Development of an Ontology of Animals in Context Within the OBO Foundry Framework from a SNOMED-CT Extension and Subset

Suzanne Lamar Santamaria

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

**Master of Science
In
Biomedical and Veterinary Sciences**

**Jeff R. Wilcke
Julie M. Green
Loren P. Rees
Kurt L. Zimmerman**

**April 25, 2012
Blacksburg, VA**

Keywords: veterinary informatics, SNOMED-CT, OBO, biomedical ontologies

Development of an Ontology of Animals in Context Within the OBO Foundry Framework from a SNOMED-CT Extension and Subset

Suzanne Lamar Santamaria

ABSTRACT

Animal classification needs vary by use and application. The Linnaean taxonomy is an important animal classification scheme but does not portray key animal identifying information like sex, age group, physiologic stage, living environment and role in production systems such as farms. Ontologies are created and used for defining, organizing and classifying information in a domain to enable learning and sharing of information. This work develops an ontology of animal classes that form the basis for communication of animal identifying information among animal managers, medical professionals caring for animals and biomedical researchers involved in disciplines as diverse as wildlife ecology and dairy science.

The Animals in Context Ontology (ACO) was created from an extension and subset of the Systematized Nomenclature of Medicine – Clinical Terms (SNOMED-CT). The principles of the Open Biological and Biomedical Ontologies (OBO) Foundry were followed and freely available tools were used.

ACO includes normal development and physiologic animal classes as well animal classes where humans have assigned the animal's role. ACO is interoperable with and includes classes from other OBO Foundry ontologies such as the Gene Ontology (GO). Meeting many of the OBO Foundry principles was straightforward but difficulties were encountered with missing and problematic content in some of the OBO ontologies. Additions and corrections were submitted to four ontologies. Some information in ACO could not be represented formally because of inconsistency in husbandry practices. ACO classes are of interest to science, medicine and agriculture, and can connect information between animal and human systems to enable knowledge discovery.

DEDICATED TO

I dedicate this thesis to my family and friends. This work would not have been possible without your support and encouragement - thank you.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

INTRODUCTION

Ontologies are formal sets of entities and relations between entities used to organize and classify information in effort to increase knowledge about a domain. There has been a recent dramatic increase in numbers and types of biological and biomedical ontologies. Many of these ontologies overlap in content, yet there are still gaps in coverage. Many of these ontologies were built in isolation and without plans for integration with other ontologies. In general, many lack formal definitions and are not based on formal logic. Although there are fewer standard terminologies and ontologies in the animal arena, these problems still exist.

An open source community called the Open Biological and Biomedical Ontologies (OBO) Foundry is attempting to apply the scientific method to ontology development. The OBO Foundry provides principles for collaborative, coordinated development and integration of biomedical ontologies[1,2]. The OBO Foundry was developed based on the successes of the Gene Ontology (GO)[3]. GO was created when researchers realized they needed to present and record data in a way that was compatible between species of animals.

Identifying animal information at various levels of granularity from production role, sex and age group to Linnaean classification is needed for various electronic applications in science and medicine. These applications include vaccine and drug labels, gene set mapping, species preservation, and veterinary medical records. Experts in many subject areas use the Linnaean classification scheme as a guide to create authoritative taxonomy hierarchies for their area of expertise. In field usage, sometimes a more granular Linnaean term is known (species or subspecies) and other times a less granular Linnaean term is more appropriate (family). The veterinarian at a dog show may know the Linnaean subspecies *Canis lupus familiaris* but the dog owners and managers will refer to them as dogs. The layperson reporting a dead animal to a

1

health authority may only be able to identify it as a bird, which corresponds to the Linnaean

Class *Aves*. In addition, animal descriptions outside the strict Linnaean hierarchy including sex,

production class and age group are used to identify animals. For some uses the Linnaean

classification is too specific. There are several common groupings of animals that do not

correspond with any one Linnaean class. For example, ducks are members of some but not all

genera in Family *Anatidae* because this family also contains swans. The current Linnaean

system does not include a single appropriate taxon equivalent to "all animals that are called

ducks." Biomedical research and veterinary medicine uses animals of the class "Duck" as the

animal source of vaccines developed from duck tissues or vaccines that are approved for use

only in ducks.

An ontology containing extra-Linnaean classification of animals is needed, but the OBO

Foundry does not currently offer one. The two ontologies on the OBO Foundry website that

contain animal classes – the NCBI Taxonomy and the NCI Thesaurus - are currently inadequate

for representing animals for various reasons: 1) lack of a mechanism for including non-Linnaean

defining characteristics such as production role, age group and living environment; 2)

inaccuracies stemming from lack of management by animal husbandry subject matter experts;

and 3) uncertain integration of some classes from an alternative classification system

(Cladistics). The Systematized Nomenclature of Medicine – Clinical Terms (SNOMED-CT)

contains animal terminology and a mechanism to define additional characteristics but it is not an

OBO Foundry ontology and does not integrate with OBO Foundry ontologies.

This research project involved editing a pre-existing SNOMED-CT extension and subset

of extra-Linnaean animal classes to create an ontology that conforms to as many of the OBO

Foundry principles as possible. This work has value for both the animal and human communities

because it highlights the successes and challenges of this process and because the resulting ontology will integrate with other OBO-based ontologies. Animal data gathered using an ontology developed according to OBO principles should facilitate data analysis and reuse [4]. This should facilitate biomedical researchers' access to animal science and veterinary research as well as production and health records for comparative analysis purposes, leading to discovery of new associations between phenotypic and gene traits[5]. Finally, this work is important because it is expensive to build and maintain biomedical ontologies and collaborating and combining resources may help to decrease costs associated with ontology development and maintenance.

BACKGROUND

Why Animals in Context?

Two classification systems exist for animal nomenclature. The Linnaean taxonomy system classifies organisms by phenotypical biological similarity into 7 or more divisions. Cladistics is a newer method of classifying organisms based on common ancestry. Although both systems are important, they are not practical for direct use in some animal records because many people are unfamiliar with the names and neither system contains utility-based classes of animals such as cattle and laying hen. At present, electronic records for animal health identify subjects of care using various unspecified combination of Linnaean classification, breed, age, sex, utility, living environment and phenotypic group. A well rendered ontology will facilitate proper and consistent identification of the animal allowing for connecting clinical, regulatory and research data.

Why SNOMED-CT?

SNOMED-CT is a large, robust, international medical terminology managed by the International Health Terminology Standards Development Organization (IHTSDO) [6]. It contains

the following hierarchies: clinical finding, procedure, observable entity, body structure,

organism, substance, pharmaceutical/biological product, specimen, special concept, linkage

concept, physical force, event, environment or geographic location, social context, situation with

explicit context, staging and scales, physical object, qualifier value, and record artifact.

SNOMED-CT has a well-developed organism hierarchy that has been in use for many years in

many applications. It contains thousands of Linnaean classification concepts[i] (e.g., *Aves*) with

common name synonyms (e.g. Bird) as well as a few common non-Linnaean grouping concepts

(e.g., *Pheasant*). Animal concepts in the organisms hierarchy of SNOMED are used by

numerous organizations including the United States Department of Agriculture – Animal and

Plant Health Inspection Services – Veterinary Services (USDA-APHIS-VS), the Virginia

Department of Health (VDH) and Veterinary Medical Databases (VMDB).

Why OBO and BFO?

The OBO Foundry, led by Dr. Barry Smith, has established itself as the leader in modern

biomedical ontology development and improvement. The OBO Foundry was built on the success

of the Gene Ontology (GO). The OBO Foundry maintains a website with over 100 ontologies,

resources, and links for collaborative discussion groups[7]. The principles of the Foundry are

enumerated on their website (see Table 1 for the list). Basic Formal Ontology (BFO)[8] was

developed as the upper level scientific ontology for OBO. BFO ensures that all ontologic

classes[ii] are placed in the correct kind of hierarchy. BFO and OBO both highly value

---

[i] The IHTSDO defines concept as 'A clinical idea to which a unique ConceptId has been assigned[13].' In this paper SNOMED entities are referred to as concepts.

[ii] Dr. Barry Smith, head of the OBO Foundry, defines class as 'A maximum collection of particulars referred to by a general term[31].' This includes classes which correspond to what is general in reality (universals such as *animal*) as well as defined classes which do not correspond with a universal (*animal names beginning with C*). In this paper OBO ontology entities are referred to as classes.

interoperability of ontologies due to the following benefits: lowered cost, combination of resources, increased chance for data reuse, and the ability to test ontologies.

Integrating SNOMED with OBO

Although SNOMED features many characteristics of an ontology, SNOMED has been criticized for not being a true ontology and it does not currently conform to all the OBO Foundry principles[9]. Clinical data recorded with SNOMED remains inaccessible to projects that utilize OBO ontologies such as GO. SNOMED does not interoperate with other OBO ontologies because it lacks a common syntax, common upper level hierarchy, common relations besides **is a** and it is not created or defined using classes from other ontologies. Although it is developed collaboratively, it is not truly open because a license is required for use. Some SNOMED concepts do not correspond to entities that exist in reality (e.g. *Finding of foot (finding)).* SNOMED does contain unique identifiers and has mechanisms for formal definitions, text definitions, and identification of successive versions. It is extremely well documented and has a large and increasing number of users worldwide. SNOMED would benefit from increased interoperability and an even higher number of users if it were aligned with OBO ontologies[10]. However, given SNOMED's size and complexity, this is a tremendous undertaking and therefore modular alignment of certain portions of hierarchies is a more realistic goal. In 2008 Hogan mapped 14/19 of the top hierarchies in SNOMED to BFO classes with significant reorganization but determined that subtypes of these top hierarchical concepts would need to be evaluated individually for placement in BFO[11]. Alignment with BFO is a significant step toward integrating with OBO ontologies.

