The OBO Foundry: Coordinated Evolution of Ontologies to Support Biomedical Data Integration

Barry Smith,[a] Michael Ashburner,[b] Cornelius Rosse,[c] Jonathan Bard,[d] William Bug,[e] Werner Ceusters,[f] Louis J. Goldberg,[g] Karen Eilbeck,[h] Amelia Ireland,[i] Christopher J Mungall,[j] the OBI Consortium,[k] Neocles Leontis,[l] Philippe Rocca-Serra,[i] Alan Ruttenberg[m], Susanna-Assunta Sansone,[i] Nigam Shah,[n] Patricia L. Whetzel,[o] Suzanna Lewis[j]

forthcoming in *Nature Biotechnology*

Abstract

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 has set in train a strategy to overcome this problem. Existing OBO ontologies, including the Gene Ontology, are undergoing a process of coordinated reform and new ontologies 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, logically well-formed, and to incorporate accurate representations of biological reality. We describe the OBO Foundry initiative, and provide guidelines for those who might wish to become involved.

[a] Department of Philosophy and New York State Center of Excellence in Bioinformatics and Life Sciences, University at Buffalo, Buffalo, NY 14260, USA (Corresponding author). Please address all correspondence to: Barry Smith <phismith@buffalo.edu>.

[b] Department of Genetics, University of Cambridge, Downing Street, Cambridge, CB2 3EH, UK

[c] Department of Biological Structure, Box 357420, University of Washington, Seattle, WA 98195, USA

[d] Department of Biomedical Sciences, The University of Edinburgh, Edinburgh EH8 9XD, Scotland, UK

[e] Department of Neurobiology and Anatomy, Drexel University College of Medicine, Philadelphia, PA 19129, USA

[f] Department of Psychiatry and New York State Center of Excellence in Bioinformatics and Life Sciences, University at Buffalo, Buffalo, NY 14260, USA

[g] Department of Oral Biology and New York State Center of Excellence in Bioinformatics and Life Sciences, University at Buffalo, Buffalo, NY 14260, USA

[h] Department of Human Genetics, University of Utah, Salt Lake City, UT 84112, USA

[i] European Bioinformatics Institute, Hinxton, Cambridge, CB10 1SD, UK

[j] Life Sciences Division, Lawrence Berkeley National Lab, 1 Cyclotron Road, Berkeley, CA 94720, USA

[k] http://obi.sf.net/

[l] Department of Chemistry, Bowling Green State University, Bowling Green, OH 43403, USA

[m] Science Commons, c/o MIT CSAIL, Building 32-386D, Cambridge, MA 02139 USA

[n] Stanford University School of Medicine, Medical Informatics, Stanford, CA 94305, USA

[o] Center for Bioinformatics and Department of Genetics, University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA

The problem of data integration

In the hunt for what is biologically and clinically significant in the swarms of data being unleashed by today's high-throughput technologies, a standard 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 processing.[1]

The world's most successful such endeavor, when measured in terms of both numbers of users and of 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 GO,[3] 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 (NIH/NHGRI 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 linear text. Even where clinical data are available in a more systematically composed form, there is a long-standing complaint to the effect that the use of local coding schemes means that these data do not cumulate in ways useful to research.[4] One approach to solving this problem is the Unified Medical Language System,[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 results tremendously useful for applications such as indexing and retrieval of documents. But because the separate vocabularies have no common architecture,[6,7] its mappings do not meld their terms together into any single system,[8] and we know of no strategy for coordinated improvement that would foster more coherent integration.

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. Two conspicuous products of this trend are the National Cancer Institute's Cancer Biomedical Informatics Grid (caBIG) project[9] and HL7's Reference Information Model (http://hl7.org). caBIG has the goal of creating a path to integration for all cancer research data through a common cyberinfrastructure focused on regimentation of the ways biomedical 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 are failing 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 to biomedical research.[7,10]

A collaborative experiment in science-based ontology development

In 2001 Ashburner and Lewis initiated a strategy to address this object-level question by creating OBO (Open Biomedical Ontologies), an umbrella body for the developers of life science ontologies applying the key principles underlying the success of the GO, namely that ontologies be *open*, *orthogonal*, *instantiated in a well-specified syntax*, and such as to *share a common space of identifiers*.[11] Ontologies must be *open* in order that they and the bodies of data described in their terms can be applied to new purposes without restriction. They must be *orthogonal* to ensure additivity of annotations and to bring 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 ensure backwards compatibility with legacy annotations as the ontologies evolve.

