

The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration

Barry Smith¹, Michael Ashburner², Cornelius Rosse³, Jonathan Bard⁴, William Bug⁵, Werner Ceusters⁶, Louis J Goldberg⁷, Karen Eilbeck⁸, Amelia Ireland⁹, Christopher J Mungall¹⁰, the OBI Consortium¹¹, Neocles Leontis¹², Philippe Rocca-Serra⁹, Alan Ruttenberg¹³, Susanna-Assunta Sansone⁹, Richard H Scheuermann¹⁴, Nigam Shah¹⁵, Patricia L Whetzel¹⁶ & Suzanna Lewis¹⁰

The value of any kind of data is greatly enhanced when it exists in a form that allows it to be integrated with other data. One approach to integration is through the annotation of multiple bodies of data using common controlled vocabularies or 'ontologies'. Unfortunately, the very success of this approach has led to a proliferation of ontologies, which itself creates obstacles to integration. The Open Biomedical Ontologies (OBO) consortium is pursuing a strategy to overcome this problem. Existing OBO ontologies, including the Gene Ontology, are undergoing coordinated reform, and new ontologies are being created on the basis of an evolving set of shared principles governing ontology development. The result is an expanding family of ontologies designed to be interoperable and logically well formed and to incorporate accurate representations of biological reality. We describe this OBO Foundry initiative and provide guidelines for those who might wish to become involved.

In the search for what is biologically and clinically significant in the swarms of data being generated by today's high-throughput technologies, a common strategy involves the creation and analysis of 'annotations' linking primary data to expressions in controlled, structured vocabularies, thereby making the data available to search and to algorithmic processing¹. The most successful such endeavor, measured both by numbers of users and by reach across species and granularities, is the Gene Ontology (GO)². There exist over 11 million annotations relating gene products described in the UniProt, Ensembl and other databases to terms in the GO³, of which half a million have been manually verified by specialist curators in different model-organism communities on the basis of the analysis of experimental results reported in 52,000 scientific journal articles (<http://www.ebi.ac.uk/GOA/>). Data related to some 180,000 genes have been manually annotated in this way, an endeavor now being refined and systematized within the Reference Genome Project (US National Institutes of Health National Human Genome Research Institute grant 2P41HG002273-07), which will provide comprehensive GO annotations for both the human genome and a representative set of model-organism genomes in support of research on the primary molecular systems affecting human health.

From retrospective mapping to prospective standardization

The domain of molecular biology is marked by the availability of large amounts of well defined data that can be used without restriction as inputs to algorithmic processing. In the clinical domain, by contrast, only limited amounts of data are available for research purposes, and these still consist overwhelmingly of natural language text. Even where more systematic clinical data are available, the use of local coding schemes means that these data do not cumulate in ways useful to research⁴. One approach to solving this problem is the Unified Medical Language System (UMLS)⁵, a compendium of some 100 source vocabularies combined through a process of retrospective mapping based on the identification of synonymy relations between constituent terms. The UMLS has yielded very useful results for applications such as indexing and retrieval of documents. But because the separate vocabularies have no common architecture^{6,7}, UMLS mappings do not meld their terms together into any single system⁸.

Increasingly, therefore, the need is being recognized for strategies of prospective standardization designed to bring about the progressive improvement and reciprocal alignment of the frameworks employed for the management, description and publication of biomedical data.

