Developing the Animals in Context Ontology

Suzanne L. Santamaria[1,*] Maureen Fallon[1], Julie M. Green[1],
Stefan Schulz[2,3] and Jeffrey R. Wilcke[1]

[1]Veterinary Medical Informatics Laboratory at Virginia Tech, Blacksburg, Virginia, US

[2] Institute for Medical Informatics, Statistics and Documentation, Medical University, Graz, Austria

[3] Institute of Medical Biometry and Medical Informatics, University Medical Center, Freiburg, Germany

ABSTRACT

Animals are classified by any number of various characteristics including Linnaean rank, physiologic features, purpose and place. The Animals in Context Ontology (ACO) was developed by editing a subset of the Systematized Nomenclature of Medicine Clinical Terms (SNOMED-CT©) to follow the Open Biological and Biomedical Ontologies (OBO) Foundry Principles. It includes animals classified by Linnaean ranking as well as practical uses that are of interest to science, medicine and agriculture. ACO was built within the ontological framework of the Basic Formal Ontology (BFO) and the Relations Ontology (RO) and uses classes from other ontologies including the Phenotypic Quality Ontology (PATO), the National Center for Biotechnology Information (NCBI) Taxonomy, the Environment Ontology (EnvO), and the Gene Ontology (GO). ACO includes 216 unique classes in an OWL format.

Availability: http://vtsl.vetmed.vt.edu/aco/Ontology/aco.zip.

1 INTRODUCTION

Animal classification needs vary by user and purpose. The Linnaean hierarchy is the international standard for animal nomenclature. However, it has notable shortcomings for many applications. It lacks the common identifying characteristics for sex, production role such as meat or milk for human food, age, diet and living environment necessary to describe many animals that are subjects in scientific research, patients in veterinary clinics, and animals in production units such as farms. At the same time, it is too specific for some common animal classes which do not correspond with a single Linnaean taxonomic equivalent and which could refer to more than one taxonomic group. For example, in the United States, “cattle” could refer to *Bos taurus* or *Bison-Bos taurus* crosses. Elsewhere, “cattle” might refer to non-*Bos taurus* species. However, all cattle throughout the world are members of *Bovinae* that have some use.

An ontology that represents the way animals in practical uses are described, from “Cattle” to “Beef heifer raised in confinement,” is needed for various applications. These applications include vaccine and drug labels, gene set mapping, species preservation, and veterinary medical records. There are two ontologies listed on the Open Biological and Biomedical Ontologies (OBO) Foundry[1] (Smith *et al.*, 2007) website that carry animal classifications - the National Center for Biotechnology Information (NCBI) Taxonomy[2] and the National Cancer Institute (NCI) Thesaurus[3] - but both are inadequate for representing animals in practical use. The reason is that both lack a structure and mechanism for representing animal classes with non-Linnaean defining characteristics such as sex and production role. In addition, these ontologies contain some imprecise classes unsuitable for use in animal production and husbandry.

The Animals in Context Ontology (ACO)[4] was developed to fill this need for identifying animals in extra-Linnaean ways. In this paper we describe the development of ACO, the resulting ontology, and future work.

2 METHOD

2.1 Source for ACO Development

A subset of animal classes was previously developed using the organism hierarchy of the Veterinary Terminology Services Laboratory (VTSL)[5] extension of the Systematized Nomenclature of Medicine Clinical Terms (SNOMED-CT©)[6], a large, international medical terminology. The subset was originally populated from animals needed by the United States Food and Drug Administration’s Center for Veterinary Medicine (FDA CVM) and the United States Department of Agriculture - Animal and Plant Health Inspection Services - Veterinary Services (USDA APHIS VS) and was stored in VTSL’s database. Current known users of a portion of the animal classes include two branches of USDA APHIS VS for animal disease surveillance and the Virginia Department of Health for rabies reporting.

