

## EDITORIAL

# Ontology and the future of dental research informatics

**Over time, more and more researchers will see the value of employing a common resource in annotating their data and in designing new databases in which to capture their research results.**

**H**ow do we find what is clinically significant in the swarms of data being generated by today's diagnostic technologies? As electronic records become ever more prevalent—and digital imaging and genomic, proteomic, salivaomics, metabolomics, pharmacogenomics, phenomics and transcriptomics techniques become commonplace—different clinical and biological disciplines are facing up to the need to put their data houses in order to avoid the consequences of an uncontrolled explosion of different ways of describing information.

Fortunately, a new strategy to advance the consistency of data in the dental research community is emerging. The strategy is based on the idea that existing systems for data collection in dental research will continue to be used, but proposes a methodology in which past, present and future data will be described using a consensus-based controlled structured vocabulary called the Ontology for Dental Research (ODR). The ODR initiative is modeled on a series of existing biomedical ontology projects and will adopt best-practice principles that already have been thoroughly tested in areas such as molecular biology, model organism research, proteomics and genetic disease.<sup>1</sup>

An “ontology,” in this context, is a controlled, logically structured vocabulary created by experts in a given area as a strategy for promoting consistency in the way primary data (for example, in the form of experimental results or clinical records) are described. Specialist biocurators create “annotations” in the form of HTML tags linking such primary data to expressions in the ontology, thereby making the data available to search and to algorithmic processing.

Each ontology contains a taxonomy at its heart, and its logical structure is built around the hierarchy defined by its taxonomic (subtype) relation. But an ontology contains also definitions of its terms, along with additional relations such as parthood, connection and participation, as well as functional relations. These additional relations make the data searchable not only through the use of terms in the ontology, but also through logically related terms. Thus, the ontology can be used to retrieve data associated with

**Barry Smith, PhD;  
Louis J. Goldberg, DDS, PhD;  
Alan Ruttenberg;  
Michael Glick, DMD**

terms referring to parts of specific anatomical entities, to anatomical entities immediately connected to specific anatomical entities, or to biological processes in which specific anatomical entities participate.

We can conceive ODR, in the first place, as providing an evolving standard set of key words for all aspects of dental research. Initially, these key words can be used to annotate both published literature and existing research databases. Such annotation will enable easier access to research results and allow also first steps toward the semantically enhanced publishing of the future.<sup>2</sup> The long-term goals of ODR, however, are much more ambitious. ODR will include not only English-language definitions of its terms for human use and for human quality control of the ontology, but also logical definitions for use by computers. With the latter, ODR then can be used as a computational resource for enhanced search and integration of data, and for reasoning—not only with dental research data but also with data annotated using other biological and biomedical ontologies with which ODR will be linked logically.

The key idea behind ODR is rooted in 10 years of experience using ontologies in support of biomedical research. Ontologies in biomedicine began in the model organism community, which faced a problem of inconsistency in the ways in which the results of functional genomics experiments on different kinds of organisms were being described. To address these problems, a group of leading model organism databases came together in 1999 to create the Gene Ontology (GO), a con-

trolled structured vocabulary for describing different attributes of gene products.<sup>3</sup>

The GO is designed to be species neutral. It provides a set of some 30,000 common terms for describing different kinds of cellular constituents, biological processes and molecular functions in all kinds of organisms—terms such as “mitochondrion” or “cell division” or “binding.” Since its inception, more than \$100 million has been invested in the use of the GO to annotate references to gene products in databases and in the scientific literature. There are more than 11 million annotations relating gene products described in the UniProt, Ensembl and other databases and in more than 50,000 scientific journal articles to terms in the GO.<sup>4</sup> The information in huge numbers of dispersed resources is hereby being made accessible through resources such as AmiGO and GOPubMed. Increasingly, the availability of this huge body of integrated information also is having an influence on clinical research, and a simple PubMed search on “gene ontology” reveals a variety of different ways in which the GO and the data annotated in its terms are being used in support of research on human health and disease.

Important features of ODR include the following:

- **It will be built to work with the GO and with other high-quality ontologies developed by the biomedical community.** This means that ODR will follow the best practices identified through 10 years of testing by the GO and by its sister ontologies participating in the Open Biomedical Ontologies (OBO) Foundry initiative.<sup>5</sup>
- **It will be built with terms**

**used by dental researchers, and it will be created and managed by the dental research community itself.**

The more an ontology is used, the more the ontology and the data described in its terms increase in value, and the more research groups in the future will be motivated to use the ontology in describing their data. The key to ontology success, therefore, is incentivizing users, and to this end it is important that potential users feel that they have ownership of the ontology, that the ontology is populated using the terms that they need and uses definitions that conform to their understanding of these terms. ODR is being initiated by the leading informatician groups within the dental research community in such a way that it will, from the very start, be in a position to serve as an attractor for multiple expanding groups of users whose members will have strong incentives not only to invest resources directed toward ensuring that it is developed in ways that keep pace with scientific advance, but also to recommend it to other users—thereby increasing the value of their own investment in the resource.

■ **It can be corrected easily in light of new research discoveries.** One key presupposition for the success of an ontology project is its ability to integrate previously annotated data with new terms and relations brought to light by ongoing scientific discovery. This process ensures that previously annotated (legacy) data do not lose their value. To this end, the biomedical ontology community has developed a methodology based on careful versioning of ontologies and annotations, com-

bined with software tools to ensure consistent updating of existing annotation resources with each new version of the ontology.

■ **It can be extended easily to incorporate new kinds of data.** The organization of OBO ontologies is based on the use of a simple and highly flexible tree-like hierarchy structure. This can be extended at will to comprehend new domains of entities as science evolves, and thereby allow the annotation of new kinds of data in ways consistent with existing annotations.