METHODS

Common Animal Names Subset as Starting Source

In 2006, approximately 200 new animal classification concepts were created in the Veterinary Terminology Service Laboratory (VTSL) extension [12] of SNOMED-CT. These classes were outside the scope of Linnaean taxonomy and included terms commonly used in animal health systems, animal medical records and animal science research. From these and from existing concepts that have common language equivalents (*Aves* = bird), the Common Animal Names (CAN) subset of SNOMED-CT was developed. The subset was based on animals appearing on drug labels approved by the US FDA Center for Veterinary Medicine (CVM). Coverage included the common classes chickens, ducks, pheasant, quail, turkeys, sheep, goats, horses, zebras, donkeys, ponies, cattle, elk, moose, antelope, deer, foxes, mink, pigs, dogs, cats, nonhuman primates, bears, honeybees, fish, lobster, shrimp, rabbit, mice and the grouping categories of food animal, animal for breeding, animal in fiber production, zoo animal, wildlife, lab animal, and exhibit animal. The extension concepts included animals in human-assigned roles (*Beef cattle),* animals in typical physiologic states (*Mare*), animals by living environment (*Aquarium animal*) and non-Linnaean animal names used to physically identify animals (*Duck*). Each concept was given a hierarchical **is a** parent of its most distal Linnaean taxonomic concept containing all members of that group and any additional grouping parents. No Linnaean concepts were allowed as descendents of common animal concepts during hierarchy construction." See Figure 1 for an example of the hierarchical arrangement of subset concepts. An extension hierarchical designator of "organism in context" was used in the fully specified names of these CAN concepts. "Organism in context" was an extension of the Organism hierarchy. The extension concepts seamlessly integrate with the international core release of SNOMED-CT in

the VTSL relational database. Added additional descriptions (synonyms) were added to the concepts. The CAN concepts were defined using relation attributes that were not approved in SNOMED as the organisms hierarchy in SNOMED currently does not have any sanctioned relations besides simple **is a** subtypes.   Relations were created for the specific purpose of this extension of SNOMED. These included combining the extension relations **has physiologic state** and **has life circumstance** with physiologic concept values (e.g., *Female (finding)*) and values from the animal life circumstance hierarchy (e.g*., Produces milk for human consumption (animal life circumstance)*), respectively.

Initial Preparation of Source Subset

In the initial stages of this research project, the CAN concepts were edited as follows: 1) the concept fully specified name format of "genus, differentia" (e.g*. Cattle, beef*) was changed to a format consistent with SNOMED and OBO guidelines of plain English without punctuation (e.g. *Cattle for beef production*) and all content edited to conform; 2) The extension SNOMED hierarchical identifier of "organism in context" was changed to the core hierarchical identifier "organism" by direct replacement; 3) Additional appropriate synonyms (descriptions) were added to the animal concepts (e.g. *Beef cattle*).

New text definitions were created for the animal concepts that were not Linnaean concepts. The text definitions were reviewed by 4 VTSL lab members and 2 external subject matter experts. SNOMED has adopted a mechanism whereby text definitions are used to clarify concept meanings[13]. SNOMED-CT uses a text definition like an additional description and these animal text definitions will likely be added to the VTSL Extension in the near future.

Creating the Ontology from the Source Subset

The final phase of the project involved transforming the CAN subset into an ontology using the OBO Foundry principles as a guide. The OBO Foundry principles were identified on the OBO website. Each principle was reviewed and changes to the subset needed to comply with the principles were made. See Table 1 for the list of principles. Compliance with the principles is presented in the results section. Use of singular nouns, use of common English, and existence in reality (instantiability) were investigated by reviewing each concept in the CAN subset. After consulting with the OBO Foundry community through a listserv, "The Animals in Context Ontology" was chosen as the name of the ontology. It is abbreviated as ACO in the remainder of this paper.

I reviewed the documentation for every ontology on the OBO Foundry website to determine if the ACO was orthogonal to other OBO ontologies (Principle 1.6). The site lists the title, domain, prefix, and date of last change for each ontology. Additional information such as contact name, current activity, website address, relevant organism and documentation is provided on a separate page. The quality of documentation varied between ontologies. Any ontology that appeared to have similar content based on the initial assessment was investigated further by reading the documentation and searching the content for animal classes that are in ACO.

Placement of the concepts into the hierarchy of the upper level scientific ontology Basic Formal Ontology (BFO) was illustrated on a concept map during the evaluation phase (see Figure 2). Relations, or attributes, in the CAN subset were examined and suitable relations from the OBO Relations Ontology (RO)[14] were identified. Three lab terminologists reviewed the relations.

Stanford University's ontology editor Protégé 4.1 for Windows 64 bit[15] was used to create ACO in the OWL-DL format. The whole of BFO was imported as the upper level scientific ontology. Classes from OBO ontologies that were previously identified as replacement classes were prepared for import into ACO using the web service OntoFox[16,17] which creates import files in the OWL format by extracting class information from OBO ontologies (see Figure 4). These files were then imported to ACO using the "direct import" function in Protégé. Animal classes were created in Protégé as subclasses of the appropriate imported classes from the NCBI Taxonomy[18] and the animal classes were formally defined using a combination of the imported classes and ACO-specific classes.

I utilized the online resources Ontology Lookup Service[19] and the NCBO Bioportal[20] to locate classes in other OBO ontologies to use for defining ACO classes to meet the OBO Foundry principle 2.7. Each candidate class was investigated by reviewing the text definition and placement in the original ontology to ensure it represented the meaning needed to define ACO classes. Examples and numerous possible substitutions were communicated in a document and three terminologists discussed controversial or unclear classes. When necessary I collaborated with ontologists through the OBO listservs and individual emails. The SNOMED concept identifier and OBO class identifier were initially stored in a document. See Table 4 for examples of the substitutions made. When OBO ontologies lacked a necessary class, the class and text definition were temporarily created in ACO and added to a list for eventual addition to the OBO ontology. See Table 7 for a list of these additions to four OBO ontologies.

Initially RO was imported into the ontology using the direct import function and the relations were used to build the ACO. BioTopLite[21], a top-domain ontology for biomedicine,

was subsequently imported into ACO and most of the relations used are the same as the relations used in the RO.

The animal classes in ACO are descendents of *Metazoa* (Kingdom Animalia) from the NCBI Taxonomy ontology. *Metazoa* imports as a subclass of *Organism* in BioTopLite. Originally, I imported all classes from the needed distal taxonomic class to *Metazoa,* 133 classes of Linnaean and Cladistic terms. I was advised that it was not necessary to import all intermediate classes so I then reimported in OntoFox attaching the most distal taxonomic class needed from NCBI directly as a subclass of *Metazoa*, eliminating 93 unneeded intermediates.

The animal classes were built under the imported NCBI Taxonomy classes by the following rules. Refer to Figure 6 for examples of rules 1 and 2 in Protégé.

1. ACO animal class is a subclass of the most distal NCBI Taxonomy class that includes all potential members of animal class. Example: The ACO class *Zoo bear* is a subclass of *Ursidae* because that is the most distal NCBI Taxonomy class including all members that would be classified as bears by animal managers and medical professionals throughout the world (Sun bear in *Helarctos*, Black bear (*Ursus americanus*), etc.). *Ursidae* is imported as a direct subclass of *Metazoa*.

2. If NCBI Taxonomy classes are identified as necessary for particular ACO classes and they are part of a natural hierarchy in the NCBI Taxonomy, they are imported and the hierarchy is left intact. Example: *Ursus americanus* from the NCBI Taxonomy is needed because that is the most distal Taxonomy class including all members of the ACO class *Zoo black bear* (Louisiana black bear, Florida black

bear, etc). Its ancestor *Ursidae* is needed for a different ACO class (*Zoo bear* as described above), so *Ursus americanus* is imported as a subclass of *Ursidae,* which is imported as a direct subclass of *Metazoa*.

3. If NCBI Taxonomy does not include the most distal taxonomic ancestor known which includes all members that animal managers and medical professionals might consider a member of the ACO animal class, then I created the needed distal taxonomic class in ACO and imported the most distal NCBI Taxonomy class that subsumes this needed distal class as a direct subclass of *Metazoa.* Example: *Suinae* is the most distal known taxonomic class that includes all animals 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 subclass of *Suidae*, which is imported as a direct subclass of *Metazoa*. The ACO class *Pig* is a subclass of *Suinae*. Suinae is a valid Linnaean class according to the taxonomic reference Mammal Species of the World[22].

I adhered to similar rules when importing classes from the Phenotypic Quality Ontology (PATO)[23], GO, and the Environment Ontology (ENVO)[24] as described above for importing classes from the NCBI Taxonomy.

To create a single **is a** inheritance structure in ACO, the editor chose one superclass that was biologically rather than production oriented. For example, in the CAN subset, *Chicken for breeding* had two parents: *Gallus gallus* and *Breeding animal*. In ACO, *Chicken for breeding* was given a single parent of *Gallus gallus* and it infers a parent of *Breeding animal* after reasoning.

The original CAN subset concepts were defined by taxon rank ("genus" level). For ACO I used the structure of the taxdemo ontology[25] to describe taxon name and rank[26]. Because the NCBI Taxonomy does not include taxon qualities, I created a taxon quality hierarchy of needed content under the BioTopLite class *taxon quality* in ACO. Each taxon quality class is a subclass of its rank quality and its Linnaean ancestor quality. For example, the S*ubfamily Bovinae quality* is a subclass of *subfamily quality* and is a subclass of *Family Bovidae quality*. Each animal class was defined as bearer of an appropriate taxon quality class. For example, the class *Gallus gallus* was given a formal definition of **bearer of** *species Gallus gallus quality*. For the non-Linnaean animal classes the taxon quality of the most distal Linnaean class including all members animal managers and medical professionals throughout the world might consider to belong to that non-Linnaean class was used rather than creating a common animal taxon quality. For example, *Cattle* inherits a taxon quality of *Subfamily Bovinae quality* from its superclass *Bovinae*, its most distal Linnaean ranking.

I created a googlecode site[27] for the ontology and use a subversion tool (TortoiseSVN)[28] to synchronize files with the googlecode site.

