frontiers in
GENETICS

EDITORIAL
published: 04 February 2014
doi: 10.3389/fgene.2014.00018

Editorial: biological ontologies and semantic biology

John M. Hancock*

Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, UK
**Correspondence: jmhancock@gmail.com*

Edited and Reviewed by:
Richard D. Emes, University of Nottingham, UK

Keywords: semantic biology, biological ontologies, semantic web, data representation, data analysis

As the amount of biological data and its diversity accumulates massively there is a critical need to facilitate the integration of this data to allow new and unexpected conclusions to be drawn from it.

The Semantic Web comprises web-based technologies that allow linking of data between diverse data sets. Semantic Biology is the application of semantic web technology in the biological domain (including medical and health informatics). The Special Topic in Biological Ontologies and Semantic Biology brings together papers in this broad area—which spans computer science, computational biology and bioinformatics—providing a platform for strengthening what is still a new and underappreciated area of research.

A key aspect of semantic biology is the description of biological, and biology-related, entities using ontologies. Ontologies are a critical requirement for such integration as they allow conclusions drawn about biological experiments, or descriptions of biological entities, to be understandable and integratable despite being contained in different databases and analyzed by different software systems. Ontologies are the standard structures used in biology, and more broadly in computer science, to hold standard terminologies for particular domains of knowledge. They consist of sets of standard terms, which are defined and may have synonyms for ease of searching and to accommodate different usages by different communities. These terms are linked by standard relationships, such as "is_a" (an eye "is_a" sense organ) or "part_of" (an eye is "part_of" a head). In this way more detailed (granular) terms can be linked to broader terms, allowing computation to be carried out that takes these relationships into account.

The classical biological ontology is the Gene Ontology (GO) (Ashburner et al., 2000) which addresses aspects of gene function, the processes in which they participate and the localization of gene products. Increasingly, semantic biology requires the linkage of these concepts to other biological features. Three such biological entities are included in the Special Topic. The Anatomical Entity Ontology (AEO) (Bard, 2012) provides a typology of anatomical entities across species that is linked to cell types (via links to the cell ontology). Amongst others things, this allows linkage of anatomical structures across species, allowing inferences of homology and comparison of features such as gene and protein expression across species.

Another cross-species ontology, and one that complements work on anatomy, is described by Giudicelli and Lefranc (2012). They provide an update on the IMGT-Ontology which is an ontology of immunogenetics and immunoinformatics used in the international ImMunoGeneTics information system® (http://www.imgt.org). The IMGT-Ontology describes a range of immunogenetics concepts (immunoglobulins or antibodies, T cell receptors, major histocompatibility (MH) proteins of humans and other vertebrates, proteins of the immunoglobulin superfamily and MH superfamily, related proteins of the immune system of vertebrates and invertebrates, therapeutic monoclonal antibodies, fusion proteins for immune applications, and composite proteins for clinical applications).

A key problem for semantic biology is linking data on phenotypic measurements between model organisms, used to understand human disease, and clinical observations made in humans. This has been an active area of research in recent years (Hancock et al., 2009; Schofield et al., 2010). Shimoyama et al. (2012) make an important contribution to this area by describing a set of ontologies used to describe clinical measurements, measurement methods and experimental conditions for traits common to rat and man (and, by extension, in other mammalian model systems such as mouse and, potentially, more distantly related species). These measurements are similar to those used in large-scale phenotyping experiments (Hancock and Gates, 2011) so that this ontology system provides a potentially valuable mechanism for the study of genotype-phenotype relations in mammals.

Going beyond the underlying ontological structures used to describe biological data Imam et al. (2012) describe an integrated set of ontologies used within the Neuroscience Information Framework (www.neuinfo.org/), which describe major domains in neuroscience, including diseases, brain anatomy, cell types, sub-cellular anatomy, small molecules, techniques, and resource descriptors. This application provides a valuable insight into how sets of existing ontologies can be integrated with novel, more application-specific ontologies and structures to underpin a semantic-based knowledge system. NIF links logically consistent sets of terms into single structures but forms links between these logically consistent sets using bridging modules. Deb (2012) argues for an alternative approach using a single upper level (foundational) ontology to link specific biological domain ontologies.

A key issue that any such framework raises is how to compare and choose appropriate ontologies for any given system. A typical default position in biological applications is to accept the ontologies held in the open biological ontologies set (Smith et al., 2007). Here Klie and Nikoloski (2012) argue that ontology choice is to a degree application-specific and that domain-specific ontologies may in some cases be more useful than general ontologies such as the GO.