¹Department of Philosophy and New York State Center of Excellence in Bioinformatics and Life Sciences, University at Buffalo, 701 Ellicott Street, Buffalo, New York 14203, USA. ²Department of Genetics, University of Cambridge, Downing Street, Cambridge, CB2 3EH, UK. ³Department of Biological Structure, Box 357420, University of Washington, Seattle, Washington 98195, USA. ⁴Department of Biomedical Sciences, The University of Edinburgh, 1 George Square, Edinburgh EH8 9JZ, Scotland, UK. ⁵Department of Neurobiology and Anatomy, Drexel University College of Medicine, 2900 Queen Lane, Philadelphia, Pennsylvania 19129, USA. ⁶Department of Psychiatry and New York State Center of Excellence in Bioinformatics and Life Sciences, University at Buffalo, 701 Ellicott Street, Buffalo, New York 14203, USA. ⁷Department of Oral Biology and New York State Center of Excellence in Bioinformatics and Life Sciences, University at Buffalo, 701 Ellicott Street, Buffalo, New York 14203, USA. ⁸Eccles Institute of Human Genetics, University of Utah, 15 North 2030 East, Salt Lake City, Utah 84112, USA. ⁹European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, UK. ¹⁰Life Sciences Division, Lawrence Berkeley National Lab, 1 Cyclotron Road, Berkeley, California 94720, USA. ¹¹<http://obi.sourceforge.net/community/index.php>. ¹²Department of Chemistry, Bowling Green State University, 212 Physical Sciences Laboratory Building, 1001 East Wooster Street, Bowling Green, Ohio 43403, USA. ¹³Science Commons, c/o Massachusetts Institute of Technology Computer Science and Artificial Intelligence Laboratory, Building 32-386D, 32 Vassar Street, Cambridge, Massachusetts 02139, USA. ¹⁴Department of Pathology, University of Texas Southwestern Medical Center, Harry Hines Blvd., Dallas, Texas 75390 USA. ¹⁵Stanford Medical Informatics, Stanford University School of Medicine, 251 Campus Drive, Stanford, California 94305, USA. ¹⁶Center for Bioinformatics and Department of Genetics, University of Pennsylvania School of Medicine, 423 Guardian Drive, Philadelphia, Pennsylvania 19104, USA. Correspondence should be addressed to B.S. (phsmith@buffalo.edu).

Table 1 Coverage of initial Foundry ontologies

Granularity	Continuant			Occurrent
	Independent	Dependent		
Organ and organism	Organism (NCBI taxonomy or similar)	Anatomical entity (FMA, CARO)	Organ function (Physiology ontology, to be determined)	Organism-level process (GO)
Cell and cellular component	Cell (CL, FMA)	Cellular component (FMA,GO)	Cellular function (GO)	Cellular process (GO)
Molecule	Molecule (ChEBI, SO, RnaO, PRO)		Molecular function (GO)	Molecular process (GO)

Down the left column is the granularities (spatial scales) of the entities represented in the ontologies; along the top is a division corresponding to the ways these entities exist in time⁴⁷. 'Continuants' endure through time. 'Occurrents' (processes) unfold through time in successive stages. Continuants are divided into physical things, on the one hand, and qualities and functions, on the other. The latter are dependent continuants: a quality such as the shape of a fly's wing depends for its existence on, and endures through time in tandem with, the wing that is its bearer; a function, such as the function of an enzyme to catalyze reactions of a certain type, similarly endures through time in tandem with the enzyme itself and exists even when it is not being exercised in any instance of that reaction. NCBI, US National Center for Biotechnology Information; CL, Cell Ontology; SO, Sequence Ontology; RnaO, RNA Ontology; PRO, Protein Ontology.

Two conspicuous products of this trend are the US National Cancer Institute's Cancer Biomedical Informatics Grid (caBIG) project⁹ and HL7's Reference Information Model (RIM) (<http://hl7.org>). caBIG seeks to integrate all cancer research data in a common cyberinfrastructure by standardizing the ways in which such data are acquired, formatted, processed and stored. The HL7 RIM, similarly, offers a standard for the exchange, management and integration of all information relevant to healthcare, from clinical genomics to hospital billing. However, because both caBIG and HL7 focus on the meta-level question of how data and information should be represented in computer and messaging systems, it can be argued that they fail to do justice to the object-level question of how best to represent the proteins, organisms, diseases or drug interactions that are of primary interest in biomedical research^{7,10}.