The ODR will benefit the research community in a number of ways:

■ It is designed to work well with existing ontologies in all areas of clinical and translational science, and thus allows dental research data to be easily integrated with other kinds of data.

■ It is designed to work well with the Semantic Web, providing access to all data resources through unique Web URLs associated with each ontology term.<sup>6</sup>

■ It provides a pretested and well-defined set of terms, selections from which can be used in the design of new databases.

■ It can incorporate, where needed, sets of synonyms deriving from legacy term sets and nomenclatures such as the Systematized Nomenclature of Medicine—Clinical Terms and Systematized Nomenclature of Dentistry vocabularies.<sup>7,8</sup>

■ To ensure high quality and continued maintenance, the ODR controlled vocabulary will be subject to a process of governance and peer review.

■ Organizations such as the National Institutes of Health are requiring definitions of com-

mon standards to ensure that the results obtained through funded research are more easily accessible to external groups. ODR will be created in such a way that its use will meet these common standards. It is designed also to allow information presented in its terms to be usable in satisfying regulatory purposes—submissions to the U.S. Food and Drug Administration, for example.

ORD will contain several subontology components, including the Salivaomics Ontology,<sup>9</sup> a Dental Anatomy Ontology based on the Foundational Model of Anatomy<sup>10</sup> and an Oral Pathology Ontology. In addition, vocabulary resources are being developed, based on the Ontology for General Medical Science (OGMS)<sup>11</sup> and the Ontology for Biomedical Investigations,<sup>12</sup> to represent dental disease and dental procedures, and to allow a seamless connection between the use of ODR in the dental domain and the use of existing ontology resources developed in other areas of medicine.

The use of ODR to describe data will be entirely voluntary. However, we anticipate that over time, more and more researchers will see the value of employing a common resource both in annotating their data and, progressively, in designing new databases in which to capture their research results. ■

Dr. Smith is a State University of New York Distinguished Professor and the director, National Center for Ontological Research, Department of Philosophy, University at Buffalo, N.Y.

Dr. Goldberg is a professor, Oral Diagnostic Sciences, School of Dental Medicine, University at Buffalo, N.Y.

Mr. Ruttenberg is a principal scientist,

Science Commons, Cambridge, Mass.

Dr. Glick is the dean, School of Dental Medicine, University at Buffalo, N.Y. He also is the editor of *The Journal of the American Dental Association*. Address reprint requests to Dr. Glick at School of Dental Medicine, University at Buffalo, 325 Squire Hall, Buffalo, N.Y. 14214-8006, e-mail “glickm@ada.org”.

1. Smith B, Ashburner M, Rosse C, et al. The OBO Foundry: Coordinated evolution of ontologies to support biomedical data integration. *Nat Biotechnol* 2007;25(11):1251-1255.

2. Kidd R. Semantic enrichment boosts information retrieval. *Res Inform* 2007; April/May. “www.researchinformation.info/features/feature.php?feature\_id=127”. Accessed Aug. 25, 2010.

3. Ashburner M, Ball CA, Blake JA, et al. Gene ontology: tool for the unification of biology—The Gene Ontology Consortium. *Nat Genetics* 2000;25(1):25-29.

4. European Molecular Biology Laboratory, European Bioinformatics Institute. Gene Ontology Annotation (UniProtKB-GOA) Database. “www.ebi.ac.uk/GOA/”. Accessed Aug. 25, 2010.

5. The Open Biological and Biomedical Ontologies. “www.obofoundry.org”. Accessed Aug. 25, 2010.

6. Ruttenberg A, Clark T, Bug W, et al. Advancing translational research with the semantic Web. *BMC Bioinform* 2007;8(suppl 3):S2. “http://www.biomedcentral.com/1471-2105/8/S3/S2”. Accessed Aug. 25, 2010.

7. International Health Terminology Standards Development Organisation. Welcome to IHTSDO. “www.ihtsdo.org/”. Accessed Aug. 31, 2010.

8. Goldberg LJ, Ceusters W, Eisner J, Smith B. The significance of SNODENT. *Stud Health Technol Inform* 2005;116:737-742.

9. Ai J, Smith B, Wong D. Saliva ontology: an ontology-based framework for a salivaomics knowledge base. *BMC Bioinform* 2010;11:302. “www.biomedcentral.com/1471-2105/11/302”. Accessed Aug. 28, 2010.

10. Rosse C, Mejino JLV. The foundational model of anatomy ontology. In: Burger A, Davidson D, Baldock R, eds. *Anatomy Ontologies for Bioinformatics: Principles and Practice*. London: Springer; 2007:59-117.

11. Ontology for General Medical Science (OGMS). “www.acsu.buffalo.edu/~ag33/ogms.html”. Accessed Aug. 28, 2010.

12. The Ontology for Biomedical Investigations. “http://obi-ontology.org/page/Main\_Page”. Accessed Aug. 25, 2010.

#### Additional Readings

Biomedical ontology. “http://ontology.buffalo.edu/biomedical.htm”. Accessed Aug. 31, 2010.

Stevens R. Ontologies in biology. “www.cs.man.ac.uk/~stevensr/menupages/background.php”. Accessed Aug. 31, 2010.

Introduction to biomedical ontologies. “http://bioontology.org/wiki/index.php/Introduction\_to\_Biomedical\_Ontologies”. Accessed Aug. 31, 2010.

Cambridge Healthtech Institute. Pharmaceutical ontologies & taxonomies glossary & taxonomy. “www.genomicglossaries.com/content/ontologies.asp”. Accessed Aug. 31, 2010.