RESULTS

I created an ontology of 216 animal classes that conforms to the principles of the OBO Foundry with few exceptions. Certain remaining principles can only be met after the ontology is accepted into the OBO Foundry and published on the website. The total number of classes including imports is 500. See Table 2 for a list of ontologies from which classes were imported for use in ACO along with specific examples. Table 3 lists the number of imported classes by ontology. Figure 3 demonstrates a view of a portion of the ACO hierarchy in Protégé and the text

definition, synonym, SNOMED identifier, and formal definitions of one animal class. A list of

the principles and a description of ACO's compliance with each principle follows.

Principle 1.1: The ontology must be open and available to be used by all without any constraint other than (a) its origin must be acknowledged and (b) it is not to be altered and subsequently redistributed under the original name or with the same identifiers.

This principle is met because ACO is currently open and available on a googlecode

website and on the VTSL website. If ACO is accepted into the OBO Foundry it will be available

openly on the OBO Foundry website as well. Use of the original SNOMED subset and extension

requires a license. ACO contains the SNOMED concept identifiers that link it back to the

original CAN subset.

Principle 1.2: The ontology is in or can be expressed in a common shared syntax.

Since ACO was produced in the common shared syntax of the Web Ontology Language

(OWL), this principle is met.

Principle 1.3: The ontology possesses a unique identifier space within the OBO Foundry.

This principle is partially met. We are in negotiations to obtain the prefix "ACO" from

OBO. Each class in ACO was automatically given a unique Uniform Resource Identifier (URI)

by Protégé in the form http://purl.obolibrary.org/obo/aco_nnnnnnnnn (e.g.,

http://purl.obolibrary.org/obo/aco.owl#Animal_for_breeding).

Principle 1.4: The ontology provider has procedures for identifying distinct successive versions.

This principle is met. Versions of the ontology are tracked numerically using the

googlecode site. In addition, the original CAN subset of SNOMED uses the method for handling

versions through the SNOMED history table. This will be used to track history and versions of

the CAN subset.

Principle 1.5: The ontology has a clearly specified and clearly delineated content.

I believe this principle is met. The classes in ACO are limited to common groupings of animals and the accompanying scope statement describes the content of ACO:

> The Animals in Context Ontology (ACO) includes classes that are created when
> animals are managed by humans. These classes relate directly to terms understood
> by individuals involved in animal husbandry, production and veterinary care. A
> few of these animal classes are imported from the recognized Linnaean hierarchy
> and are actually common synonyms for the scientific names (*Aves* is equivalent to
> Bird). Most of the ontology extends the Linnaean hierarchy by referring to
> various life stages (e.g., calf), production classes (e.g., broiler chicken),
> circumstances of husbandry (e.g., zoo animal), physiologic and age-based
> classifications (e.g., heifer) and common non-Linnaean groupings (e.g., duck).

Principle 1.6: The ontology must be orthogonal to other ontologies already lodged within OBO.

I believe this principal is met based on a review of the ontology content information provided on the OBO Foundry website.

Principle 1.7: The ontologies include textual definitions for all term AND *Principle 2.4: Textual definitions will use the genus-species form: An A =def. a B which Cs, where B is the parent of the defined term A and C is the defining characteristic of those Bs which are also As.*

These principles are met because text definitions were created for all ACO-specific classes in the genus-species form during this project.

Principle 1.8: The ontology uses relations which are unambiguously defined following the pattern of definitions laid down in the OBO Relation Ontology.

This principle is met. All of the relations used in ACO are in the RO except **has locus**, which is in the top domain ontology BioTopLite. I imported all the relations in BioTopLite but only use the **participates in** relation to define an organism and the process it participates in and **bearer of** to define an organism and the quality, role, or disposition it bears. These relations also are found in the RO. No additional relations were required.

Principle 1.9: The ontology is well documented.

This principle is met. A three page description of the ontology was prepared based on the GO documentation. This description of ACO is available on the ACO portal of the VTSL website which is open to the public. The CAN subset had been presented to a client in 2007 and presented at the 2011 AVMA Convention in St. Louis, MO. A paper was written about the CAN subset for the convention.

Principle 1.10: The ontology has a plurality of independent users.

This principle is partially met. The Vaccine Ontology (VO) is already using classes from ACO to indicate the animal for which a vaccine is approved. There is a potential for additional users because a few members of the OBO listserv indicated ACO would be useful in their work. The concepts in the CAN subset are being used by multiple groups including two branches of USDA for animal disease surveillance and reporting and the VDH for rabies reporting. The FDACVM is evaluating the CAN subset for possible use in the Structured Product Label (SPL).

Principle 1.11: The ontology will be developed collaboratively with other OBO Foundry members.

This principle is met. ACO was created through collaboration with OBO members via the OBO listservs, web conferences, emails and in-person discussions at the 2011 International Conference on Biomedical Ontology (ICBO) in Buffalo, NY. Assistance in using OntoFox was

provided by a collaborator from the University of Michigan, Dr. Asiyah Yu Lin. Dr. Stefan

Schulz reviewed ACO in January, 2012, and offered multiple suggestions for improvement,

including the use of the top domain ontology BioTopLite and general class axioms. The CAN

subset was originally developed in house at VTSL with input from FDA CVM, USDA, Stefan

Schulz and other members of the SNOMED organisms committee.

Principle 2.1: All candidate Foundry ontologies will appoint a person responsible for liaison

with the OBO Foundry coordinating editors.

This principle is met. Dr. Jeff Wilcke and I will be responsible for liaison with the OBO

Foundry coordinating editors and are listed as contacts on the ACO website. When ACO is

presented formally to OBO, Dr. Wilcke and I will provide contact information and agree to

become official liaisons.

Principle 2.2: All candidate Foundry ontologies will provide a tracker for additions and

corrections, and a help desk for inquiries from users.

This principle is unmet. I prepared a website for ACO where interested parties can email

us with corrections and additions. Once ACO is presented to OBO a tracker will be obtained on

SourceForge where the other OBO ontology trackers reside.

Principle 2.3 Textual definitions will, by degrees, be complemented with equivalent formal

definitions.

This principle is met. Of the 216 animal classes in ACO, 170 are fully defined. The

designation of fully defined means that defined classes have one or more necessary and sufficient

conditions. ACO classes are formally defined by bearing certain qualities, dispositions and roles

and by participating in certain processes that distinguish them from other classes. See Table 6 for

an example of a fully defined class. ACO includes some general class axioms to further define

the animals in roles. See Table 5 for an example. It shows how animals bearing a certain role can be considered equivalent to animals that participate in certain processes. At the inception of the project there were already formal definitions in the CAN subset for the majority of the concepts, however I increased the number of fully defined classes by adding classes that were not present in the subset.

Principle 2.5: Single is_a inheritance: ontologies will distinguish a backbone ('asserted') is_a hierarchy subject to the principle of single inheritance.

This principle is met as the ACO contains only single **is a** inheritance. Multiple **is a** relations are inferred after classification with a reasoner. The CAN subset of SNOMED has multiple **is a** relations asserted. The FACT++ reasoner was used to classify ACO in this project.

Principle 2.6: Instantiability: Terms in an ontology should correspond to instances in reality.

This principle is met. One concept from the CAN subset was identified as not corresponding to an instance in reality – *Animal in context* – so it was retired from the CAN subset and not included in ACO.

Principle 2.7: Terms should be created and defined using terms and relations drawn from other Foundry ontologies.

This principle is met. The CAN subset concepts did not use terms from ontologies besides SNOMED in their formal definitions. See Tables 2 and 3 for information on external ontologies used in ACO. See Table 4 for an example of substitutions made.

Principle 2.8: Each Foundry ontology should be subject to evaluation.

This principle is partially met. ACO was reviewed by three VTSL members. Two external ontologists evaluated ACO in the spring of 2012. The next step is to submit the ontology for a formal evaluation by OBO Foundry members.

Principle 2.9: Use of Basic Formal Ontology (BFO).

This principle is met. The classes in ACO are descendants of the upper level scientific hierarchy Basic Formal Ontology (BFO). In addition, BioTopLite, a top domain ontology for biomedicine, was imported into ACO along with a BFO-BioTopLite bridge. The CAN subset was not aligned with BFO at the inception of the project.

Principle 2.10: Preferred terms should consist entirely of singular nouns.

This principle is met with two kinds of exceptions: names where singular and plural are the same word (deer) and plural nouns that do not have a singular form (cattle).

Principle 2.11: Preferred terms should be nouns and noun phrases belonging to ordinary English as extended by technical terms already established in the relevant discipline.

This principle was met. Each class in ACO was reviewed by two terminologists and it was determined that all preferred names are noun or noun phrases belong to ordinary English. Some discipline-specific technical terms such as "starter pig" are included in ontology.

During this process, I identified and corrected numerous errors and omissions in the original CAN subset and extension. See Table 8 for some examples of corrected errors.

DISCUSSION

The OBO Foundry principles provide solid guidelines to develop a structurally and logically sound biomedical ontology that integrates with other similarly designed ontologies. In addition, tools are freely available to assist with ontology development, including the ontology editor Protégé, the class extractor OntoFox, the ontology browser Ontology Lookup Service, and OBO Foundry listservs. Each ontology provides documentation and support. Although the structure put forth by the OBO Foundry is sound, I encountered some challenges when implementing the principles during development of the Animals in Context Ontology (ACO).

The issues encountered while attempting to meet each principle are discussed below.

Meeting the OBO Foundry Principles

Principle 1.2: The ontology is in or can be expressed in a common shared syntax.

There were some challenges associated with learning and using the tools for ontology development. Even though the Protégé group provides tutorials and extensive documentation online, it took a few weeks to become comfortable using Protégé. OntoFox also provided documentation and a tutorial, but it was not sufficient for a first-time user. A conversation and demonstration by an OntoFox representative was necessary to use the system properly. When using OntoFox a couple times I discovered it was missing the most recent versions of the ontologies and notified the OntoFox staff who then added the latest versions. When OBO changed the class URI format, I needed to recreate all the import files in OntoFox.

Principle 1.6: The ontology must be orthogonal to other ontologies already lodged within OBO.