A collaborative experiment in ontology development

In 2001, Ashburner and Lewis initiated a strategy to address this object-level question by creating OBO, an umbrella body for the developers of life-science ontologies. OBO applies the key principles underlying the success of the GO, namely, that ontologies be open, orthogonal, instantiated in a well-specified syntax and designed to share a common space of identifiers¹¹. Ontologies must be open in the sense that they and the bodies of data described in their terms should be available for use without any constraint or license and so be applicable to new purposes without restriction. They are also receptive to modification as a result of community debate. They must be orthogonal to ensure additivity of annotations and to bring the benefits of modular development. They must be syntactically in good order to support algorithmic processing. And they must employ a common system of identifiers to enable backward compatibility with legacy annotations as the ontologies evolve.

OBO now comprises over 60 ontologies, and its role as an ontology information resource is supported by the NIH Roadmap National Center for Biomedical Ontology (NCBO) through its BioPortal¹². At the same time, the developers of a subset of OBO ontologies have initiated the OBO Foundry, a collaborative experiment based on the voluntary acceptance by its participants of an evolving set of principles (available at <http://obofoundry.org>) that extend those of the original OBO by requiring in addition that ontologies (i) be developed in a collaborative effort, (ii) use common relations that are unambiguously defined, (iii) provide procedures for user feedback and for identifying successive versions and (iv) have a clearly bounded subject-matter (so that an ontology devoted to cell components, for example, should not include terms like 'database' or 'integer'). A graphical representation of the coverage of the initial Foundry ontologies is provided in **Table 1**.

Progress thus far

Since the OBO Foundry was established, ontologies such as the GO and the Foundational Model of Anatomy (FMA)¹³ have been reformed and new ontologies created on the basis of its principles^{14–16}. Perhaps most importantly, ontologies have been laid to rest. Before the OBO Foundry there existed at least four cell-type ontologies: one from Bard, Rhee and Ashburner¹⁷, another from Kelso *et al.*¹⁸, a third implicit within the GO and the fourth a subontology within the FMA. The first three now form a single cell-type ontology (CL)¹⁹, which is itself being integrated with the cell-type representations contained within the FMA.

The Foundry initiative also serves to align ontology development efforts carried out by separate communities, for example in research on different model organisms. The potential of such research to yield results valuable for the understanding of human disease rests on our ability to make reliable cross-species comparisons. Because so much model-organism data is localized to anatomical structures, drawing inferences on the basis of such comparisons has been hampered by the lack of coordination in anatomy ontology development among different communities. Some ontologies represent structure, others represent function, yet others represent stages of development, and some draw on combinations of these, in ways that close off opportunities for automatic reasoning. The Foundry has created a roadmap for the incremental resolution of this problem through the initiation of the Common Anatomy Reference Ontology (CARO)¹⁴, which is providing guidelines both for model-organism communities with legacy anatomy ontologies who wish to initiate reforms in the direction of compatibility and for communities who wish to build new ontologies from scratch. CARO is based on the top-level types of the FMA and is serving as a template for the creation of the Fish Multi-Species, Ixodidae and Argasidae (tick), mosquito and *Xenopus* anatomy ontologies, and also as basis for reforms of the *Drosophila* and zebrafish anatomy ontologies¹⁹.

The Ontology for Biomedical Investigations (OBI) addresses the need for controlled vocabularies to support integration of experimental data, a need originally identified in the transcriptomics domain by the Microarray Gene Expression Data Society (MGED), which developed the MGED Ontology²⁰ as an annotation resource for microarray data. In response to the recognition of convergent needs in areas such as protein and metabolite characterization, this effort was broadened to become what was initially known as FuGO (Functional Genomics Investigation Ontology)²¹. FuGO was further expanded in 2006 to include clinical and epidemiological research, biomedical imaging and a variety of further experimentation domains to become what is today OBI, an ontology designed to serve the coordinated representation of designs, protocols, instrumentation, materials, processes, data

and types of analysis in all areas of biological and biomedical investigation. Twenty-five groups are now involved in building OBI (<http://obisf.net/community>), and the Foundry discipline has proven essential to its distributed development.