OBO now comprises over 60 ontologies, and its role as ontology information resource is supported by the National Center for Biomedical Ontology (NCBO) through its BioPortal.[12] 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 here: http://obofoundry.org) extending those of the original OBO by requiring in addition that ontologies be *developed in a collaborative effort*; *use common relations which are unambiguously defined*; *provide procedures for user feedback and for identifying successive versions*; and have a *clearly bounded subject-matter* (so that an ontology devoted to, say, cell components should not include terms like 'database' or 'integer').

<table>
<tr><td>RELATION TO TIME

GRANULARITY</td><td colspan="3">CONTINUANT</td><td rowspan="2">OCCURRENT</td></tr>
<tr><td colspan="2">INDEPENDENT</td><td colspan="2">DEPENDENT</td></tr>
<tr><td>ORGAN AND ORGANISM</td><td>Organism (NCBI Taxonomy or similar)</td><td>Anatomical Entity (FMA, CARO)</td><td>Organ Function (placeholder for physiology ontology)</td><td rowspan="3">Phenotypic Quality (PaTO)</td><td>Organism-Level Process (GO)</td></tr>
<tr><td>CELL AND CELLULAR COMPONENT</td><td>Cell (CL, FMA)</td><td>Cellular Component (FMA,GO)</td><td>Cellular Function (GO)</td><td>Cellular Process (GO)</td></tr>
<tr><td>MOLECULE</td><td colspan="2">Molecule (ChEBI, SO, RnaO, PRO)</td><td>Molecular Function (GO)</td><td>Molecular Process (GO)</td></tr>
</table>

Table 1: Coverage of initial Foundry Ontologies

A graphical representation of the coverage of the initial Foundry ontologies is provided in Table 1. Down the left are the levels of granularity of the entities represented in the ontologies; along the top is a division corresponding to the ways these entities exist in time.[13] On the one hand are *continuants*, which *endure* through time. On the other hand are *occurrents* (processes), which *unfold* through time in successive stages. Continuants are divided in turn 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 which 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.

ONTOLOGY	SCOPE	URL	CUSTODIANS
Mature ontologies undergoing incremental reform			
Cell Ontology (CL)	cell types from prokaryotes to mammals	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	JLV Mejino Jr., Cornelius Rosse
Zebrafish Anatomical Ontology (ZAO)	anatomical structures in *D. rerio*	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://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
Early versions exist			
Ontology for Clinical Investigations (OCI)	clinical trials and related clinical studies	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://envoc.org	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 3D structures, sequence alignments, and interactions	http://roc.bgsu.edu/	RNA Ontology Consortium

Table 2: OBO Foundry Ontologies (as of April 2007)

Progress thus far

Since the OBO Foundry was established, ontologies such as the GO and the Foundational Model of Anatomy (FMA)[14] have been reformed and new ontologies created on the basis of its principles[15,16,17] (Table 2). Perhaps most importantly, ontologies have been laid to rest. Prior to the OBO Foundry there existed at least four cell type ontologies, one from Bard, Rhee and Ashburner[18], another from Kelso, *et al.*[19], a third implicitly stated within the GO, and the fourth a sub-ontology within the FMA. The first three now form a single cell type ontology (CL),[18] which is itself being subjected to a process of integration with the cell type representations contained within the FMA.