The major purpose of developing biological ontologies (rather than simpler controlled vocabularies) is to make use of the relations implicit in ontologies to facilitate analysis and annotation. These topics are addressed by two papers in this series. Ross et al. (2013) describe the use of the PRotein Ontology to carry out cross-species comparisons of function in the spindle checkpoint pathway. Bastos et al. (2013) consider the use of subsets of functionally coherent proteins to improve functional annotation in a protein family.

Finally, advances in technology provide new opportunities for the use of semantically-enriched data in applications that are only minimally ontology-aware. Dönitz and Wingender (2012) describe a web-based service that can be accessed from any application to make use of standard ontologies, removing a significant burden to application development. At a higher level, Deb and Srirama (2013) provide us with a view of how the data and ontologies currently being produced might be linked and accessed via cloud infrastructures and describe some of the problems this raises in the domain of human eHealth.

REFERENCES

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. The gene ontology consortium. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556

Bard, J. B. L. (2012). The AEO, an ontology of anatomical entities for classifying animal tissues and organs. *Front. Genet.* 3:18. doi: 10.3389/fgene.2012.00018

Bastos, H. P., Clarke, L. A., and Couto, F. M. (2013). Annotation extension through protein family annotation coherence metrics. *Front. Genet.* 4:201. doi: 10.3389/fgene.2013.00201

Deb, B. (2012). An ontological analysis of some biological ontologies. *Front. Genet.* 3:269. doi: 10.3389/fgene.2012.00269

Deb, B., and Srirama, S. N. (2013). Social networks for eHealth solutions on cloud. *Front. Genet.* 4:171. doi: 10.3389/fgene.2013.00171

Dönitz, J., and Wingender, E. (2012). The ontology-based answers (OBA) service: a connector for embedded usage of ontologies in applications. *Front. Genet.* 3:197. doi: 10.3389/fgene.2012.00197

Giudicelli, V., and Lefranc, M. P. (2012). Imgt-ontology 2012. *Front. Genet.* 3:79. doi: 10.3389/fgene.2012.00079

Hancock, J. M., and Gates, H. (2011). "The informatics of high-throughput mouse phenotyping: EUMODIC and beyond," in *Mouse as a Model Organism–From Animals to Cells*, eds *C. Brakebusch and T. Pihlajaniemi (Berlin: Springer)*, 77–88.

Hancock, J. M., Mallon, A. M., Beck, T., Gkoutos, G. V., Mungall, C., and Schofield, P. N. (2009). Mouse, man, and meaning: bridging the semantics of mouse phenotype and human disease. *Mamm. Genome* 20, 457–461. doi: 10.1007/s00335-009-9208-3

Imam, F. T., Larson, S. D., Bandrowski, A., Grethe, J. S., Gupta, A., and Martone, M. E. (2012). Development and use of ontologies inside the neuroscience information framework: a practical approach. *Front. Genet.* 3:111. doi: 10.3389/fgene.2012.00111

Klie, S., and Nikoloski, Z. (2012). The choice between mapman and gene ontology for automated gene function prediction in plant science. *Front. Genet.* 3:115. doi: 10.3389/fgene.2012.00115

Ross, K. E., Arighi, C. N., Ren, J., Natale, D. A., Huang, H., and Wu, C. H. (2013). Use of the protein ontology for multi-faceted analysis of biological processes: a case study of the spindle checkpoint. *Front. Genet.* 4:62. doi: 10.3389/fgene.2013.00062

Schofield, P. N., Gkoutos, G. V., Gruenberger, M., Sundberg, J. P., and Hancock, J. M. (2010). Phenotype ontologies for mouse and man; bridging the semantic gap. *Dis. Model. Mech.* 3, 281–289. doi: 10.1242/dmm.002790

Shimoyama, M., Nigam, R., McIntosh, L. S., Nagarajan, R., Rice, T., Rao, D. C., et al. (2012). Three ontologies to define phenotype measurement data. *Front. Genet.* 3:87. doi: 10.3389/fgene.2012.00087

Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., et al. (2007). The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat. Biotechnol.* 25, 1251–1255. doi: 10.1038/nbt1346

Received: 09 January 2014; accepted: 21 January 2014; published online: 04 February 2014.

Citation: Hancock JM (2014) Editorial: biological ontologies and semantic biology. Front. Genet. ***5****:18. doi: 10.3389/fgene.2014.00018*

This article was submitted to Bioinformatics and Computational Biology, a section of the journal Frontiers in Genetics.