Unlike most OBO ontologies, which use the OBO file format and the associated OBO-Edit software favored by model-organism and other biologist communities, OBI uses the OWL-DL Web Ontology Language. The need to make OWL and OBO ontologies interoperable has sparked the creation of bidirectional OBO–OWL conversion tools²² that integrate data annotated in terms of the GO and other OBO ontologies with the bodies of data coming onstream within the framework of the Semantic Web²³ an influential initiative to exploit OWL ontologies to encode knowledge in distributed computer systems²⁴.

Models of good practice

Each Foundry ontology forms a graph-theoretic structure, with terms connected by edges representing relations such as ‘is_a’ or ‘part_of’ in assertions such as ‘serotonin is_a biogenic amine’ or ‘cytokinesis part_of cell proliferation’. Because relations in OBO ontologies were initially used in inconsistent ways²⁵, the OBO Relation Ontology (RO)²⁶ was developed to provide guidelines to ontology builders in the consistent formulation of relational assertions. These guidelines are already proving useful—for example, in the representation of anatomical change²⁷ and in linking diverse image collections to phylogenetic datasets²⁸.

Other areas in which the Foundry is providing guidelines include naming conventions²⁹ and pathway representations³⁰. The model of good practice in the formulation of definitions is the FMA¹³, a representation of types of anatomical entities built around two backbone hierarchies of ‘is_a’ and ‘part_of’ relations. The FMA imposes a rule whereby all definitions take the genus-species form:

an A = def. a B that C’s where B is the ‘is_a’ parent of A, and C are the differentia marking out that subfamily of Bs which are also As. For example,

cell = def. an anatomical structure that has as its boundary the external surface of a maximally connected plasma membrane

plasma membrane = def. a cell component that has as its parts a maximal phospholipids bilayer in which instances of two or more types of protein are embedded.

Anchoring definitions in the ‘is_a’ hierarchy in this way diminishes the role of opinion in determining where terms should be placed in the hierarchy, thereby fostering consistency both within and between ontologies and helping to prevent common errors^{6,7,26}.

To maximize cross-ontology coordination, compound terms should be built as far as possible out of constituent terms drawn from Foundry ontologies linked using relational expressions from the RO³¹. This methodology of cross-products is being applied, in one of the biological projects driving the NCBO, to the annotation of *Drosophila*, zebrafish and human alleles for genes implicated in disease^{12,32}. Specialist curators associate these alleles with phenotype descriptions formulated using terms drawn from more than one OBO Foundry ontology—for example, composing the Phenotypic Quality Ontology (PATO) term ‘increased concentration’ with the FMA term ‘blood’ and the ChEBI term ‘glucose’ to represent increased blood glucose phenotypes. Such creation of terms through explicit composition avoids the bottlenecks created where, as for example in the Mammalian Phenotype Ontology, each new term must be approved for inclusion in the ontology before it can be used in annotations. But the approach will work only if the resultant terms are unambiguous, and here the Foundry helps provide the necessary rigor. The orthogonality principle helps to reduce the need for arbitrary decisions between equivalent-seeming terms drawn from different ontologies, the PATO phenotypic-quality ontology provides templates

for term formation, and the RO provides formally coherent glue for combination³³.

The current scope of the OBO Foundry initiative is summarized in **Table 2**. Foundry ontologies are created and maintained by biologists with a thorough knowledge of the underlying science. Where domain experts jointly control ontology, data, and annotations (as in the case of the GO/Uniprot collaboration), all three can be curated in tandem in a way that provides a reality check at each stage of the process³⁴. As results of experiments are described in annotations, this leads to extensions or corrections of the ontology, which in turn lead to better annotation³⁵. The results of the Foundry’s work can then be applied by external groups as benchmarks—for example, to help identify genes mutated at significant frequencies in human cancers³⁶ or to identify cellular components involved in antigen processing³⁷ or, in general, to refine otherwise noisy results of text- and data-mining^{38–41}.

