

Multi-Species Anatomy Ontology Development Requires a Pluralist Approach to Label-Class Mapping

István Mikó, Matthew J. Yoder, Matthew A. Bertone,
Katja C. Seltmann, Andrew R. Deans

North Carolina State University, Department of Entomology, Raleigh, NC, USA

Abstract. Scientists use labels to reference complex anatomical entities that could, in practice, be represented by spelling out their circumscription in excruciating detail. The referential approach, however, potentially limits the utility of the work by forcing the consumer to interpret the underlying meaning sought by the author. These problems of interpretation are exacerbated when the user communities grow to include experts with disparate training and conventions, *e.g.*, the typical community of multispecies anatomy ontology consumers. Anatomy ontologies are poised to address these problems by providing logically parsable definitions for classes and by reconciling the labels and their contexts. Here we discuss mechanisms used by the Hymenoptera Anatomy Ontology project to solicit broad buy-in and promote harmony amongst a very broad array of users.

Keywords: biomedical ontologies, semantic conflicts, preferred terms, Hymenoptera Anatomy Ontology

1 Representing Common Classes, Rather than Forcing Common Vocabularies

A strong argument for the development of an ontology is that it can stand as a controlled vocabulary that unifies within and cross-domain knowledge [1, 3]. There are obvious benefits to a 1:1 relationship between labels and classes [2]: improved readability and navigability of ontology class hierarchies and the facilitation of alignment and integration. An emphasis on 1:1 label-class relationships, however, inhibits broad buy-in.

Any mechanism that promotes the adoption of ontologies by domain experts is valuable [4]. We propose that conflicts arising from debates focused on “preferred terms” or absolute disambiguation are more likely to occur within multi-species ontologies, in which the spectrum of expertise across domain experts is particularly broad, and in which synonymy and homonymy are more prevalent due to isolation of research in particular sub-domains. In a practical context, ignorance of ontologies by domain experts might easily turn to aversion if they are ordered to follow a

terminology that is represented in a difficult and cutting edge construct (ontology). Eliminating or minimizing this conflict (*i.e.* the emphasis of the importance of a 1:1 relationship) should result in an increased acceptance of the ontology.

The requirement for unique labels also introduces problems across ontologies, which results in the proposal of *ad hoc* solutions, like appending organismal or other context-based strings in front of labels (see discussion on the OBO-listserv from 2010 [5]). Application-level mechanisms exist already for the disambiguation of concepts, *e.g.*, meaningless identifiers, and so debates about unique labels ultimately displace effort that could otherwise be spent on class clarification. Aspiring for ontological exactness using referents that can be trivially misinterpreted will ultimately fail. The term “process”, for example, will never have a single meaning, either functional or anatomical. However, a meaningless, globally unique identifier – ID:012345 – stands a chance of stable mapping. Together these arguments emphasize the importance of a shared set of common classes rather than a common vocabulary.

2 Facilitating Adoption of Multi-Species Anatomical Ontologies – A Pluralist Approach and Currency of Adoption

A pluralist approach to ontology development is one that de-emphasizes the importance the 1:1 relationship between label and classes. The Hymenoptera Anatomy Ontology (HAO) embraces and facilitates this approach via the “sensu” model [6]. A “sensu” is simply a combination of class, label, and reference, which points to an author and the context. A domain-user mechanism for interacting with an ontology employing sensus is, minimally, a text “URI table” with three columns that can be included in a “materials and methods” section of a published work. This table links the annotator’s or author’s text to the ontological meaning intended. The first column is required and contains a label for the annotation or publication. The second column is optional and contains the URI of the class being referenced if the author has found a match in the ontology. In the case of the HAO we have provided a simple search interface that guides the user towards the definition being sought, and, when found, provides the necessary URI. The third column, required only when the URI is empty, contains a definition, preferably presented as a genus-differentia and formatted according to the ontology’s guidelines (though this is not necessary). This definition represents a new class or a proposed correction if the second column is also populated. This type of information is typically already provided within publications authored by multi-species domain experts, like taxonomists or phylogeneticists, the only necessary addition being the inclusion of a URI. A discussion of the table can be included, which would ideally articulate arguments for why definitions were altered or why a label was used in a novel combination.

The URI table acts as currency for interaction between domain experts and ontology curators. It is easily interpreted by the ontology curators, who better understand the logical constraints of the ontology. By facilitating interaction, it further promotes interaction with the ontology, which will ultimately shape the value of the ontology,

and it guides people onto a path towards the ultimate adoption of the central principles for ontology building. Furthermore, when formally included, a URI table acts as the basis for various metrics by which a “preferred” term could be computed, should it be deemed necessary (e.g. number of usages, number of conflicting usages, temporal distribution of use).

A pluralist approach has various inherent advantages. It frees domain expert curators to focus the more important aspects of the ontology – its class definitions and relations – and to act as supervisors rather than dictators. This approach provides domain experts the freedom to use their own labels and to explicitly accept or reject individual classes (*i.e.* logical definitions) of a given ontology without concern as to the label representing that class. In this framework authors are free to propose new labels with confidence, understanding that others will be able interpret their work based on definitions, they are also given full credit for their additions to the ontology (via the sensu).

At its core, a pluralist approach emphasizes the distinction between the referencing mechanisms (IDs, labels), and the underlying meaning of an ontology (its class definitions and relations). Ontologies with definitions that are inconsistently written, poorly conceived, or at worst non-existent, while referenceable, are destined to confuse, or in a worse case scenario, positively mislead. *Domain experts are the primary source of meaningful input* to expert bodies of anatomical data, as such we must minimize the hurdles they must leap to engage an ontology.

We propose that the historical importance of a controlled vocabulary be downplayed, recognizing that there are many application and publication level mechanisms for referencing ontologies that do not require a 1:1 bound of label to class.

References

1. Noy, N.F., McGuinness, D.L.: *Ontology Development 101: A Guide to Creating Your First Ontology* by Noy, N.F. and McGuinness, D.L. SMI technical report SMI-2001-0880, Stanford University (2001)

2. Schober, D., Smith, B., Lewis, S.E., Kusnierczy, W., Lomax, J., Mungall, C., Taylor, C.F., Rocca-Serra, P., Sansone, S.A.: Survey-based naming conventions for use in OBO Foundry ontology development. *BMC Bioinformatics*, 10, 125 doi:10.1186/1471-2105-10-125 (2009)
3. Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L.J., Eilbeck, K., Ireland, A., Mungall, C.J., OBI Consortium, Leontis, N., Rocca-Serra, P., Ruttenberg, A., Sansone, S.A., Scheuermann, R.H., Shah, N., Whetzel, P.L., Lewis, S.: The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology*, 25, 1251–5. doi:10.1038/nbt1346 (2007)
4. Swartout, W.R., Patil, P., Knight, K., Russ, T.: Toward Distributed Use of Large-Scale Ontologies. In: *Proceedings of the 10th Knowledge Acquisition for Knowledge-Based Systems Workshop*. Banff, Canada (1996)
5. Multiple authors. http://sourceforge.net/mailarchive/message.php?msg_id=26794461. Obo-discuss listserv at <https://lists.sourceforge.net/lists/listinfo/obo-discuss> (2010)
6. Yoder, M.J., Mikó, I., Seltmann, K.C., Bertone, M.A., Deans, A.R.: A Gross Anatomy Ontology for Hymenoptera. *PLoS ONE* 5 (12), e15991. doi:10.1371/journal.pone.0015991 (2010)