The current organism hierarchy in SNOMED-CT core does not contain any non-taxonomic defining relationships; however, the organism classes in the VTSL extension were defined using additional characteristics including sex, age group, production role, and taxonomic rank. The subset had a stated poly-hierarchical structure so animals could be classified by taxonomy (e.g., *Bovinae*) and common role grouping (e.g., *Food animal*), but lacked text definitions.

* Address correspondence to: slsantamaria@vt.edu.

[1] http://obofoundry.org

[2] http://www.ncbi.nlm.nih.gov/taxonomy

[3] http://www.cancer.gov/cancertopics/cancerlibrary/terminologyresources

[4] http://vtsl.vetmed.vt.edu/aco

[5] http://vtsl.vetmed.vt.edu

[6] http://www.ihtsdo.org/snomed-ct

2.2 Importing External Ontology Classes

ACO imports the upper level ontologies Basic Formal Ontology (BFO)[7] and BioTopLite[8] as well as a bridge between them. BioTopLite was chosen because it is a top-domain ontology for biomedicine, and because it includes numerous object properties (relations), some of them mapped to the Relations from the OBO Relations Ontology,[9] together with numerous constraints such as domain/range restrictions. See Figure 1 for placements of ACO classes in BioTopLite.

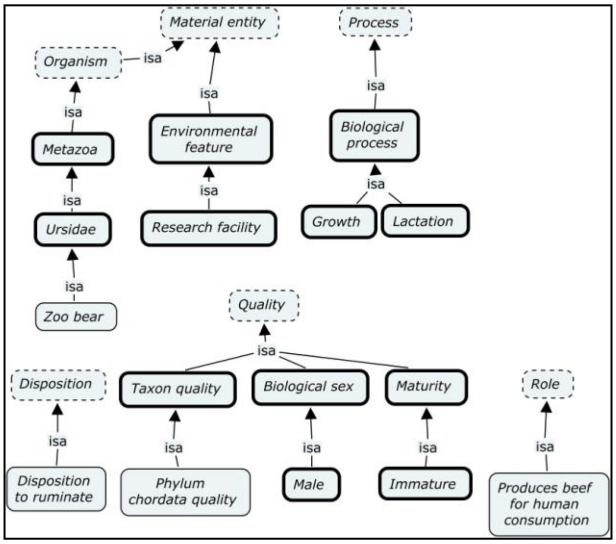

Fig. 1. Placement of ACO classes in an upper level ontology. Dashed boxes contain BioTopLite classes. Bold boxes contain classes imported from other ontologies and the thin lined boxes are classes in ACO. Image composed with CmapTools.

The taxdemo ontology[10] (Schulz, *et al.*, 2008) had been proposed as an example of how to build organism ontologies that refer to biological taxa. The basic idea had been to represent taxa as qualities, which can inhere in populations, in a single organism, as well as in organism parts. For the purpose of ACO, taxon quality classes were created and related to the ACO animal classes as proposed in taxdemo. The Ontology Lookup Service[11] and the Bioportal[12] were used to locate appropriate external ontologies for class re-use. We used OntoFox[13] to create files to import classes from external ontologies such as the Phenotypic Quality Ontology (PATO)[14] and the Gene Ontology (GO)[15] into ACO. See Table 1 for a summary of the classes in ACO. Collaboration with OBO members was necessary in some instances to pick the appropriate classes from external ontologies and to learn the OntoFox program.

Ontology	No.	Use in ACO	Example
ACO	216	Practical animal classes	*Female adult horse*
ACO	58	Taxon qualities	*Subfamily bovinae quality*
ACO	12	Roles	*Produces milk for human food*
ACO	1	Disposition	*Disposition to ruminate*
Imported Full Ontologies			
Basic Formal Ontology (BFO)	39	Upper level hierarchy	*Process*
BioTopLite	49	Upper level hierarchy and relations	**participates in**
BioTopLite-BFO bridge	39	Connects BFO and BioTopLite	
From External Ontologies			
NCBI Taxonomy	40	Organism taxonomy (Linnaean)	*Bovinae*
Gene Ontology (GO)	7	Biological processes	*Lactation*
Environment Ontology (EnvO)	10	Environment sites	*Aquatic habitat*
Phenotypic Quality Ontology (PATO)	13	Phenotypic qualities	*Female*

Table 1. Summary of classes created for or imported into ACO.