The OBO Foundry applied

Neurophysiology. A demonstration of the utility of the Foundry methodology is provided by ongoing work to create the NeuronDB database within the Senselab project (<http://senselab.med.yale.edu/>). NeuronDB encompasses three types of neuronal property: voltage-gated conductances, neurotransmitters and neurotransmitter receptors. An initial representation of neurotransmitters defined an ‘is_a’ hierarchy with classes such as ‘neurotransmitter receptor’ and subclasses such as ‘GABA receptor’. In this initial ontology, receptors were not defined, and strictly speaking one would not have known, for example, whether a receptor was a protein or a protein complex. The Foundry provided a set of principles and at least one task that may be evaluated in making such choices: namely, the scope of each ontology should be clearly bounded and (by orthogonality) no term should appear in more than one ontology. Reviewing the existing ontologies, we found that the GO Molecular Function (GO MF) ontology already had classes such as ‘receptor activity’ (GO:0004872) and a number of subclasses that described receptor activities that were referred to in NeuronDB.

We reviewed one hundred thirty resultant receptor classes. Where they existed, we reused MF classes; where they did not, we created subclasses of existing MF classes and submitted the results to GO for future inclusion. Arranging NeuronDB to interoperate transparently with GO provided the further benefit that we can now take advantage of GO annotations to find the proteins that correspond to the receptor classes by searching annotations to the MF terms. This is a model for how small ontology builders can constructively contribute to the growth of shared resources while simultaneously benefiting users of their own ontologies.

Neuroanatomy. In support of research on neurodegenerative and neurological disease within the Biomedical Informatics Research Network (BIRN)⁴², the BIRN Ontology Task Force is applying the Foundry principles to formally represent several large domains, including (i) neuroanatomy⁴³, where annotations must capture not only the structural systems of parthood and topological connection but also cytoarchitectural parcellations such as the CA1, CA2 and CA3 regions of the hippocampus, (ii) functional systems, such as the basal ganglion circuits for motor planning and motor memory and (iii) neurochemistry (for example, of brainstem monoamine nuclei). The members of the BIRN Ontology Task Force see the Foundry as providing a framework within which these distinct axes can be algorithmically combined, and they are incorporating the results into BIRN’s neuro-image atlasing project and using them to integrate spatially mapped microarray expression data with mouse imaging results.

Table 2 OBO Foundry ontologies (as of April 2007)

Ontology	Scope	URL	Custodians
Mature ontologies undergoing incremental reform			
Cell Ontology (CL)	Cell types from prokaryotic to mammalian	http://obofoundry.org/cgi-bin/detail.cgi?cell	Michael Ashburner, Jonathan Bard, Oliver Hofmann, Sue Rhee
Gene Ontology (GO)	Attributes of gene products in all organisms	http://www.geneontology.org	Gene Ontology Consortium
Foundational Model of Anatomy (FMA)	Structure of the mammalian and in particular the human body	http://fma.biostr.washington.edu	J.L.V. Mejino, Jr., Cornelius Rosse
Zebrafish Anatomical Ontology (ZAO)	Anatomical structures in <i>Danio rerio</i>	http://zfin.org/zf_info/anatomy/dict/sum.html	Melissa Haendel, Monte Westerfield
Mature ontologies still in need of thorough review			
Chemical Entities of Biological Interest (ChEBI)	Molecular entities which are products of nature or synthetic products used to intervene in the processes of living organisms	http://www.ebi.ac.uk/chebi	Paula Dematos, Rafael Alcantara
Disease Ontology (DO)	Types of human disease	http://diseaseontology.sf.net	Rex Chisholm
Plant Ontology (PO)	Flowering plant structure, growth and development stages	http://plantontology.org	Plant Ontology Consortium
Sequence Ontology (SO)	Features and properties of nucleic acid sequences	http://www.sequenceontology.org	Karen Eilbeck
Ontologies for which early versions exist			
Ontology for Clinical Investigations (OCI)	Clinical trials and related clinical studies	http://www.bioontology.org/wiki/index.php/CTO:Main_Page	OCI Working Group
Common Anatomy Reference Ontology (CARO)	Anatomical structures in all organisms	http://obofoundry.org/cgi-bin/detail.cgi?caro	Fabian Neuhaus, Melissa Haendel, David Sutherland
Environment Ontology	Habitats and associated spatial regions and sites	http://www.obofoundry.org/cgi-bin/detail.cgi?id=envo	Norman Morrison, Dawn Field
Ontology for Biomedical Investigations (OBI)	Design, protocol, instrumentation and analysis applied in biomedical investigations	http://obi.sf.net	OBI Working Group
Phenotypic Quality Ontology (PATO)	Qualities of biomedical entities	http://www.phenotypeontology.org	Michael Ashburner, Suzanna Lewis, Georgios Gkoutos
Protein Ontology (PRO)	Protein types and modifications classified on the basis of evolutionary relationships	http://pir.georgetown.edu/pro	Protein Ontology Consortium
Relation Ontology (RO)	Relations in biomedical ontologies	http://obofoundry.org/ro	Barry Smith, Chris Mungall
RNA Ontology (RnaO)	RNA three-dimensional structures, sequence alignments, and interactions	http://roc.bgsu.edu/	RNA Ontology Consortium