The OBO Foundry does not provide methods or metrics on how to assess orthogonality. Ghazvinian et al utilized a prototype tool (Obomap) to analyze the overlap and reuse of classes in OBO ontologies in 2010[29]. As this was a prototype tool restricted to OBO Foundry ontologies, Obomap was not available to test ACO for orthogonality. However, it is unlikely that animal classes in ACO overlap with classes in other OBO ontologies based on a review of the information provided about each OBO ontology and an adequate search of each ontology.

Many ontologies listed on the OBO Foundry website are for the anatomy and development of specific research and model animals (Xenopus, Zebrafish, Fly, etc.). These animals are not within the scope of ACO.

Two ontologies on the OBO Foundry website contained some animal taxonomic content. Linnaean classes from the National Center for Biotechnology Information (NCBI) Taxonomy ontology were imported into ACO and ACO classes were developed as subtypes of these NCBI Taxonomy classes. NCBI Taxonomy contains the Linnaean hierarchy but does not contain common non-Linnaean groupings of animals needed for many applications nor a mechanism to extend the hierarchy. For example, NCBI Taxonomy has a concept for *Equus caballus* (horse) but there are no classes generally used by those responsible for husbandry of horses (e.g., mare for adult female horse). Similarly, NCBI Taxonomy lacks important production classes for animals like producers of meat and milk for human food. Another concern is that some of the Linnaean classes in the NCBI Taxonomy have inaccurate synonyms. For example, "cow" is a synonym of *Bos taurus* in NCBI but this is inaccurate because "cow" actually refers to the adult females of *Bos taurus*, *Bison bison*, or *Bos taurus* and *Bison bison* crosses. The term "cow" is also used for animals besides cattle (e.g., female whales). I will recommend that NCBI Taxonomy remove "cow" as a synonym of *Bos taurus*.

The National Cancer Institute Thesaurus (NCIt)[30] is listed on the OBO website and contains some animal classes but is not suitable for animal description purposes in veterinary medicine and animal husbandry. As an example, NCIt contains a code for "Cow" with text definition of "Domesticated ungulates that are raised for beef, dairy products, and leather, and used for labor" that does not mention the animal is an adult female. Many inappropriate synonyms are applied to this code, including outdated Linnaean classification terms (*Bos bovis*, *Bos primigenius taurus*) and misaligned terms (*Bos taurus*, Bovine, and Cattle.) In one sense, the Linnaean class *Bos taurus* is too narrow to correspond with cow because it does not include non-*Bos taurus* species and crosses which some laypeople would consider to be "cows." In another

sense, it is too broad because it does not include the sex and maturity qualities in the true definition of cow as adult female cattle. Bovine does not align as a synonym with cow because bovine includes males and younger females. Plus it is an adjective and most likely refers to members of *Bovinae* rather than *Bos taurus*. Cattle does not align with cow because it refers to both sexes and all ages groups. In addition, NCIt lacks a mechanism for formal definitions and is not curated by animal subject matter experts. I will submit a recommendation that NCIt correct their "Cow" class.

Principle 1.8: The ontology uses relations which are unambiguously defined following the pattern of definitions laid down in the OBO Relation Ontology.

There were no relations in RO to indicate where an organism lived so I initially used a placeholder relation of **lives in** to connect an organism with its living environment. Questions concerning existence of a suitable relation to connect an organism and its environment in RO were posed to the ENVO and the RO listservs however, no concrete suggestions were offered and multiple members posted that they needed similar relations. Stefan Schulz subsequently reviewed ACO and recommended using the relation **has locus** from BioTopLite instead of **lives in** so I made this substitution in ACO.

Principle 1.10: The ontology has a plurality of independent users.

Meeting the principle of multiple users seems difficult for new ontologies. Perhaps OBO expects that ontologies need to be out in the user community for some time to gain acceptance before submission for formal evaluation and inclusion into the OBO Foundry. This was not clear on the OBO website. I assume that initial potential interest in the ontology followed by subsequent actual use is allowed. I stated ACO met this principle because the subset already has multiple users and the ontology has one current user and a few interested parties.

21

Principle 1.11: The ontology will be developed collaboratively with other OBO Foundry

members.

Collaboration with OBO members was valuable during this project. I communicated with

other ontologists through email, web conferences, the OBO listservs, and in-person meetings at

the International Conference on Biomedical Ontology (ICBO). Even though most of the

collaborators lacked animal and veterinary medicine expertise, they were able to provide helpful

suggestions when given a detailed description of the issues encountered when developing ACO.

Feedback was sought on multiple issues and consensus was reached quickly in some cases. In a

few cases, there was no consensus among collaborators and I decided how to proceed taking

their comments under advisement.

I was unsure which classes to use in ENVO for water and nature so I consulted

with the ENVO listserv and were directed toward the appropriate ENVO classes *aquatic*

habitat and *terrestrial biome*, respectively.

I identified around 25 classes which should be added to existing OBO ontologies (see

Table 7). To date, *apiary* has been submitted through the ENVO tracker for addition (see Figure

5), *rumination* has been submitted through the GO tracker for addition, and four classes have

been submitted through the PATO tracker for addition. I have not heard back on the disposition

of the ENVO request but the GO and PATO requests were accepted.

A consensus was not reached on the OBO listserv for a preferred name for the ontology

so I choose the name "Animals in Context Ontology" over many alternatives for the following

reasons: 1) "Animals in context" refers to animals described according to their specific situation

or circumstance whereas "common" is ambiguous because it could refer to animals in common

use or to the common names associated with scientific names and 2) "Animals in context" does

not limit the scope to animals in particular environments such veterinary clinics, agricultural use, or zoos.

Principle 2.1: All candidate Foundry ontologies will appoint a person responsible for liaison with the OBO Foundry coordinating editors AND *Principle 2.2: All candidate Foundry ontologies will provide a tracker for additions and corrections, and a help desk for inquiries from users.*

OBO Foundry principles 2.1 and 2.2 impose a necessity for ongoing support and maintenance of the ontology. Although I agree that long-term maintenance is necessary, I note that no source of funding is mentioned. There does not seem to be a common source of funding for the ontologies and it is difficult to determine if the ontologies will continue to be funded. Some of the ontologies are driven by and supported by outside needs and projects and are subject to funding loss if the projects are not supported. Therefore, the future of the OBO Foundry ontologies seems uncertain. It is difficult to guarantee continued maintenance of ACO (and other OBO ontologies) if funding is not secure. Additional funding for developed ontologies could assist with maintenance and could lead to more curators and higher quality ontologies.

Principle 2.3 Textual definitions will, by degrees, be complemented with equivalent formal definitions.

Although the original CAN subset consisted of a majority of fully defined classes, I increased the number of fully defined classes in ACO by adding additional classes.

1. A class of *Preproduction role* was created that was not included in the CAN subset. This role is used for replacement animals when their production role has been predefined by the animal manager but is not realized yet. This created an additional 7 fully defined classes in ACO.

2. A class of *Disposition to ruminate* was created that was not included in the CAN subset. This disposition is applied to all animals that have a multi-compartmented stomach and have the organs necessary to ruminate, even if they are not currently ruminating (e.g. preruminant calves). This affected two classes directly (*Bovidae* and *Cervidae*) and around 50 classes indirectly (e.g., *Lactating cattle*).

ACO includes some general class axioms to further define the animals in roles. See Table 5 for an example. It shows how animals bearing a certain role can be considered equivalent to animals that participate in certain processes.

Principle 2.4: *Textual definitions will use the genus-species form: An A =def. a B which Cs, where B is the parent of the defined term A and C is the defining characteristic of those Bs which are also As.*

Writing text definitions in the genus-species format was not intuitive or easy at first, but became more natural with time and practice. This format of text definition was acceptable to one VTSL client but not to another client, which could hinder adoption of ACO. It might be necessary to add additional definitions to the ontology in a format that is more natural and acceptable to the non-ontologist layperson. An additional annotation property "has additional definition" might be necessary.

Principle 2.5: Single is_a inheritance: ontologies will distinguish a backbone ('asserted') is_a hierarchy subject to the principle of single inheritance.

New formal definitions were added in some areas to fulfill the single inheritance principle. In the past a Linnaean term *Ruminantia* consisted of many grazing mammals including those in *Bovidae, Cervidae*, and *Antilocapridae* but not *Camelidae;* however, this Linnaean term is now considered outdated by many taxonomists. I recognized the confusion over the use of the

term "ruminant" since it could refer to the outdated *Ruminantia* classification, an animal which possesses a rumen, or an animal which possesses the ability to ruminate. I consulted a veterinary anatomist at Virginia Tech who concurred that camelids are not of the former *Ruminantia* but are considered ruminants because they ruminate. In the absence of a Linnaean term for ruminants, a common grouping concept of *Ruminant* was used with a text definition of "Animal of the Order Artiodactyla which has a multi-compartmented stomach with the function to ruminate." In the CAN subset, ruminant concepts like *Bovidae* have two parents – their taxonomic parent and the concept *Ruminant.* To maintain a single **is a** hierarchy in ACO, the formal definition of **bearer of** some *disposition to ruminate* was given to the class *Ruminant* and to each class which is a ruminant (*Bovidae* and *Cervidae*). Therefore upon reasoning *Bovidae* and *Cervidae* are given additional **is a** connections to the class *Ruminant.*

Principle 2.6: Instantiability: Terms in an ontology should correspond to instances in reality.

I questioned the instantiability and candidacy for formal representation of the class *Non-human primate* so I sought input from the OBO listserv. A member referred us to a paper that described the use of the relations **never in taxon** and **only in taxon** with taxonomic classes as values. It seems feasible that *Non-human primate* could be formally represented as a subtype of *Order Primates* that is **only in taxon** *Order Primates* and **never in taxon** *Homo sapiens;* however, I did not choose to implement it during this project.

I could have saved some development time and possible error insertion by automating the transfer from the SNOMED concepts to OWL classes. I investigated this superficially but chose not to pursue this route for several reasons. I needed experience building an ontology in Protégé. Problems with SNOMED-CT's description logic and the extension concepts use of non-sanctioned relations prevented a seamless transfer. Finally, the need to evaluate each concept

individually for substitution of classes in the formal definitions precluded a streamlined automated transfer.