2.3 ACO-Specific Classes

We developed ACO following the OBO Foundry principles. ACO specific classes and related definitions were entered manually into the Protégé 4.1 ontology editor.[16] All classes unique to ACO are given URIs. The original SNOMED-CT class identifier has been retained as a cross reference using the *alternativeId* annotation property. The preferred names are in plain English and are mainly singular nouns with the exception of cattle (explained in Discussion). The preferred description in SNOMED-CT was used as the preferred name in ACO. The scope includes classes of those animals that are put to practical use. Text definitions in the genus-species differentia format were added with the *hasDefinition* annotation property for each ACO specific class. ACO uses a common shared syntax of OWL-DL. Description logic definitions were added for most of the classes

[7] http://www.ifomis.org/bfo
[8] http://purl.org/biotop
[9] http://obofoundry.org/ro
[10] http://purl.org/biotop/taxdemo/dev
[11] http://www.ebi.ac.uk/ontology-lookup/
[12] http://bioportal.bioontology.org
[13] http://ontofox.hegroup.org
[14] http://www.obofoundry.org/cgi-bin/detail.cgi?id=quality
[15] http://www.geneontology.org
[16] http://protege.stanford.edu

in ACO. Appropriate, commonly used synonyms were added using the *hasExactSynonym* annotation property.

3 RESULTS

ACO contains 510 classes, 286 of which are unique to ACO. See Table 1 for a listing of classes by ontology. We imported classes from external ontologies to avoid duplication of existing content. Table 2 shows an example of a class from ACO and its associated axioms.

Preferred name:	*Castrated male cattle for beef production*	
Synonym:	"Beef steer"	
Text definition:	"Beef cattle which are male and castrated"	
Formal definition:		
equivalentTo	*Cattle for beef production* and	
	bearer of some *Castrated male quality*	
Inherited:		
subClassOf	**bearer of** some	*Subfamily bovinae quality*
subClassOf	**bearer of** some	*Disposition to ruminate*

Table 2. Example of ACO class. For brevity, many of the inherited anonymous classes are excluded from this table.

3.1 ACO Top Structure

The animal classes in ACO denote descendents of *Metazoa* from the NCBI Taxonomy. *Metazoa* corresponds with *Kingdom Animalia* and is the class that encompasses all potential animal classes in ACO. *Metazoa* imports as a direct child of *Organism* in BioTopLite. Originally, we imported all classes from the needed distal taxonomic class in NCBI Taxonomy (superclass of an ACO-specific class subclass) to *Metazoa* in NCBI Taxonomy. This included many intermediate classes and a mixture of Linnaean and cladistic classes that proved unwieldy. We then reimported in OntoFox attaching the most distal taxonomic class needed from NCBI directly as a child of *Metazoa*, eliminating the intermediates. The following rules were used. Refer to Figure 2 for an illustration of the examples.