The Minimum Information for Biological and Biomedical Investigations (MIBBI). This initiative represents the first new standards effort that takes OBO and the OBO Foundry as its role model⁴⁴. MIBBI provides information resources to promote the consolidation of the many prescriptive checklists that specify core metadata items to be included when reporting results in a variety of experimentation domains⁴⁵. The proliferation of such 'minimum information' checklists has made it increasingly difficult to obtain an overview of existing specifications, unnecessarily duplicating efforts and creating problems when third parties try to use described information. The MIBBI Portal operates analogously to OBO and the NBCO Biportal as an open information resource for all initiatives addressing these problems; the MIBBI Foundry fosters collaborative development and integration of checklists into orthogonal modules⁴⁶.

How to join

Like OBO, the OBO Foundry is an open community. Any individual or group working in the domain of biomedicine wishing to join the initiative is encouraged to do so, and all discussion forums (listed at <http://obofoundry.org>) are open to all interested parties without restriction. The recommended first step is to join one or more mailing lists in salient areas as a way to become familiar with the Foundry's collaborative methodology and identify members with overlapping

expertise. Those with new ontology resources are invited to submit them for informal consideration by existing members; this will be followed by a period in which compliance with the Foundry principles is addressed, especially as concerns potential conflicts in areas of overlap. Membership in the Foundry initiative then flows from a commitment to incremental implementation of these principles as they evolve over time, with the Foundry coordinators (currently Ashburner, Lewis, Mungall and Smith) serving as analogs of journal editors, whereby the division of labor that results from orthogonality helps ensure that development decisions are made by the authors of single ontologies. By joining the initiative, the authors of an ontology commit to working with other members to ensure that, for any particular domain, there is convergence on a single ontology. Criticism, too, is welcomed: the Foundry is an attempt to apply the scientific method to the task of ontology development, and thus it accepts that no resource will ever exist in a form that cannot be further improved.

Our long-term goal is that the data generated through biomedical research should form a single, consistent, cumulatively expanding and algorithmically tractable whole. Our efforts to realize this goal, which are still very much in the proving stage, reflect an attempt to walk the line between the flexibility that is indispensable to scientific advance and the institution of principles that is indispensable to successful coordination.

ACKNOWLEDGMENTS

The Foundry is receiving *ad hoc* funding under the BISC, Gene Ontology Consortium, MGED, NCBO and RNA Ontology grants. We are grateful to all of these sources, and also to the ACGT Project of the European Union and to the Humboldt and Volkswagen Foundations.

Published online at <http://www.nature.com/naturebiotechnology>
Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions/>