The Common Anatomy Reference Ontology (CARO). The potential of research on model organisms to yield results valuable for the understanding of human disease rests on our ability to make reliable cross-species comparisons. Since so much of model organism data is localized to anatomical structures, the drawing of inferences based on such comparisons has been hampered by the fact that different communities have developed their anatomy ontologies in uncoordinated fashion. Some represent structure, others function, yet others stages of development, and some draw on combinations of these, in ways which close off opportunities for automatic reasoning. A roadmap for the incremental resolution of this problem has been created through the institution of CARO,[15] which is providing guidelines both for those model organism communities with legacy anatomy ontologies who wish to initiate reforms in the direction of compatibility, and also for those who wish to build new ontologies from scratch. CARO is based on the top-level types of the FMA and is serving as 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 Ontologies.[20]

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, which developed the MGED Ontology[21] as 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).[22] FuGO was further expanded in 2006 to include clinical and epidemiological research, biomedical imaging and other domains to become what is today OBI, an ontology designed to serve the coordinated representation of design, protocols, instrumentation, materials, processes, data and types of analysis performed in all areas of biological and biomedical investigation. Twenty-five groups are now involved in building OBI (http://obi.sf.net/community), and the Foundry discipline has proved essential to its distributed development.

OBI stands out also because, where most OBO ontologies have been developed using the OBO file format and the associated OBO-Edit software favored by biologist communities, OBI employs the OWL-DL Web Ontology Language. The need to make OWL and OBO ontologies interoperable has sparked the creation of bi-directional OBO/OWL conversion tools (http://tinyurl.com/34ddl2) now serving to 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 Web.[23]

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 ways,[24] the OBO Relation Ontology (RO)[25] was developed to provide guidelines to ontology builders in the consistent formulation of relational assertions. These guidelines are already proving useful for example in ontology-based approaches to the representation of anatomical change[26] and in the creation of an efficient means for linking diverse image collections to phylogenetic data sets.[27]

In other areas, too, the Foundry is providing guidelines, for example in the field of *naming conventions*[28] and of *pathway representations*.[29] The model of good practice in the formulation of *definitions* is the FMA,[14] 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 which C's*,

where *B* is the *is_a* parent of *A*, and *C* the *differentia* marking out that subfamily of *B*s which are also *A*s, for example in:

> *cell* =def. an *anatomical structure* which has as its boundary the external surface of a maximally connected plasma membrane
>
> *plasma membrane* = def. a *cell component* which 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 serves to diminish 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.[7,7,24]

To maximize cross-ontology coordination, compound terms should be built as far as possible out of constituent terms drawn from Foundry ontologies linked via relational expressions from the RO. This *methodology of cross-products* is being applied in one of the driving biological projects of the NCBO to the annotation of drosophila, zebrafish and human alleles for genes implicated in disease.[12,30] Specialist curators associate these alleles with phenotype descriptions formulated using terms drawn from multiple OBO Foundry ontologies, for example composing the 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 combination.[31]

Foundry ontologies are created and maintained by biologist-curators with a thorough knowledge of the underlying science. Where, as in the case of the GO/Uniprot collaboration, the Foundry methodology is applied by domain experts who enjoy joint control of ontology, data and annotations, all three can be curated in tandem in a way which serves to provide a reality check at each stage of the process.[32] As results of experiments are described in annotations, this leads to extensions or corrections of the ontology, which in turn lead to better annotation.[33]

By providing a body of humanly validated assertions, the results of the Foundry's work can then be applied by external groups as benchmarks, for example, in helping to identify genes mutated at significant frequencies in human cancers,[34] or revealing cellular components involved in antigen processing,[35] or in general by allowing refinement of otherwise noisy results of text- and data-mining.[36,37,38,39]

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 comprehends three types of neuronal properties: voltage gated conductances, neurotransmitter receptors, and neurotransmitter substances. An initial representation of neurotransmitters defined an *is_a* hierarchy with *neurotransmitter* as root and subclasses such as *Gaba receptor*. In this initial ontology, receptors were not defined, and strictly speaking one would not know, for example, whether a receptor was in fact a protein or a protein complex. The Foundry provides a set of principles and at least one task that may be evaluated in the making of such choices, namely that 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 find 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.[40]

130 resultant receptor classes were reviewed. Where they existed, MF classes were reused; where not, existing MF classes were subclassed and the results submitted to GO for future inclusion. Arranging NeuronDB to be able to interoperate transparently with GO provides the further benefit that we can now take advantage of GO Annotations to find the proteins that correspond to the receptor classes via annotations to the MF terms. This is a model for how a small ontology builder can constructively contribute to the growth of shared resources while simultaneously benefiting users of their own ontology.

