ontology (ATO)<sup>\*\*\*\*\*</sup>.

To facilitate the use of our amphibian anatomical ontology, we are developing a Web site ([www.amphibanat.org](http://www.amphibanat.org)) that includes documentation on all aspects of the project such as a monthly newsletter and answers to frequently asked questions, discussion boards, links to contacts and mailing lists, and downloadable tools for using the ontology including a Java-based API and a user interface for searching, browsing, and navigating the ontology.

#### 4. Knowledge Integration via the Anatomical Ontology

A long-term goal of this project is to integrate the amphibian anatomical ontology knowledge base with systematic, biodiversity, embryological and genomic resources. Interoperability with other media resources has been considered in the design and implementation of our knowledge base. We currently are developing a Java-based API with several functions, including: searching for a particular term, iterating through all terms related to a specific term, finding citations for literature associated with the use of a term, etc. By standardizing amphibian anatomical terminology and providing a platform- and implementation-independent API to access the ontology from existing applications, we are developing a means to facilitate integration of phylogenetic, anatomical, embryological, and gene expression data.

We plan to demonstrate the usefulness of this API by integrating it with the querying facilities of *MorphologyNet*<sup>\*\*\*\*\*</sup>, a digital library of 3D visualizations of anatomy developed by Leopold and Maglia [16]. The *MorphologyNet* query interface (which is currently being developed, and will be available in November 2006) allows searching for morphological images by any combination of: taxonomic reference, anatomical reference, developmental stage, accession number, and contributor name. Integrating the amphibian

---

\*\*\*\* <http://sourceforge.net/projects/cob>

\*\*\*\*\* <http://www.animalgenome.org/bioinfo/projects/ATO>

\*\*\*\*\* <http://www.morphologynet.org>

anatomical ontology with the *MorpholgyNet* database will provide users with the option of searching for anatomical structures using the controlled vocabulary, retrieving images using synonym matching, or accessing images that are hierarchically-related to the search term in the ontology.

## 5. Summary

The amphibian anatomical ontology provides a terminological and hierarchical framework for amphibian anatomy. By standardizing the lexicon used for diverse biological studies related to anatomy (e.g., gene expression, embryology, systematics), we hope to facilitate the integration of anatomical data representing all orders of amphibians, thus enhancing our knowledge of amphibian biology and diversity. Our ontology will provide a powerful tool that will facilitate cross database querying and foster consistent use of vocabularies in the annotation of amphibian morphology. Thus, it could allow a morphologist determine the preferred name for a given anatomical structure, an evolutionary biologist to find similar morphological structures of phylogenetic significance present across different species, or an embryologist study/compare how gene expression guides the development of embryos in different taxonomic groups. By using good practices of ontology development, we hope to integrate the amphibian anatomical ontology with many different types of Internet-distributed databases (including anatomical ontologies representing other organisms), thus helping to realize fully a Semantic Web within the biological domain.

## References

1. Davies, J., Fensel, D., and Van Harmelen, F. (eds.). *Towards the Semantic Web: Ontology-driven Knowledge Management*. John Wiley & Sons., West Sussex, England (2004)
2. Soldatova, L., and King, R. Are the current ontologies in biology good ontologies? *Nature Biotech.* 23:1095–1098 (2005)
3. Rosse, C., Shapiro, L. G., Brinkley, J. F. *The Digital Anatomist Foundational Model: principles for defining and structuring its concept domain*. *Proc AMIA Symp* 820–4 (1998)
4. Knublauch, H. Fergerson, R. W., Noy, N. F., Musen, M. A. *The Protégé OWL Plugin: An Open Development Environment for Semantic Web Applications*. Third International Semantic Web conference – ISWC 2004, Hiroshima, Japan (2004)
5. Pease, A., Niles, I., and Li, J. *The Suggested Upper Merged Ontology: A Large Ontology for the Semantic Web and its Applications*. In

- Working Notes of the AAAI-2002 Workshop on Ontologies and the Semantic Web, Edmonton, Canada (2002)
6. Fellbaum, C. WordNet: An Electronic Lexical Database (Language, Speech, and Communication). MIT Press: Boston. 423 pp. (1998)
  7. Duellman, W. E. and Trueb, L. Biology of Amphibians. Johns Hopkins University Press, Baltimore (1994)
  8. Maglia, A. M., Pugener, L. A., and Trueb, L. Comparative Development of Anurans: Using Phylogeny to Understand Ontogeny. *Amer. Zool.*, 41:538–551 (2001)
  9. Reilly, S. M. and Altig, R. Cranial ontogeny in *Siren intermedia* (Caudata: Sirenidae): Paedomorphic, metamorphic, and novel patterns of heterochrony. *Copeia* 1996: 29–41 (1996)
  10. Maglia, A. M. and Pugener, L. A. Skeletal development and adult osteology of *Bombina orientalis* (Anura: Bombinatoridae). *Herpetologica* 54: 344–363 (1998)
  11. Smith, B., and Ross, C. The Role of Foundational Relations in the Alignment of Biomedical Ontologies. 11th World Congress on Medical Informatics (MEDINFO) (2004)
  12. Gauch, S., and Chandramouli, A. Semi-Automatic Update of Existing Taxonomy using Text Mining. 5th International Conference on Ecological Informatics (ISEI5) (2006)
  13. Antoniou, G., and Van Harmelen, F. A Semantic Web Primer. Boston, MIT Press. (2004)
  14. Bao, J., Hu, Z., Caragea, D., Reecy, J., and Honavar, V. A Tool for Collaborative Construction of Large Biological Ontologies. Fourth International Workshop on Biological Data Management – DEXA 2006, Krakow, Poland (2006)
  15. Leopold, J., Maglia, A., and Hoefl, T. Interactive Anatomy Online: The *MorphologyNet* Digital Library of Anatomy. *IEEE Potentials* 24(2):39–41 (2005)