Principle 2.7: Terms should be created and defined using terms and relations drawn from other Foundry ontologies.

Numerous challenges were encountered while evaluating classes in other ontologies to include in ACO, including incomplete ontologies, incorrect classes in ontologies, and uncertainty over which classes to use. A discussion concerning an analysis of the four ontologies used to define ACO classes follows.

Gene Ontology

The Gene Ontology (GO) contained classes of biological processes needed to define ACO animal classes such as *lactation, growth*, and *adult feeding behavior* but GO was missing classes related to rumination (*rumination*, *disposition to ruminate*, and *pre-rumination stage*.) Questions concerning these desired classes and use of classes were submitted to the GO-discuss listserv. Suggestions for improvement were provided and resulted in submission of *rumination* as an addition to GO through the GO Tracker. GO does not carry disposition classes so *disposition to ruminate* will stay in ACO at this point.

I identified the GO Extension class *suckling behavior* to define several ACO classes (*Unweaned growing dog*, *Suckling pig for meat production*, etc); however, I had concerns with its text definition and one of its parents. I agree with its parent *Feeding behavior* but not with *Behavioral interaction between organisms* because a suckling animal could be bottle fed milk. I think the text definition ("Specific actions of a newborn or infant mammal that result in the derivation of nourishment from the breast") is too

26

specific because nursing does not necessarily result in nourishment (although that is the purpose) and breast is a human specialization of the mammary gland and does not technically apply to animals. I suggest that GO remove the undesired parent and edit the text definition to: A feeding behavior of a newborn or infant mammal for the purpose of deriving nourishment from milk from the mammary gland.

Environment Ontology

The Environment Ontology (ENVO) contained classes of some living environments needed to define ACO animal classes such as *research facility, farm* and *aquatic habitat*, but was missing common living environments for animals such as a *feedlot, cage*, and *exhibit*. ENVO also needs some extensive editing and one of the ENVO developers mentioned this. One example of an ENVO class that needs some attention is *pasture*. 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 I suggest the ENVO curators should either: 1. Move this class from *grassland* to *terrestrial habitat* and edit the text definition to: "Terrestrial habitat used for grazing, foraging or browsing by animals" or 2. Edit this class name to "grassland ranch pasture" and leave the text definition as is, then create a new class *pasture* with a text definition of "Terrestrial habitat used for grazing, foraging or browsing by animals."

Phenotypic Quality Ontology

The Phenotypic Quality Ontology (PATO) contained classes of organismal qualities needed to define ACO animal classes such as *female, neonatal* and *prepubescent,* but was

27

missing classes related to reproductive status such as *intact male*, *intact female*, and *castrated male*.

<u>NCBI Taxonomy</u>

The NCBI Taxonomy ontology contained Linnaean classes such as *Aves* and *Apis* to serve as superclasses of ACO animal classes but was missing some needed taxonomic classes such as *Suinae* and *Myomorpha*. In addition, it does not include taxon quality attributes. It is unclear how the mixture of Linnaean and Cladistic classes in NCBI Taxonomy is to be used.

I did not give NCBI Taxonomy classes a stated superclass of a common animal grouping class for a few reasons. One is to maintain a single **is a** inheritance to meet Principle 2.5. Second, I am unsure if there is enough agreement in the world about which Linnaean classes belong to a common animal group to make these connections. Lastly, it would be necessary to properly describe and formally define the phenotypic, physiologic and other characteristics that differentiate Linnaean classes to determine membership in a common animal group. This analysis was beyond the scope of this project, but would allow for inferred **is a** relationships from Linnaean classes to ACO common animal classes.

Principle 2.10: Preferred terms should consist entirely of singular nouns.

I identified three categories of non-compliance in the source terminology: 1. Plural nouns where singular form exists (eggs); 2. Single noun and plural noun are same (deer); 3. Mass/plural noun where no singular form exists (cattle and broodstock). All classes with the plural "eggs" in the preferred name were changed to the singular "egg." Deer and other similar classes were left as is since no exclusive singular form exists and the names are in fact proper singular forms. All classes with "broodstock" in the preferred name were edited to include "breeding" instead and broodstock terms were retained as synonyms. The issue of a singular form of cattle was

presented to the OBO listserv. Multiple suggestions were given and the most promising was "head of cattle" for a singular form. Although this is technically correct, it is not how people in the animal discipline talk (which would violate Principle 2.11) so I chose to keep "cattle" in ACO's classes. In addition, the primary audience of ACO usually does not refer to a singular "head of cattle" but instead to a more specific name for the animal such as heifer, cow, or calf.

Comparing Animal Classes in SNOMED and OWL

This work provided an opportunity to investigate the functionality of SNOMED-CT's proposed organism model. Organisms in ACO are defined using both intrinsic (physiologic or phenotypic) characteristics such as sex and extrinsic, human defined characteristics such as production role. Defined classes were created to infer additional grouping connections. As an example, *Cattle as human food* is a subclass of *Cattle* in ACO. After reasoning, it classifies under the *Food Animal* because of its formal definitions. This demonstrates how organisms in SNOMED might be classified by a combination of various intrinsic and extrinsic qualities such as Linnaean classification, morphology, Gram stain quality and antigenic formulae in bacterial nomenclature.

There are multiple advantages of having the animal classes in an OWL-DL format over the SNOMED-CT subset. In addition to the benefits of using a common structure, classes and relations that enable interoperability with hundreds of biomedical ontologies, the OWL-DL ontology includes more options to define, manipulate and annotate classes. OWL-DL includes more expressive class and property axioms, and domain, range, existential, universal and other cardinality restrictions. The OWL Viz plug-in for Protégé allows graphic visualization of the ontology (see Figure 6 for an example). ACO uses a reasoner for classification and creation of a

poly-hierarchical structure that may be more desirable for some applications. ACO is openly

available while SNOMED is open but requires a license to use.

There are some advantages of the SNOMED-CT extension and subset. The organization

that manages SNOMED-CT, the International Health Terminology Standards Developing

Organization (IHTSDO), defines models for concepts for each hierarchy. They also distribute

extensive documentation for technical reference, implementation, and editorial guidelines twice a

year. In addition, many organizations worldwide have adopted SNOMED. However, ACO is

linked to SNOMED-CT through the annotation property **has alternative id** with value of the

SNOMED-CT concept identifier so users can still benefit from some of these advantages

provided they comply with the SNOMED-CT licensing requirements. If the proposed SNOMED

organisms model is adopted, continuous alignment of ACO with SNOMED-CT will be

maintained with ease. Adoption of a different organisms model at SNOMED may affect the

current alignment approach of ACO with SNOMED.

Future Directions

There are many areas for further work. As use of ACO increases, I anticipate the need for

addition, deletion and refinement of classes based on user input and requests. Further study is

needed on how to define animal specific production classes such as *Broiler chicken* and

Finishing turkey. These classes have inconsistent husbandry practices and lack universal

definitions for age ranges, sizes and weight ranges. Additional useful grouping classes such as

Shellfish, Cold blooded animal, *Duck* and *Nonhuman primate* should be investigated for formal

representation so they could infer Linnaean and other classes as subclasses. These classes are

currently defined by **is a** relations and text definitions, which preclude the inclusion of relevant

subclasses. Further analysis of the relations used to define the relationship between an animal

and its living environment is needed. **Has locus**, currently used in ACO, could be given further refined and defined subrelations.

It was suggested by a couple OBO members that the scope of ACO is too broad and that I should consider dividing ACO into three distinct ontologies: 1) An ontology of animal production classes; 2) An ontology of common animal groupings; and 3) An ontology of animal development stages. Perhaps division based on characterization by intrinsic or extrinsic factors would be a better choice, with resultant ontologies of: 1) An ontology of animal production classes; and 2) An ontology of common animal groupings by phenotypic and physiologic factors. However, division of ACO might not make the most sense given the limited resources and expertise needed to maintain animal ontologies.

Additional analysis of how to best represent organism characteristics external to the organism is needed. A study to determine if the Linnaean and Cladistic classes are properly organized in the NCBI Taxonomy is a direction for future work. Increasing the automation of ACO development and maintenance is another area for future work. This includes the automated bidirectional transfer of classes and changes between the SNOMED-CT extension and ACO, as well as the automated update of the OntoFox import files.

Formal evaluation of ACO by OBO members and inclusion on the OBO Foundry or NCBO Bioportal website are future goals. ACO needs to be delivered to the communities of veterinary medicine, animal production and husbandry and biomedical research so they can start utilizing classes that enable data collation and comparison.

CONCLUSION

Following the OBO Foundry principles, I created an ontology of animal classes conforming with most of the standards for modern biomedical ontologies. This project highlighted the successes and challenges of this process and provided suggestions for improvements. OBO Foundry members were willing to help with the development of the ontology. Tools were available to search and import classes from OBO ontologies. However, I encountered challenges trying to follow some of the principles. The ontologies utilized on the OBO Foundry website are very much a work in progress - they do not contain all of the content needed for defining animals and some need improved organization and maintenance. Additions to and changes needed in the OBO ontologies identified in this work should facilitate future ontology development involving animals. The future of funding ontology development and maintenance through the OBO Foundry is unclear. The resulting ontology is an improvement over the source SNOMED-CT extension and subset as its format and structure makes it interoperable with more ontologies and systems and boasts greater opportunities for classification and data analysis.

