

# BioPortal: A Web Repository for Biomedical Ontologies and Data Resources

[Demonstration]

Natalya F. Noy  
Michael Dorf  
Michael J. Montegut

Nigam H. Shah  
Nicholas Griffith  
Daniel L. Rubin  
Mark A. Musen

Benjamin Dai  
Clement Jonquet  
Cherie Youn

Stanford Center for Biomedical Informatics Research, Stanford University  
Stanford, California 94305

{noy, shah, bdai1, mdorf, ngriff, jonquet, mmontegut, dlubin, cyoun, musen}@stanford.edu

## ABSTRACT

Biomedical ontologies provide essential domain knowledge to drive data integration, information retrieval, data annotation, natural-language processing, and decision support. The National Center for Biomedical Ontology is developing BioPortal, a Web-based system that serves as a repository for biomedical ontologies. BioPortal defines relationships among those ontologies and between the ontologies and online data resources such as PubMed, ClinicalTrials.gov, and the Gene Expression Omnibus (GEO). BioPortal supports not only the technical requirements for access to biomedical ontologies either via Web browsers or via Web services, but also community-based participation in the evaluation and evolution of ontology content. BioPortal enables ontology users to learn what biomedical ontologies exist, what a particular ontology might be good for, and how individual ontologies relate to one another. BioPortal is available online at <http://alpha.bioontology.org>.

## 1. ONTOLOGY REPOSITORIES AND BIO-PORTAL

As the number of ontologies available for Semantic Web applications grows, so does the number of ontology repositories that index and organize the ontologies. Some repositories crawl the Web to collect ontologies (e.g., Swoogle [?], Watson [?] and OntoSelect [?]). In other repositories, users submit their ontologies themselves (e.g., the DAML ontology library<sup>1</sup> and SchemaWeb<sup>2</sup>). These repositories provide a gateway for users and application developers who need to find ontologies to use in their work. In our laboratory, we have developed BioPortal<sup>3</sup>—an open repository of biomedical ontologies. Researchers in biomedical informatics submit their ontologies to BioPortal and others can access the ontologies in their web browsers through the BioPortal user interface or through web services. The BioPortal users can browse and search the ontologies, update the ontologies in the repository that they authored by uploading new versions, comment on any ontology (or portion of an ontology) in the repository, evaluate it, describe their experience in

<sup>1</sup><http://www.daml.org/ontologies/>

<sup>2</sup><http://www.schemaweb.info/>

<sup>3</sup><http://alpha.bioontology.org>

using the ontology, or make suggestions to ontology developers. This focus on enabling members of the community to contribute actively to BioPortal content and to increase its value to other users, distinguishes BioPortal from other ontology repository.

Most researchers in biomedicine, however, are interested in biomedical data and the ontologies per se. Indeed, ontologies provide the means for them to access and integrate the data. Thus, one of the key features of BioPortal is the Open Biomedical Repository (OBR). To create OBR, we automatically index important biomedical data sets available online (e.g., entries in PubMed, GEO, ClinicalTrials.gov) on the basis of metadata annotations, and link the underlying data sets to the terms in the ontologies in BioPortal.

At the time of this writing, BioPortal has 72 biomedical ontologies with more than 300,000 classes. Ontology authors add new content regularly. While the BioPortal content focuses on the biomedical domain, the BioPortal technology is domain-independent.

## 2. SYSTEM DESCRIPTION

Ontologies in BioPortal may be represented in OWL, RDF, OBO Format, or the Protégé frame language. BioPortal uses the Mayo Clinic's LexGrid system<sup>4</sup> to store ontologies in OBO Format and to access standard biomedical terminologies, such as UMLS. Protégé<sup>5</sup> serves as the backend for OWL and RDF ontologies.

### 2.1 Key Features

In this demonstration, we will highlight several key BioPortal features.

#### *Ontology navigation and browsing.*

In its main browsing interface for ontologies, BioPortal displays the ontology class hierarchy in a tree display. When a user selects a class, BioPortal shows the details of the class definition. We provide different visualization methods for links between classes, such as nodes-and-links diagrams. Incorporation of the Jambalaya ontology-visualization system from the University of Victoria offers sophisticated graphics

<sup>4</sup><http://informatics.mayo.edu/LexGrid>

<sup>5</sup><http://protege.stanford.edu>

and animation for cognitive support of ontology navigation and perusal.

### *Marginal Notes.*

Users can add notes to all ontology classes in BioPortal, discussing the rationale for modeling decisions, pointing out problems with definitions, requesting changes from ontology authors, and so on. We plan to export these notes to the Protégé Changes and Annotations ontology [?], so that ontology developers can see the notes in the user interface for Collaborative Protégé [?], and can edit the ontologies accordingly to address the comments that users have stored in BioPortal. Similarly, if ontology developers add their own notes when developing an ontology in a tool such as Collaborative Protégé (e.g., providing references for a class definition, or explaining a design decision), they can choose to export their notes and make them visible as marginal notes in BioPortal.