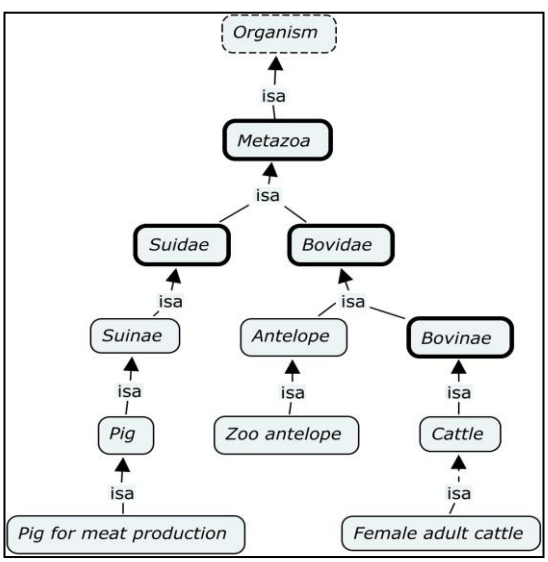

Fig. 2. Choice of upper level classes from other ontologies. Organism is from BioTopLite. Bolded boxes are classes imported from NCBI Taxonomy and thin lined boxes are classes in ACO.

(1) ACO animal class is a subclass of the most distal NCBI Taxonomy class that includes all members of animal class. As shown in Figure 2, the ACO class *Antelope* is a subclass of *Bovidae* because that is the most distal NCBI Taxonomy class that denotes all members considered antelope by mammalogists and taxonomists of authority such as Mammal Species of the World[17] (Four-horned antelope in *Bovinae*, grey rhebok of *Peleinae*, etc.). *Bovidae* is imported as a direct child of *Metazoa*.

(2) If two needed NCBI Taxonomy classes are part of a natural hierarchy in the NCBI Taxonomy, they are imported retaining the hierarchy. In Figure 2, the ACO class *Cattle* is a subclass of *Bovinae* from the NCBI Taxonomy because that is the most distal NCBI Taxonomy class that includes all members considered cattle by taxonomists throughout the world (*Bos taurus*, *Bison*, etc.). Because *Bovidae* is needed for a different ACO class (*Antelope* as described above), *Bovinae* is imported as a child of *Bovidae*, which is imported as a direct child of *Metazoa*.

(3) If NCBI Taxonomy does not include the most distal taxonomic ancestor known that includes all members taxonomists consider to be a member of the ACO animal class, then we created the needed distal class in ACO and imported the most distal NCBI Taxonomy class that subsumes this needed distal class as a direct child of *Metazoa*. As shown in Figure 2, *Suinae* is the most distal known taxonomic class that includes all species considered to be pigs. *Suinae* does not exist in NCBI Taxonomy so the most distal taxonomic ancestor (*Suidae*) was imported from NCBI Taxonomy and a class *Suinae* was created in ACO as a child of *Suidae*, which is imported as a direct child of *Metazoa*. The ACO class *Pig* is a child of *Suinae*.

BioTopLite Relation	Value Example	Ontology Source
participates in	*Lactation*	GO
bearer of	*Female*	PATO
bearer of	*Produces milk for human food*	ACO
bearer of	*Subfamily caprinae quality*	ACO

Table 3. Defining relationships for ACO class: *Lactating ewe for milk production.* The relationship in the first row distinguishes this class from its parent. Other relationships are inherited. Some inherited relationships are not shown to due space limitations.

[17] http://www.vertebrates.si.edu/msw/mswcfapp/msw/index.cfm

3.2 ACO Defining Classes

Many classes used in the formal definitions of ACO-specific classes were imported from external ontologies or were created in ACO but identified as probable additions to external ontologies. An appropriate source for the animal roles in external ontologies was not found so they remain in ACO. Taxon quality classes were created in ACO. See Table 3 for an example of a formal definition of an ACO class.

3.3 Added Classes to Infer Structure

ACO has a single **isa** asserted inheritance structure, expressed by subclass relations in OWL-DL. Animal classification and organization which do not obey a biological taxonomy is desired for grouping by common classes such as *Food animal*. This provides useful classification hierarchies for the users of the ontology. ACO includes the following classes that infer members based on formal definitions: *Animal for breeding, Animal in fiber production, Exhibition animal, Aquarium animal, Zoo animal, Food animal, Laboratory animal,* and *Wildlife*. ACO classifies with both the Fact++ and HermiT reasoners in Protégé 4.1.