FIGURES
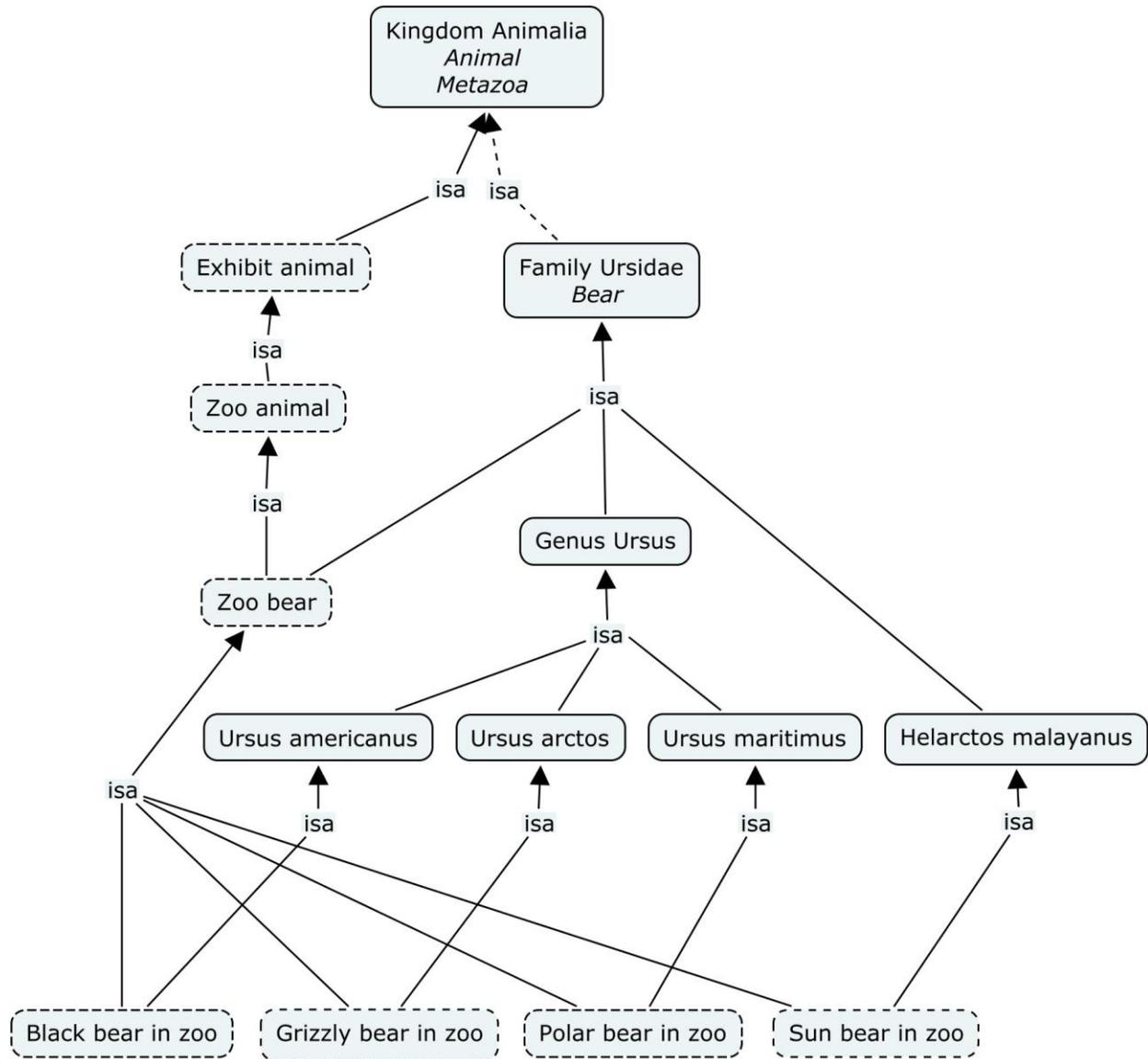
Figure 1. Concept map representation of a portion of the Common Animal Names (CAN) subset concepts and their hierarchical arrangement. Dotted line boxes represent non-Linnaean concepts and solid line boxes represent Linnaean concepts in SNOMED-CT. The bottom level extension concepts (e.g., *Black bear in zoo***) contain a common grouping parent (***Zoo bear***) and a Linnaean parent (***Ursus americanus***). The hierarchical classification between** *Family Ursidae* **and** *Kingdom Animalia* **is eliminated from this example for brevity.**
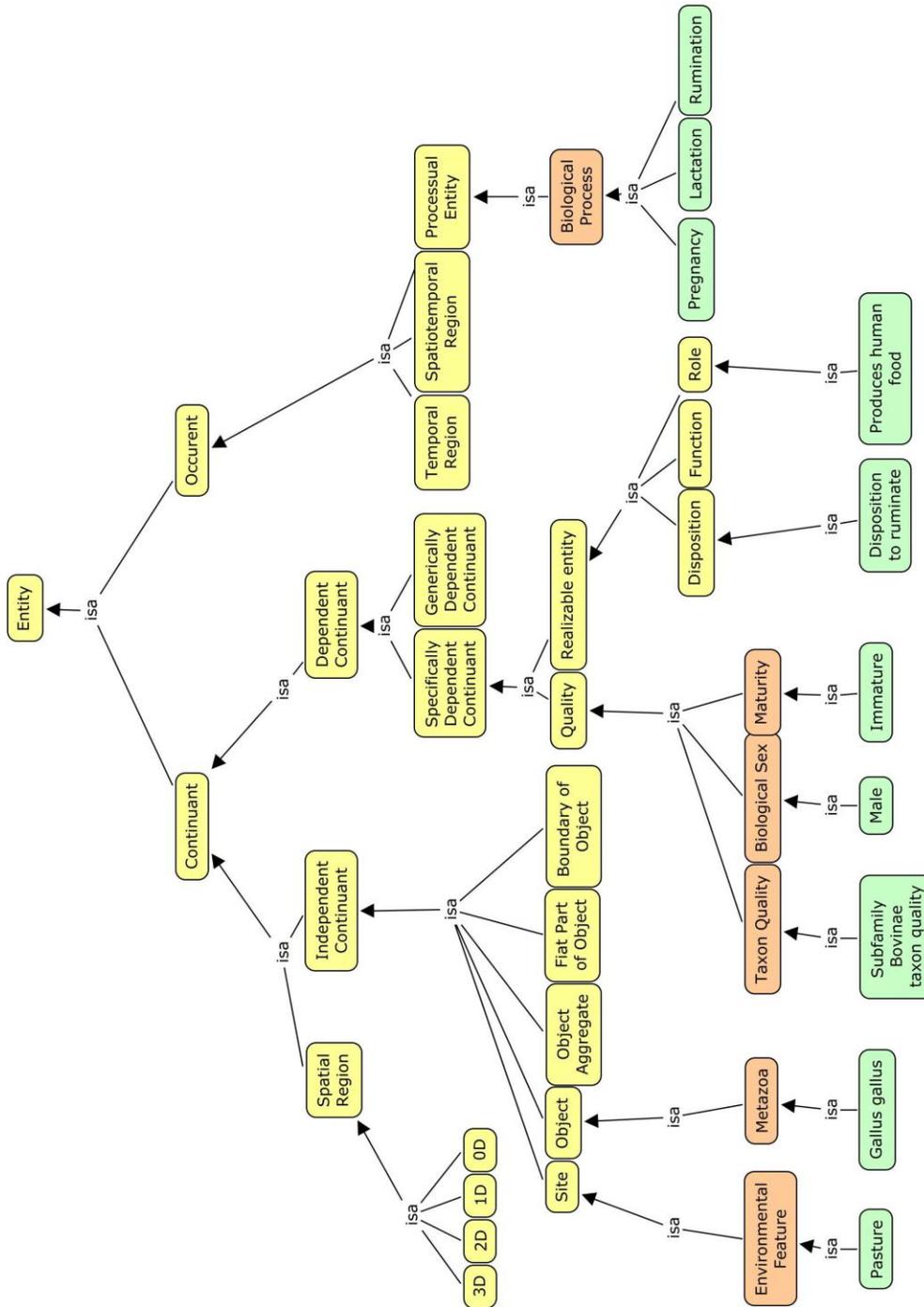
Figure 2. Example classes used in the Animals in Context Ontology (ACO) organized in the Basic Formal Ontology (BFO). Classes in yellow are from BFO. Classes in green are used to define ACO classes. Classes in orange serve as intermediate organizational classes between more granular classes and upper level BFO classes. Some intermediate classes are not shown due to space limitations.
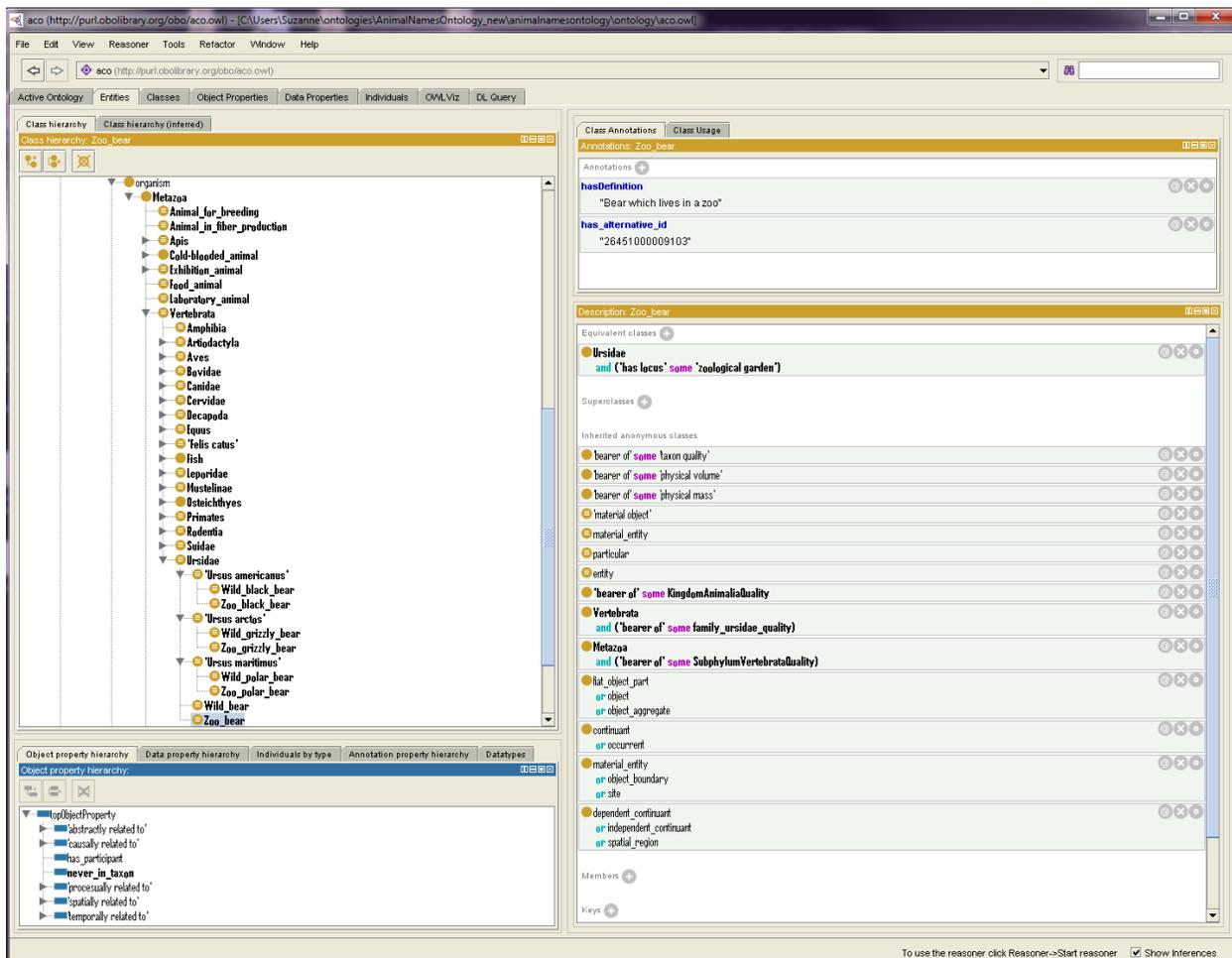
Figure 3. Part of bear hierarchy in the Animals in Context Ontology (OWL file) viewed with Protégé. The area on the left shows the classes organized in a hierarchy. Upper right hand frame shows the annotations for the text definition and alternative id (SNOMED concept id) the highlighted class *Zoo bear*. The lower right hand frame shows the formal definition of the class.