1. Yue, L. & Reisdorf, W.C. Pathway and ontology analysis: emerging approaches connecting transcriptome data and clinical endpoints. *Curr. Mol. Med.* **5**, 11–21 (2005).
2. Gene Ontology Consortium. The Gene Ontology (GO) project in 2006. *Nucleic Acids Res.* **34** (database issue), D322–D326 (2006).
3. Camon, E. *et al.* The Gene Ontology Annotation (GOA) Project. *Genome Res.* **13**, 662–672 (2003).
4. Kohane, I.S. *et al.* Building national electronic medical record systems via the World Wide Web. *J. Am. Med. Inform. Assoc.* **3**, 191–207 (1996).
5. Bodenreider, O. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Res.* **32** (database issue), D267–D270 (2004).
6. Ceusters, W., Smith, B., Kumar, A. & Dhaen, C. Mistakes in medical ontologies: where do they come from and how can they be detected? *Stud. Health Technol. Inform.* **102**, 145–164 (2004).
7. Ceusters, W., Smith, B. & Goldberg, L. A terminological and ontological analysis of the NCI Thesaurus. *Methods Inf. Med.* **44**, 498–507 (2005).
8. Campbell, K.E., Oliver, D.E. & Shortliffe, E.H. The Unified Medical Language System. Toward a collaborative approach for solving terminologic problems. *J. Am. Med. Inform. Assoc.* **5**, 12–16 (1998).
9. Buetow, K.H. Cyberinfrastructure: empowering a 'third way' in biomedical research. *Science* **308**, 821–824 (2005).
10. Smith, B. & Ceusters, W. HL7 RIM: an incoherent standard. *Stud. Health Technol. Inform.* **124**, 133–138 (2006).
11. Ashburner, M., Mungall, C.J. & Lewis, S.E. Ontologies for biologists: a community model for the annotation of genomic data. *Cold Spring Harb. Symp. Quant. Biol.* **68**, 227–236 (2003).
12. Rubin, D.L. *et al.* National Center for Biomedical Ontology: advancing biomedicine through structured organization of scientific knowledge. *OMICS* **10**, 185–198 (2006).
13. Rosse, C. & Mejino, J.L.F. The Foundational Model of Anatomy ontology. In *Anatomy Ontologies for Bioinformatics* (eds. Burger, A. *et al.*) (Springer, New York, in the press).
14. Haendel, M. *et al.* CARO: the Common Anatomy Reference Ontology. In *Anatomy Ontologies for Bioinformatics* (eds. Burger, A. *et al.*) (Springer, New York, in the press).
15. Leontis, N.B. *et al.* The RNA Ontology Consortium: an open invitation to the RNA community. *RNA* **12**, 533–541 (2006).
16. Natale, D.A. *et al.* Framework for a protein ontology. *BMC Bioinformatics [online]* (in the press).
17. Bard, J., Rhee, S.Y. & Ashburner, M. An ontology for cell types. *Genome Biol. [online]* **6**, R21 (2005).
18. Kelso, J. *et al.* eVOC: a controlled vocabulary for unifying gene expression data. *Genome Res.* **13**, 1222–1230 (2003).
19. Mabee, P.M. *et al.* Phenotype ontologies: the bridge between genomics and evolution. *Trends Ecol. Evol.* **22**, 345–350 (2007).
20. Whetzel, P.L. *et al.* The MGED Ontology: a resource for semantics-based description of microarray experiments. *Bioinformatics* **22**, 866–873 (2006).
21. Whetzel, P.L. *et al.* Development of FuGO: an ontology for functional genomics investigations. *OMICS* **10**, 199–204 (2006).
22. Golbreic, C. *et al.* OBO and OWL: leveraging semantic web technologies for the life sciences. In *Proceedings 6th International Semantic Web Conference (ISWC 2007)*, (Springer, in the press).
23. Brinkley, J.F., Detwiler, L.T., Gennari, J.H., Rosse, C. & Suci, D. A framework for using reference ontologies as a foundation for the semantic web. *Proc. AMIA Fall Symposium*, 2006, 95–100.
24. Lacy, L.W. *Owl: Representing Information Using the Web Ontology Language* (Trafford Publishing, Victoria, BC, Canada, 2005).
25. Smith, B., Köhler, J. & Kumar, A. On the application of formal principles to life science data: a case study in the Gene Ontology. *Data Integration in the Life Sciences (DILS) Workshop 2004*, 79–94.