3.4 General Class Axioms

ACO includes some general class axioms to further define the animals in roles. See Table 4 below for an example. It shows how animals bearing a certain role can be considered equivalent to animals that participate in certain processes.

bearer of some *Produces fiber*
EquivalentTo **participates in** some
(*Production* and (**has outcome** some *Fiber product*))

Table 4. Shown is the general class axiom for an animal who is bearer of the role class *Produces fiber. Fiber product* will be requested as an addition to the Environment Ontology (EnvO).

Discussion of conceptual issues including upper level ontology placement, external ontology classes re-use, and text definition creation took place over the period of one year. The actual manual creation of the ontology in Protégé took one month. The linkage to a well-constrained upper-level ontology like BioTopLite was of considerable heuristic value, due to iterative validation steps using DL classifiers for consistency checking.

3.6 Availability

ACO is open and available online.

4 DISCUSSION

ACO was developed as an ontology of animal classes within the OBO Foundry framework to maximize resources, data integration, reusability and interoperability. This proved both challenging and rewarding. Tools to assist with ontology development were available without charge, including the OBO Foundry website, Protégé, OntoFox, the Ontology Lookup Service and the NCBI Bioportal. Collaboration with OBO members was very effective. Multiple people offered their opinions on questions posed to the listservs. Responses were provided within 24 hours and in some cases almost immediately. We found that ontologies listed on the OBO website are at varying stages of development, compliance with OBO principles, and curation level. We encountered several classes that need work and identified several necessary additions to the ontologies.

An example of a class that could be improved is *Pasture* in the Environment Ontology (EnvO)[18]. Its parent is *Grassland* and its text definition is “Grassland used for grazing of ungulate livestock as part of a farm or ranch.” Pasture can consist of grasses or legumes and are not always part of a managed farm or ranch. There are pastures in certain parts of the world that are open, public areas. Therefore we suggest the EnvO curators should either: 1. edit this class name to “grassland ranch pasture” and leave the text definition as is, or 2. move this class from *Grassland* to *Terrestrial habitat* and edit the text definition to: “Terrestrial habitat used for grazing, foraging or browsing by animals.”

Ontology	**No. of Additions Needed**	**Example**
GO	3	*Rumination*
EnvO	15	*Feedlot*
PATO	5	*Castrated male*
NCBI Taxonomy	3	*Suinae*

Table 5. Summary of additional classes needed in OBO ontologies for ACO. Some of these additions have been submitted through the appropriate tracker.

We discovered numerous classes for additions to existing ontologies so other ontologists can draw similar content from the same external ontology. See Table 5 for a summary of these additions and the ACO site[19] for a list of all the needed additions. We believe it is more desirable for the taxon quality classes to be included as formal definitions of the NCBI Taxonomy classes rather than included directly in ACO. Since this is a significant and debatable request, we did not include these in the additions list to NCBI Taxonomy. Another option is to interpret the NCBI Taxonomy classes as taxon qualities themselves rather than organisms. However, we did not choose this because NCBI’s documentation explicitly states that the taxonomy refers to organisms and because including ACO classes as subclasses of NCBI Taxonomy classes enables reasoning and subsumption with other ontologies using the same taxonomic resource.

[18] http://environmentontology.org
[19] http://code.google.com/p/animalnamesontology/downloads/list

Animals bearing roles were given additional general class axioms relating their production role to an outcome of a specific product. The EnvO class *Food product* includes food for human or animal consumption in its text definition, therefore additional EnvO classes specific to products for human consumption (e.g., *Egg product for human consumption*) are needed to fulfill these axioms. Classes for *Wool product* and *Fiber product* also need to be added to EnvO.