Neuroanatomy. In support of research on neurodegenerative and neurological disease within the Biomedical Informatics Research Network,[41] the BIRN Ontology Task Force is applying the Foundry principles to formally represent several large domains, including: *neuroanatomy*, where annotations must capture not only the structural systems of parthood and topological (connection) relations but also cytoarchitectural parcellations such as the CA1, CA2, and CA3 regions of the hippocampus; *functional systems* such as the basal ganglion circuits for motor planning and motor memory; and *neurochemistry* (for example of brainstem monoamine nuclei). The Foundry is seen by the BIRN OTF as providing a framework within which these distinct axes can be algorithmically combined, and the results are being incorporated into BIRN's neuroimage atlasing project and used for integration of spatially mapped microarray expression data with mouse model imaging results.

MIBBI. The Minimal Information for Biological and Biomedical Investigations initiative represents the first new standards effort taking OBO and the OBO Foundry as its role model.[42] MIBBI provides information resources to promote the consolidation of the many *prescriptive checklists* which specify core metadata items to be included when reporting results in a variety of experimentation domains.[43] The proliferation of such 'minimal information' checklists has made it increasingly difficult to obtain an overview of existing specifications, resulting in an unnecessary duplication of effort and creating problems when third parties try to use described information. The *MIBBI Portal* operates in a manner analogous to OBO and the NBCO Bioportal

as an open information resource for all initiatives addressing these problems; the *MIBBI Foundry* fosters collaborative development and integration of checklists into orthogonal modules.[44]

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 fora (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 means of becoming familiar with the Foundry's collaborative methodology and of identifying 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 analogues of journal editors, whereby the division of labor that flows from orthogonality goes far towards ensuring that development decisions can be 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 in which it 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.

Acknowledgements

The Foundry is receiving *ad hoc* funding under the 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.

References

[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) 322-6 (2006).

[3] Camon, E., *et al*. The Gene Ontology Annotation (GOA) Project. *Genome Res.* **13**, 662–672 (2003).

[4] Kohane, I.S., Greenspun, P., Fackler, J., *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) 267-70 (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 Harbor Symposia on Quantitative Biology* **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**, 185-98 (2006).

[13] Grenon, P., Smith, B., Goldberg, L. Biodynamic ontology: Applying BFO in the biomedical domain, in DM Pisanelli (ed.), *Ontologies in Medicine*, Amsterdam, IOS Press, 2004, 20–38.

[14] Rosse, C., Mejino, J.L.F. The Foundational Model of Anatomy ontology, in: A Burger, *et al*. (eds.), *Anatomy Ontologies for Bioinformatics*, New York, Springer, in press.

[15] Haendel, M., *et al*, CARO: The Common Anatomy Reference Ontology, in: A Burger, *et al*. (eds.), *Anatomy Ontologies for Bioinformatics: Principles and Practice*, New York, Springer, in press.

[16] Leontis, N.B. *et al.*,The RNA Ontology Consortium: an open invitation to the RNA community. *RNA* **12**, 533-41 (2006).

[17] Natale, D. A. *et al.*, Framework for a Protein Ontology, *BMC Bioinformatics*, in press.

[18] Bard, J., Rhee, S.Y., Ashburner, M. An ontology for cell types, *Genome Biol.* **6**, R21 (2005).

[19] Kelso, J., *et al*. eVOC: a controlled vocabulary for unifying gene expression data. *Genome Res.* **13**, 1222-30 (2003).

[20] Mabee, P.M., *et al*. Phenotype ontologies: the bridge between genomics and evolution. *Trends Ecol. Evol.* **22**, 345-350 (2007).

[21] Whetzel, P.L., *et al*. The MGED Ontology: a resource for semantics-based description of microarray experiments. *Bioinformatics,* **22**, 866-73 (2006).