Figure 4. Abbreviated image of the OntoFox program. In 2(a) *lactation* **and** *biological process* **from the Gene Ontology are identified as classes to import into ACO. In 2(b) I state that** *lactation* **will be a subclass of** *biological process* **in the import file.**

Figure 5. Image of a class request of *apiary* through the ENVO tracker. The request can be viewed at http://sourceforge.net/tracker/?group_id=76834&atid=934923.
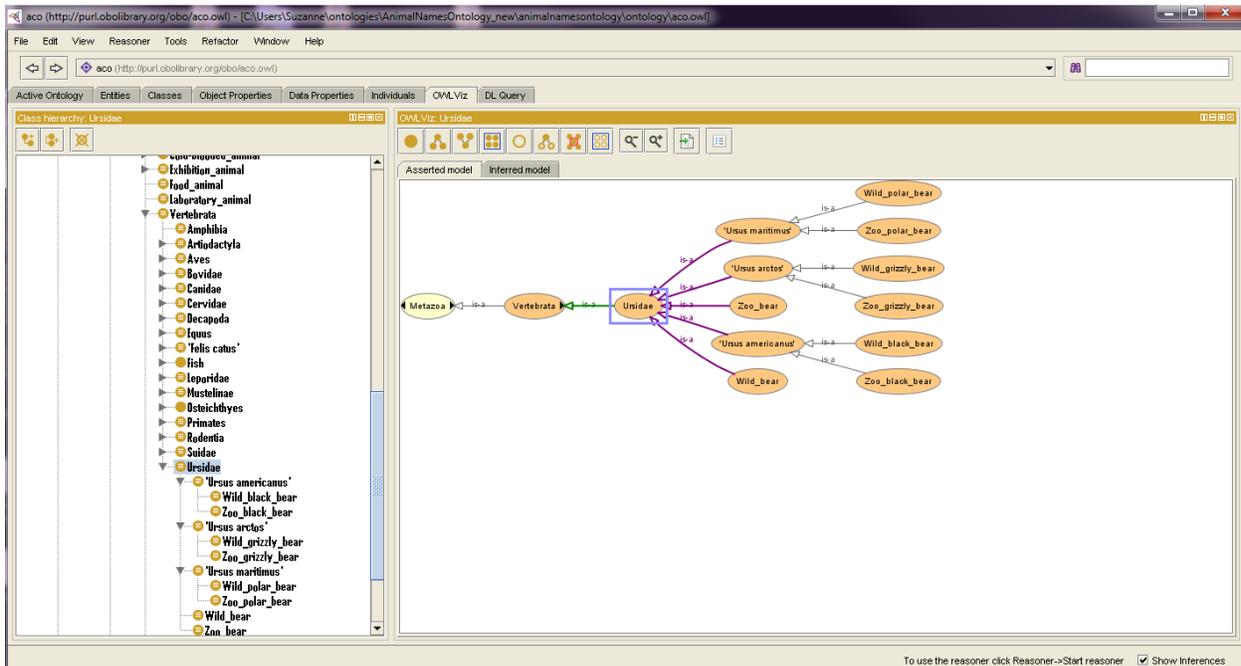

Figure 6. Image of the OWL Viz tool to visualize the animal class hierarchy in Protege. This images shows the superclass and subclasses of *Ursidae* in the Animals in Context Ontology.

TABLES

Table 1. List of the Open Biological and Biomedical Ontologies (OBO) Foundry Principles accepted as of 2006 (1.1-1.11) and 2008 (2.1-2.11).

Number	OBO Foundry Principle	Compliance of ACO with Principle
1.1	The ontology must be open and available to be used by all without any constraint other than (a) its origin must be acknowledged and (b) it is not to be altered and subsequently redistributed under the original name or with the same identifiers.	Met
1.2	The ontology is in, or can be expressed in, a common shared syntax. This may be either the OBO syntax, extensions of this syntax, or OWL.	Met
1.3	The ontology possesses a unique identifier space within the OBO Foundry. The identifier uniquely and persistently identifies a definition, which itself unambiguous identifies some type of biological entity. The identifier is for the definition: it is NOT the name and it is NOT an identifier for the name.	In negotiation
1.4	The ontology provider has procedures for identifying distinct successive versions.	Met
1.5	The ontology has a clearly specified and clearly delineated content.	Met
1.6	The ontology must be orthogonal to other ontologies already lodged within OBO. For each domain, there should be convergence upon a single reference ontology that is recommended for use by those who wish to become involved with the Foundry initiative.	Met
1.7	The ontologies include textual definitions for all terms.	Met
1.8	The ontology uses relations which are unambiguously defined following the pattern of definitions laid down in the OBO Relation Ontology.	Met
1.9	The ontology is well documented.	Met
1.10	The ontology has a plurality of independent users.	Partially met
1.11	The ontology will be developed collaboratively with other OBO Foundry members.	Met
2.1	All candidate Foundry ontologies will appoint a person responsible for liaison with the OBO Foundry coordinating editors. This person will be a member of the Foundry Board of Associate Editors.	Met
2.2	All candidate Foundry ontologies will provide a tracker for additions and corrections, and a help desk for inquiries from users.	Unmet
2.3	Textual definitions will, by degrees, be complemented with equivalent formal definitions	Met
2.4	Textual definitions will use the genus-species form: An A =def. a B which Cs, where B is the parent of the defined term A and C is the defining characteristic of those Bs which are also As.	Met

2.5	Single is_a inheritance: ontologies will distinguish a backbone ('asserted') is_a hierarchy subject to the principle of single inheritance (each term in the ontology has maximally one is_a parent in this asserted hierarchy). Further is_a relations may be inferred.	Met
2.6	Instantiability: Terms in an ontology should correspond to instances in reality. (First audit trail principle.)	Met
2.7	Terms should be created and defined using terms and relations drawn from other Foundry ontologies. (Second audit trail principle.) When the terms and relations needed to create definitions of Foundry terms do not already exist in the Foundry, they should be submitted for inclusion to the relevant ontology. In this way no useful term goes undefined, and we do not land up with definitions that are in terms that themselves are not defined (except for the upper level undefinables).	Met
2.8	Evaluation. Each Foundry ontology should be subject to evaluation (as far as possible quantitative): software (conversion OBO format <-> OWL), specialist review (OWL -> controlled English) when one version is used for a given purposes later versions should be applied to the same purpose and results compared	Partially met
2.9	Use of Basic Formal Ontology (BFO): Each Foundry ontology should be built on the basis of BFO top-level distinctions (common top level), selected along the following lines: Does your ontology recognize both things that continue to exist (continuants) and also processes in which such things participate (occurrents)? Does your ontology recognize a distinction between entities (qualities, functions, roles, dispositions, ...) which have bearers (specifically dependent continuants) and those bearers themselves (independent continuants: molecules, cells, organisms, etc.) Does your ontology recognize information artifacts (sequences in a database, protocols, publications ...)?	Met
2.10	Singular Nouns: Ontologies consist of representations of types in reality – therefore, their preferred terms should consist entirely of singular nouns.	Substantially met
2.11	Preferred terms should be nouns and noun phrases belonging to ordinary English as extended by technical terms already established in the relevant discipline – they should not employ phrases like 'EV-EXP-IGI'; they should not employ lab slang, they should not employ ellipses.	Met

Table 2. The OBO ontologies utilized in the Animals in Context Ontology.

External Ontology	Use in ACO	Example Class
Basic Formal Ontology (BFO)	Upper level hierarchy	*Continuant*
BioTopLite (BTL)	Top domain hierarchy and relations	*Organism,* **Bearer of**
Environment Ontology (EnvO)	Environment sites	*Farm*
Gene Ontology (GO)	Biological processes	*Lactation*
Phenotypic Quality Ontology (PATO)	Phenotypic qualities and functions	*Female*
NCBI Taxonomy	Taxonomy (Linnaean)	*Gallus gallus*

Table 3. Classes imported from OBO ontologies as well as additions and corrections needed in these ontologies.