26. Smith, B. *et al.* Relations in biomedical ontologies. *Genome Biol. [online]* **6**, R46 (2005).
27. Bittner, T. & Goldberg, L.J. Spatial location and its relevance for terminological inferences in bio-ontologies. *BMC Bioinformatics* **23**, 1674–1682 (2007).
28. Ramirez, M.J. *et al.* Linking of digital images to phylogenetic data matrices using a morphological ontology. *Syst. Biol.* **56**, 283–294 (2007).
29. Schober, D., *et al.* Towards naming conventions for use in controlled vocabulary and ontology engineering. *Bio-Ontologies Workshop*, ISMB/ECCB, Vienna, 20 July 2007, 87–90.
30. Ruttenberg, A., Rees, J., & Zucker, J. What BioPAX communicates and how to extend OWL to help it. *OWL: Experiences and Directions Workshop Series* <http://owl-workshop.man.ac.uk/acceptedLong/submission_26.pdf> (2006).
31. Hunter, L. & Bada. M. Enrichment of OBO ontologies. *J. Biomed. Inform.* **40**, 300–315 (2007).
32. Hill, D.P., Blake, J.A., Richardson, J.E. & Ringwald, M. Extension and integration of the Gene Ontology (GO): combining GO vocabularies with external vocabularies. *Genome Res.* **12**, 1982–1991 (2002).
33. Mungall, C.J. Obo: integrating language and meaning in bio-ontologies. *Comp. Funct. Genomics* **5**, 509–520 (2004).
34. Camon, E. *et al.* The Gene Ontology Annotation (GOA) Database: sharing knowledge in Uniprot with Gene Ontology. *Nucleic Acids Res.* **32** (database issue), D262–D266 (2004).
35. Blake, J., Hill, D.P. & Smith, B. Gene Ontology annotations: what they mean and where they come from. *Bio-Ontologies Workshop*, ISMB/ECCB, Vienna, 20 July 2007, 79–82.
36. Sjoblom, T. *et al.* The consensus coding sequences of human breast and colorectal cancers. *Science* **314**, 268–274 (2006).
37. Lee, J.A. *et al.* Components of the antigen processing and presentation pathway revealed by gene expression microarray analysis following B cell antigen receptor (BCR) stimulation. *BMC Bioinformatics [online]* **7**, 237 (2006).
38. Rebholz-Schuhmann, D., Kirsch, H. & Couto, F. Facts from text—is text mining ready to deliver? *PLoS Biol. [online]* **3**, e65 (2005).
39. Witte, R., Kappler, T. & Baker, C.J.O. Ontology design for biomedical text mining. In *Semantic Web: Revolutionizing Knowledge Discovery in the Life Sciences* (eds. Baker C.J.O. & Cheung, K.-H.) 281–313 (Springer, New York, 2007).
40. Zhang, S. & Bodenreider, O. Aligning multiple anatomical ontologies through a reference. *International Workshop on Ontology Matching (OM 2006)* 193–197 (2006).
41. Luo, F. *et al.* Modular organization of protein interaction networks. *Bioinformatics* **23**, 207–214 (2007).
42. Martone, M.E., Gupta, A. & Ellisman, M.H. E-neuroscience: challenges and triumphs in integrating distributed data from molecules to brains. *Nat. Neurosci.* **7**, 467–472 (2004).
43. Fong, L. *et al.* An ontology-driven knowledge environment for subcellular neuroanatomy. *OWL Experiences and Directions, 3rd International Workshop*, Innsbruck, Austria, June 6–7, 2007 (in the press).
44. Taylor, C.F. *et al.* Promoting coherent minimum reporting requirements for biological and biomedical investigations: the MIBBI Project. *Nat. Biotechnol.* (in the press).
45. Brazma, A. *et al.* Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. *Nat. Genet.* **29**, 365–371 (2001).
46. Sansone, S.A. *et al.* A strategy capitalizing on synergies: the Reporting Structure for Biological Investigation (RSBI) working group. *OMICS* **10**, 164–171 (2006).
47. Grenon, P., Smith, B. & Goldberg, L. Biodynamic ontology: applying BFO in the biomedical domain. In *Ontologies in Medicine* (ed. Pisanelli, D.M.) 20–38 (IOS, Amsterdam, 2004).