

Developing Frameworks and Tools for Animal Trait Ontology (ATO)

Zhi-liang Hu ¹, Jie Bao ^{2,3}, Max F. Rothschild ¹, Vasant Honavar ^{2,3}, and James M. Reecy ¹

1. Department of Animal Science and Center for Integrated Animal Genomics, 2255 Kildee Hall, 2. Department of Computer Science, and 3. Center for Computational Intelligence, Learning, and Discovery, 226 Atanasoff Hall, Iowa State University, Ames, IA 50011

Abstract

There is an urgent need for precise definition of animal trait terms (phenotypes) to capture the biologically relevant distinctions at the desired level of detail in an unambiguous fashion. Ontologies that identify and define the entities and relationships in specific domains of interest offer a powerful approach for annotating biological data in a form that allows users and software tools to retrieve, inter-relate, and extract biological knowledge. Our previous work on PigQTLdb introduced simple ontologies in the form of controlled vocabularies to describe pig phenotypes or traits. We now turn our attention to the development of software tools to facilitate the creation, editing, curation, and management of animal trait ontologies. We have developed an animal trait ontology editor (ATO editor) that overcomes the most of these problems. We have also developed database structures to manage trait ontologies for cattle, pigs, chickens, sheep and other livestock species. To date, we have entered over 300 pig traits into the ontology database. In order for such ontologies to be broadly useful to the livestock and animal genomics communities, they need to capture the knowledge and expertise of multiple experts and research groups. Hence, we propose the creation of consortia representing the relevant livestock genome communities to develop, maintain, and update trait ontologies. ATO and associated software tools for collaborative creation, editing, curation, and management of ontologies provide the infrastructure necessary for engaging the animal genomics and livestock communities in the process of creating comprehensive genomic resources for annotating, integrating, and analyzing phenotype and genomic data.

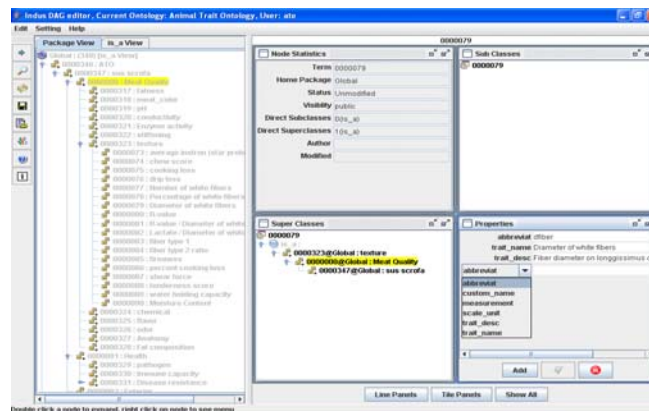


Figure 2. A snapshot view of ATO Editor loaded with current pig trait ontology.

Background and Rationale

The need for biological ontologies has risen in recent years in large part due to the rapid development of biological databases and the need for sharing information between disparate research groups. This has led to the development of ontologies such as Gene Ontology, Rice Ontology, Plant Phenotype and Trait Ontology.

Animal Trait Ontologies (ATO): Precise definition of animal trait terms (phenotypes) that capture the biologically relevant distinctions at the desired level of detail in an unambiguous fashion is a prerequisite for sharing phenotype information, performing genome annotation, and cross-species comparisons. Previous work by Hu et al (2005) on the PigQTLdb introduced a simple ontology in the form of a controlled vocabulary to describe pig phenotypes/ traits, and to link them to QTL information (Figure 1). However, real-world applications call for more expressive ontologies. Animal Trait Ontologies (ATO) are designed to address this need.

Developing Large Ontologies: In order for an ontology to be broadly useful to a certain community, it needs to draw on the collective expertise of multiple individuals and groups. Typically, a large ontology has to be built and curated by a community. Unfortunately, existing ontology editing tools, such as the DAG-Edit (Day-Richter, 2004), and OBO-Edit (Mungall, 2005) currently do not offer support for such collaborative development. Consequently, there is an urgent need for collaborative ontology development environments. Additionally, development of large ontologies calls for approaches to ontology management that can efficiently process large ontologies. Hence, there is a need for software tools for ontology storage, editing, browsing, visualization, and sharing that scale with increasingly large, collaboratively developed ontologies.

To address this need, we have been developing software tools for collaborative development and use of ontologies in general, and ATO in particular. Our approach draws on recent advances in modular ontologies (Bao and Honavar, 2004, 2005) to facilitate the creation, editing, and management of animal trait ontologies within a distributed curator model by the animal genetics/genomics community at large.

Animal Trait Ontology Editor

The ATO editor is written in Java for its universal portability across different hardware and operating systems platforms. It utilizes a backend relational database, e.g., Postgres for storing ATO.

The current version of the ATO editor supports different ATO curators/editors to work on an ATO remotely over internet. It allows multiple users to work on the same ontology (through appropriate locking mechanisms) without inadvertently overwriting the work of others.

The current version of the ATO editor can be used to introduce terms, define relationships between terms, as well as to edit the term properties.



Figure 1

A web page showing part of the simple pig trait ontology with the "Meat Quality" traits as an example. The pig traits and trait categories are retrieved from the PigQTLdb.

Summary and Discussion

Ontologies that identify and define the entities and relationships in specific domains of interest offer a powerful approach for annotating biological data in a form that allows users and software tools to retrieve, inter-relate, and extract biological knowledge from such data. With the rapid proliferation of biological ontologies, there is an urgent need for software tools for collaborative development of ontologies as well as, sharing, and integration of independently developed ontology fragments.

While there are several existing tools for developing ontologies, few support collaborative development of complex ontologies. ATO Editor and associated software tools that we are developing aim to offer capabilities that complement those that are provided by existing ontology tools.

Work in progress is aimed at developing the theoretical foundations as well as practical tools for collaborative development of modular ontologies including in particular, support for identification and resolution of inconsistencies among independently developed modules, support for partial (as opposed to all-or-nothing) reuse ontologies (Bao, Caragea, and Honavar, 2005).

Availability

The current version of ATO editor is available for download at <http://www.animalgenome.org/bioinfo/software/ATO>.

Jie Bao (baojie@cs.iastate.edu) may be contacted for technical questions and the most recent release.

References

Bao, J. and Honavar, V. (2004) Ontology Language Extensions to Support Collaborative Ontology Building In: International Semantic Web Conference (Poster Track), 2004.

Bao, J., Caragea, D., and Honavar, V. (2005). Toward Collaborative Environments for Ontology Development and Reuse. Technical Report CS-TR-06-01. Department of Computer Science, Iowa State University.

Day-Richter, John (2004). DAG-Edit: A controlled vocabulary editor <http://www.godatabase.org/dev/java/dagedit/docs/>

Hu, Zhi-Liang, Svetlana Dracheva, Wonhee Jang, Donna Maglott, John Bastiaansen, Max F. Rothschild and James M. Reecy (2005). A QTL resource and companion tool for pigs: PigQTLdb. Mammalian Genome. Volume 16(10):792-800.

Mungall, Chris (2005). Integrated ontologies for biological annotation: The National Center for Biomedical Ontologies. The First International Biocurator Meeting. Pacific Grove, CA, December 8-11, 2005