Ontology	No. Classes Imported	No. Classes Identified for Correction	No. Class Additions Needed
Basic Formal Ontology (BFO)	39	0	0
BioTopLite (BTL)	49	0	0
BioTopLite –BFO Bridge	39	0	0
Environment Ontology (EnvO)	10	2	15
Gene Ontology (GO)	7	1	3
Phenotypic Quality Ontology (PATO)	13	0	5
NCBI Taxonomy	40	2	3

Table 4.The former and current relation attributes and values for the formal definition of the concept *Breeding bull for beef production (organism)*.

CAN Subset Relation Attribute	CAN Subset Relation Value	ACO Relation Attribute	ACO Relation Value
Has life circumstance (attribute)	*Produces beef for human consumption (animal life circumstance)*	**Bearer of**	*Produces beef for human consumption*
Has life circumstance (attribute)	*Used as breeding stock (animal life circumstance)*	**Bearer of**	*Used as breeding stock*
Has physiologic state (attribute)	*Physically mature (finding)*	**Bearer of**	*mature*
Has physiologic state (attribute)	*Intact male (finding)*	**Bearer of**	*Intact male*

Table 5. General class axiom for an animal who is bearer of the role class *Produces wool*. *Produces wool* will be requested as an addition to the Environment Ontology (ENVO).

'bearer of' some *Produces_wool*
    EquivalentTo '**participates in'** some
        (*Production* and ('**has outcome'** some *Wool product*))

Table 6. Formal definitions for ACO class *Lactating dairy cattle*. The text definition is: Female dairy cattle which is currently producing milk. The relation and value in the first row distinguishes this class from its parent *Cow for milk production*. Other connections are inherited. Some inherited connections are not shown for brevity.

BioTopLite Relation	Value Example	Ontology Source
participates in	*Lactation*	GO
bearer of	*Female*	PATO
bearer of	*Mature*	PATO
bearer of	*Produces milk for human food*	ACO
bearer of	*Subfamily bovinae quality*	ACO
bearer of	*Disposition to ruminate*	ACO

Table 7. List of proposed class additions to OBO ontologies needed to define ACO classes.

Ontology	Proposed Class Name of Addition	Propose Text Definition of Addition
GO	*Rumination*	Digestive process in which food, usually grass or hay, is swallowed into a multicompartmented stomach, is regurgitated, chewed again, and swallowed again
GO	*Prerumination*	Physiologic process of young ruminant by which an esophageal groove is created by neural stimulation of esophageal muscle in response to suckling and milk consumption and anticipation which redirects the flow of milk past the rumen and into the abomasum. Intake of water and non-milk feed materials decreases this effect and allows for the development of the rumen and rumination. Once the rumen is fully developed and the ruminant is weaned (no longer suckling) the esophageal groove is no longer present and the preruminate phase is over
Unknown	*Disposition to ruminate*	Disposition, once physiologic development has progressed sufficiently, to breakdown food, usually grass or hay, via swallowing into a multicompartmented stomach, regurgitate, chew again, and swallow again, as part of the digestion of the organism

ENVO	*Confined agricultural feature*	An environmental feature for plant or animal organisms that allows for control of the nutrition and conditions in which the organism is living
ENVO	*Cage*	A confined agricultural feature where the animal is enclosed in a box-like structure which restricts the movement of an animal to a specific area
ENVO	*Drylot*	A confined agricultural feature that is free of vegetation and is used for the containment and feeding of livestock
ENVO	*Feedlot*	A confined agricultural feature that is free of vegetation and is used for the containment, feeding, and fattening of livestock
ENVO	*Exhibit*	An environmental feature that allows for viewing by people
ENVO	*Apiary*	A place where beehives are managed for commercial honey production
PATO	*Egg*	A round or oval body laid by the female of many animals, consisting of an ovum surrounded by layers of membranes and an outer casing, which is capable of nourishing and protecting a developing embryo and its nutrient reserves
PATO	*Intact female*	Female quality of having complete, unaltered reproductive organs
PATO	*Intact male*	Male quality of having complete, unaltered reproductive organs
PATO	*Castrated male*	Male quality of being incapable to reproduce because of functional loss of the testicles from surgical removal or chemical means
PATO	*Nulliparous*	Organismal quality of a viviparous organism having never given birth
NCBI Taxonomy	*Suinae*	
NCBI Taxonomy	*Myomorpha*	
NCBI Taxonomy	*Osteichthyes*	
ENVO	*Dairy product for human consumption*	
ENVO	*Food product for human consumption*	
ENVO	*Egg product for human consumption*	

ENVO	*Flesh product for human consumption*	
ENVO	*Beef product for human consumption*	
ENVO	*Veal product for human consumption*	
ENVO	*Honey product for human consumption*	
ENVO	*Fiber product*	
ENVO	*Wool product*	

Table 8. Examples of errors identified in and subsequent corrections made to the Common Animal Names (CAN) Subset of SNOMED-CT while developing the Animals in Context Ontology (ACO).

Error Type	Example of Error	Example of Solution
Concept was missing a formal definition when it could be created	*Honey bee in apiary* did not include a location of apiary in its formal definition	Created a new concept for *apiary* in the SNOMED extension and added it as a defining value to *honey bee in apiary*
Non-Linnaean concept did not have a text definition	*Pony* did not have a text definition	Created a text definition for *Pony*
Concept was duplicated in SNOMED core and extension	There were separate concepts for *zebra* in SNOMED core and the VTSL extension	Retired the extension concept *Zebra* and referred it to the core concept *Zebra*
Concept was missing a logical parent	*Alces alces* (elk) and *Cervus elaphus* (moose) did not have **is a** connections to *Cervidae* in CAN subset	Added an **is a** relationship from *Alces alces* to *Cervidae* and from *Cervus elaphus* to *Cervidae*
Concept's formal definition was self-referential	The concept *aquarium animal* had a relation **has life circumstance** to *aquarium animal* rather than *lives in water*	Corrected the formal definition by removing *aquarium animal* and replacing with *lives in water*

REFERENCES

1. The Open Biological and Biomedical Ontologies. Available at: http://obofoundry.org/.

2. Smith B, Ashburner M, Rosse C, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature biotechnology*. 2007;25(11):1251-5. Available at: http://dx.doi.org/10.1038/nbt1346.

3. The Gene Ontology. Available at: http://geneontology.org/.

4. Bodenreider O, Stevens R. Bio-ontologies: current trends and future directions. *Briefings in bioinformatics*. 2006;7(3):256-74. Available at: /pmc/articles/PMC1847325/?report=abstract.

5. Rubin DL, Shah NH, Noy NF. Biomedical ontologies: a functional perspective. *Briefings in bioinformatics*. 2008;9(1):75-90. Available at: http://bib.oxfordjournals.org/cgi/content/abstract/9/1/75.

6. The International Health Terminology Standards Development Organization. Available at: http://www.ihtsdo.org/.

7. OBO Foundry Mailing Lists and Trackers. Available at: http://www.obofoundry.org/cgi-bin/discussion.cgi.

8. Basic Formal Ontology (BFO). Available at: http://www.ifomis.org/bfo.

9. Open Biological and Biomedical Ontologies: Current Principles. Available at: http://www.obofoundry.org/crit.shtml.

10. Smith B, Brochhausen M. Putting biomedical ontologies to work. *Methods of Information in Medicine*. 2010;49(2):135-140. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20135080.

11. Hogan W. Aligning the top-level of SNOMED-CT with Basic Formal Ontology. *Nature Precedings*. 2008. Available at: http://precedings.nature.com/documents/2373/version/1.

12. Veterinary Terminology Services Laboratory at Virginia Tech. Available at: http://vtsl.vetmed.vt.edu/.

13. SNOMED-CT Technical Implementation Guide. Available at: www.snomed.org/tig .

14. OBO Relations Ontology. Available at: http://code.google.com/p/obo-relations/.

15. The Protégé Ontology Editor and Knowledge Acquisition System. Available at: http://protege.stanford.edu/.

16. OntoFox. Available at: http://ontofox.hegroup.org/.

17. Xiang Z, Courtot M, Brinkman RR, Ruttenberg A, He Y. OntoFox: web-based support for ontology reuse. *BMC Research Notes*. 2010;3(1):175. Available at: /pmc/articles/PMC2911465/?report=abstract.

18. National Center for Biotechnology Information (NCBI) Taxonomy Ontology. Available at: http://www.ncbi.nlm.nih.gov/taxonomy.

19. Ontology Lookup Service. Available at: http://www.ebi.ac.uk/ontology-lookup/.

20. National Centers for Biomedical Computing (NCBO) BioPortal. Available at: http://bioportal.bioontology.org/.

21. BioTop Ontology. Available at: http://www.imbi.uni-freiburg.de/ontology/biotop/.

22. Wilson & Reeder's Mammal Species of the World. Online database. Available at: http://vertebrates.si.edu/msw/mswcfapp/msw/index.cfm.

23. Phenotypic Quality Ontology. Available at: http://obofoundry.org/cgi-bin/detail.cgi?id=quality.

24. The Environment Ontology (EnvO). Available at: http://environmentontology.org/.

25. TaxDemo Ontology. Available at: http://www.imbi.uni-freiburg.de/ontology/biotop/.

26. Schulz S, Stenzhorn H, Boeker M. The ontology of biological taxa. *Bioinformatics (Oxford, England)*. 2008;24(13):i313-21. Available at: /pmc/articles/PMC2718636/?report=abstract.

27. Animal Names Ontology Google Code Site. Available at: http://code.google.com/p/animalnamesontology/.

28. Tortoise SVN Subversion Tool. Available at: http://tortoisesvn.net/.

29. Ghazvinian A, Noy NF, Musen MA. How orthogonal are the OBO Foundry ontologies? *Journal of biomedical semantics*. 2011;2 Suppl 2(Suppl 2):S2. Available at: /pmc/articles/PMC3102891/?report=abstract.

30. National Cancer Institute (NCI) Thesaurus. Available at: http://ncit.nci.nih.gov/.

31. Smith B. What is an ontology and what is it useful for? *Introduction to Biomedical Onotologies*. Available at: http://ontology.buffalo.edu/08/TrainingCourse/1-Introduction.pdf.