### *Peer Reviews and Ontology Evaluation.*

When evaluating whether a particular ontology is appropriate for the user's task, one of the key pieces of information that a user will want is knowledge of those projects for which the ontology has been used, whether the projects' developers concluded that the ontologies had been appropriate for the projects, whether the ontology-based tasks in those project were similar to the current tasks that the user has in mind. Thus, we developed an infrastructure and a user interface to collect peer reviews of ontologies in the context of the specific project descriptions where the ontologies have been used. An ontology user can submit a description of his ontology-based project, link the description to the BioPortal ontologies that he used in the project, and provide comments on the ontologies along several different dimensions, such as degree of formality, documentation and support, usability, domain coverage, quality of content.

### *Ontology Mappings.*

Ontologies in BioPortal, as in almost any ontology repository, overlap in coverage. Thus, **mappings** among ontologies in a repository constitute a key component that enables the use of the ontologies for data and information integration. Thus, mappings between ontology concepts are first-class objects in the BioPortal repository. Users can browse the mappings, create new mappings, upload mappings created with other tools, download mappings that BioPortal has, or comment on the mappings and discuss them [?]. Each mapping has its own set of metadata that describes who created the mapping and when, which algorithm was used to produce the mapping, application context in which the mapping might be valid, the specific mapping relationship, and other properties. At the time of this writing, the BioPortal mapping repository contains more than 30,000 mappings created by biomedical researchers in different contexts.

### *Open Biomedical Resources.*

The Open BioMedical Resources (OBR) component automatically indexes important biomedical data sets available online (e.g., entries in PubMed, GEO, ClinicalTrials.gov) on the basis of metadata annotations, and links the underlying data sets to the terms in the ontologies in BioPortal. These linkages take advantage of the semantic relationships in BioPortal, including subsumption relationships among ontology entities and mappings among ontologies. OBR thus allows

biomedical investigators to use the terms in the BioPortal ontologies to enhance their ability to search for relevant on-line data in a manner that is not possible with conventional keyword search strategies.

## **2.2 BioPortal Technology**

BioPortal adopts a layered architecture approach, which decouples the logic and domain object models between each layer. The *Presentation Tier* delivers the BioPortal user interface, which currently adopts Ruby-on-Rails technology. The *Interface Tier* consists of both REST and SOAP Web services that present all BioPortal capabilities to the upper tiers (e.g., upload ontology, download ontology, display concept, administrative functions). The Presentation Tier is driven solely by the REST services. The *Business Logic Tier* uses NetKernel technology, which enables our collaborators to insert any software implementation that conforms with NCBO-defined interfaces.

## **3. CONCLUSIONS**

BioPortal offers investigators and clinicians "one-stop shopping" on the Web for important biomedical ontologies. The incorporation of a variety of Web 2.0 features allows the system to behave not only as a comprehensive ontology repository, but also as general infrastructure to support community-based access, peer-review, mapping, and annotation of ontology content. The BioPortal technology is open-source and is domain-independent. Thus, other communities can reuse the software to maintain their own ontology repositories.

## **4. ACKNOWLEDGMENTS**

This work was supported by the National Center for Biomedical Ontology, under roadmap-initiative grant U54 HG004028 from the National Institutes of Health.

## **5. REFERENCES**

- [1] P. Buitelaar, T. Eigner, and T. Declerck. OntoSelect: A dynamic ontology library with support for ontology selection. In *Demo Session at the International Semantic Web Conference (ISWC 04)*, Hiroshima, Japan, 2004.
- [2] M. d'Aquin, C. Baldassarre, L. Gridinoc, S. Angeletou, M. Sabou, and E. Motta. Watson: A gateway for next generation semantic web applications. In *Poster session at the International Semantic Web Conference (ISWC 2007)*, Busan, Korea, 2007.
- [3] L. Ding, T. Finin, A. Joshi, R. Pan, R. S. Cost, J. Sachs, V. Doshi, P. Reddivari, and Y. Peng. Swoogle: A search and metadata engine for the semantic web. In *Thirteenth ACM Conference on Information and Knowledge Management (CIKM'04)*, Washington DC, 2004.
- [4] N. F. Noy, A. Chugh, W. Liu, and M. A. Musen. A framework for ontology evolution in collaborative environments. In *Fifth International Semantic Web Conference, ISWC*, volume LNCS 4273, Athens, GA, 2006. Springer.
- [5] N. F. Noy, N. Griffith, and M. A. Musen. Collecting community-based mappings in an ontology repository. In *7th International Semantic Web Conference (ISWC 2008)*, Karlsruhe, Germany, 2008.
- [6] T. Tudorache, N. F. Noy, S. Tu, and M. A. Musen. Supporting collaborative ontology development in protege. In *7th International Semantic Web Conference (ISWC 2008)*, Karlsruhe, Germany, 2008.