We reviewed each class in ACO to check for compliance to the OBO Foundry singular noun principle. Three categories of non-compliance were identified: 1) plural noun where singular form exists ("eggs"); 2) single noun and plural noun are the same ("deer"); 3) plural noun where singular form does not exist ("cattle" and "broodstock"). All classes with the plural "eggs" in the preferred name were changed to the singular "egg." All classes with "broodstock" in the preferred name were edited to include "breeding" instead and broodstock terms were retained as synonyms. "Deer" were left as is as there is no exclusive singular form. The issue of a singular form of cattle was presented to the OBO list serve. Multiple suggestions were given and "head of cattle" seemed the most logical and accurate of the suggestions for a singular count noun. Although this is technically correct, it is not how people engaged in animal husbandry or veterinary medicine talk and would violate the OBO Foundry principle that preferred terms should be in ordinary English as extended by technical terms already established in the relevant discipline. Therefore, we chose to keep "cattle" in our singular classes.

We built ACO manually because one researcher needed experience in ontology building and using Protégé. An effective automated transfer method between the SNOMED-CT subset and the ontology in OWL would have decreased some development time. This was investigated superficially and problems with SNOMED-CT's description logic and the extension classes' use of non-sanctioned relationships in SNOMED-CT were encountered.

In addition to the improved format and increased interoperability, this development work resulted in improvements in the original subset. We identified and corrected simple and logical errors and omissions in the original subset. Examples include retiring a class from the original subset (*Animal in context*) because it could be not be instantiated, adding a missing definition of the quality *neonatal* to *Newborn sheep for milk production*, and removing a redundant parent of *Cattle* for *Cattle on pasture for human food*, leaving *Cattle for human food* as its only parent. The original subset classes had a taxon rank attribute and value ("genus" level). This was deprecated and we plan on using the structure of the taxdemo ontology to communicate taxon quality and rank instead. We added a role of *Pre-production* to better define replacement animals and increase the number of fully defined classes in the subset.

Identifying animal information at various levels from breed and utility to Linnaean classification is needed for various electronic record applications from science to medicine. ACO integrates within the Linnaean classification system but provides common non-Linnaean groupings such as *Duck* and extends them to practical animal classes such as *Duck laying egg for human food.* Animal data recorded with ACO classes integrate and interoperate with other OBO-based scientific and medical ontologies, allowing for reasoning and classification of data captured from multiple sources and with multiple ontologies. This should encourage biomedical researchers to access animal science and veterinary research as well as production and health records for comparative analysis purposes including discovering new associations between phenotypic and gene traits. Because it is expensive to build and maintain biomedical ontologies, collaborating and using common resources may help to decrease costs associated with ontology development and maintenance. Collaborators from multiple OBO ontologies including the Vaccine Ontology have expressed interest in using ACO. ACO's format is more accessible to the broader scientific community while still maintaining its SNOMED-CT subset origin.

5 FUTURE WORK

Community use of ACO will result in the addition of classes and other changes needed to improve the ontology. Future work of the ACO development process includes: 1) analyzing representation of animal taxa specific production classes like broilers and fryers in chickens and starters, growers, and finishers in pigs; 2) considering formal definition with Linnaean and other classes for useful grouping classes such as *Antelope*, *Shellfish, Cold blooded animal*, *Duck* and *Nonhuman primate*; and 3) investigating the need to divide ACO into multiple ontologies. Formal evaluation for inclusion into the OBO Foundry, assignment of an OBO Foundry namespace, documentation development and tracker creation are future goals.

ACKNOWLEDGEMENTS

We acknowledge and appreciate funding from the US Food and Drug Administration's Center for Veterinary Medicine (FDA CVM) and the US Department of Agriculture – Animal and Plant Health Inspection Service – Veterinary Services (USDA APHIS VS) in support of this work.

REFERENCES

Schulz S., Stenzhorn H., Boeker M. The ontology of biological taxa. Bioinformatics. 2008 Jul 1;24(13):i313-21.

Smith, B. *et al*. (2007). The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat Biotechnol, 25(11), 1251–1255.