[22] Whetzel, P.L., et al. Development of FuGO: An ontology for functional genomics investigations. OMICS. 2006 ;10 (2):199-204.

[23] Brinkley, J.F., Detwiler, L.T., Gennari, J.H., Rosse, C., Suciu, D. A framework for using reference ontologies as a foundation for the semantic web. *Proc. AMIA Fall Symposium*, 2006, 95-100.

[24] 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.

[25] Smith, B., *et al*. Relations in biomedical ontologies. *Genome Biol.* **6**, R46 (2005).

[26] Bittner, T., Goldberg, L.J. Spatial location and its relevance for terminological inferences in bio-ontologies, *BMC Bioinformatics* **8**, 134-42 (2007).

[27] Ramírez, M.J., Coddington, J.A., Maddison WP,*et al*. Linking of digital images to phylogenetic data matrices using a morphological ontology, *Syst. Biol.* **56**, 283–294 (2007).

[28] Schober, D., *et al*. Towards naming conventions for use in controlled vocabulary and ontology engineering. *Bio-Ontologies Workshop*, ISMB/ECCB, Vienna, July 20, 2007, 87-90.

[29] Ruttenberg, A., Rees, J., Zucker, J. What BioPAX communicates and how to extend OWL to help it. - owl-workshop.man.ac.uk

[30] 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).

[31] Mungall, C.J. Obol: Integrating language and meaning in bio-ontologies. *Comp. Funct. Genom.* **5**, 509-520 (2004).

[32] Camon, E., Magrane, M., Barrell, B., *et al*.The Gene Ontology Annotation (GOA) Database: sharing knowledge in Uniprot with Gene Ontology. *Nucleic Acids Res*. **32** (Database issue), D262–D266 (2004).

[33] Blake, J., Hill, D.P., Smith, B. Gene Ontology annotations: What they mean and where they come from. *Bio-Ontologies Workshop*, ISMB/ECCB, Vienna, July 20, 2007, 79-82.

[34] Sjoblom, T. *et al*. The consensus coding sequences of human breast and colorectal cancers, *Science*, **314**, 268-274 (2006).

[35] Lee, J.A., *et al.* The Alliance for Cellular Signaling IC. Components of the antigen processing and presentation pathway revealed by gene expression microarray analysis following B cell antigen receptor (BCR) stimulation, *BMC Bioinformatics* **7**, 237 (2006).

[36] Rebholz-Schuhmann, D., Kirsch, H., Couto, F. Facts from text--Is text mining ready to deliver? *PLoS Biol.* **3**, e65 (2005).

[37] Witte, R., Kappler, T., Baker, C.J.O. Ontology design for biomedical text mining, in: *Semantic Web: Revolutionizing Knowledge Discovery in the Life Sciences*, Springer, New York, 2007, 281-313.

[38] Zhang, S., Bodenreider, O. Aligning multiple anatomical ontologies through a reference. *International Workshop on Ontology Matching* (OM 2006) 193-197 (2006).

[39] Luo, F., *et al*. Modular organization of protein interaction networks. *Bioinformatics* **23**, 207-214 (2007).

[40] Samwald, M., *et al*. Semantic SenseLab: implementing the vision of the Semantic Web in neuroscience. Submitted.

[41] Martone, M.E., Gupta, A., Ellisman, M.H. E-neuroscience: challenges and triumphs in integrating distributed data from molecules to brains. *Nat. Neurosci.* **7**, 467-72 (2004).

[42] Taylor, C.F., *et al*. Promoting coherent minimum reporting requirements for biological and biomedical investigations: The MIBBI Project. *Nat. Biotechnol*. in press. CHECK

[43] Brazma, A., *et al*. Minimum information about a microarray experiment (MIAME) – toward standards for microarray data. *Nat. Genet.* **29**, 365-71 (2001).

[44] Sansone, S.A., *et al*. A strategy capitalizing on synergies: the Reporting Structure for Biological Investigation (RSBI) working group. *OMICS* **10**, 164-71 (2006).