W170–W173 *Nucleic Acids Research, 2009, Vol. 37, Web Server issue* *Published online 29 May 2009*
doi:10.1093/nar/gkp440

BioPortal: ontologies and integrated data resources at the click of a mouse

Natalya F. Noy[1], Nigam H. Shah[1], Patricia L. Whetzel[1,*], Benjamin Dai[1], Michael Dorf[1], Nicholas Griffith[1], Clement Jonquet[1], Daniel L. Rubin[1], Margaret-Anne Storey[2], Christopher G. Chute[3] and Mark A. Musen[1]

[1]Stanford Center for Biomedical Informatics Research, Stanford University, [2]Department of Computer Science, University of Victoria, Victoria, Canada and [3]Department of Biomedical Informatics, Mayo Clinic, Rochester, MN, USA

Received March 6, 2009; Revised April 19, 2009; Accepted May 11, 2009

ABSTRACT

Biomedical ontologies provide essential domain knowledge to drive data integration, information retrieval, data annotation, natural-language processing and decision support. BioPortal (http://bioportal.bioontology.org) is an open repository of biomedical ontologies that provides access via Web services and Web browsers to ontologies developed in OWL, RDF, OBO format and Protégé frames. BioPortal functionality includes the ability to browse, search and visualize ontologies. The Web interface also facilitates community-based participation in the evaluation and evolution of ontology content by providing features to add notes to ontology terms, mappings between terms and ontology reviews based on criteria such as usability, domain coverage, quality of content, and documentation and support. BioPortal also enables integrated search of biomedical data resources such as the Gene Expression Omnibus (GEO), ClinicalTrials.gov, and ArrayExpress, through the annotation and indexing of these resources with ontologies in BioPortal. Thus, BioPortal not only provides investigators, clinicians, and developers 'one-stop shopping' to programmatically access biomedical ontologies, but also provides support to integrate data from a variety of biomedical resources.

INTRODUCTION

As the number of biomedical ontologies increases, so does the number of repositories that index and organize ontologies. Some repositories crawl the Web to collect ontologies (e.g. Swoogle, http://swoogle.umbc.edu;, Watson, http://watson.kmi.open.ac.uk/Overview.html; and OntoSelect, http://olp.dfki.de/ontoselect?wicket:bookmarkablePage=wicket-0:de.dfki.ontoselect.Home), some request users to submit their ontologies themselves (e.g. the DAML ontology library. http://www.daml.org/ontologies; and SchemaWeb, http://www.schemaweb.info) and others are limited to storage of ontologies based on ontology representation language [Ontology Lookup Service, http://www.ebi.ac.uk/ontology-lookup/ (1,2)]. BioPortal is an open repository of biomedical ontologies that stores ontologies developed in various formats, that provides for automatic updates by user submissions of new versions, and that provides access via Web browsers and through Web services. BioPortal users can browse and search the ontologies, submit new versions of the ontologies in the repository, comment on any ontology (or portion of an ontology) in the repository, add a review of the ontology, describe their experience in using the ontology or make suggestions to ontology developers. The focus on enabling members of the community to contribute actively to BioPortal content and to increase the value of that content to other users distinguishes BioPortal from other ontology repositories.

Most researchers in biomedicine, however, are interested in biomedical data and not the ontologies *per se*. Indeed, ontologies provide the means for users to interpret and integrate the data. One of the key features of BioPortal is the Open Biomedical Repository (OBR). To create OBR, we have indexed biomedical data sets available online (e.g. entries in GEO, ClinicalTrials.gov) with ontologies in BioPortal. The index links the underlying data sets to the terms in the ontologies in BioPortal. While the content of BioPortal focuses on the biomedical domain, the BioPortal technology is domain-independent.

*To whom correspondence should be addressed. Tel: +001 650 721 2378; Fax: +001 650 725 7944; Email: whetzel@stanford.edu

© 2009 The Author(s)
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.0/uk/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

BMJ Publishing Group Ex. 1041
BMJ Publishing Group v. Evidium

AVAILABLE DATA

Ontologies

BioPortal contains ontologies that range in subject matter such as anatomy, phenotype, experimental conditions, imaging, chemistry, and health. In March 2008, the repository included 72 ontologies (300 000 total classes) and has almost doubled in the last year to 134 ontologies (680 000 total classes). Metadata collected for each ontology include keyword terms, text descriptions, version information, release date, ontology author contact information and links to documentation. Ontology content and metadata can be updated automatically or by user submission. Prior ontology versions are able to be accessed through both the Web services and Web interface and are available for download.

Notes

Users can add notes to all ontology classes in BioPortal, discussing the rationale for modeling decisions, pointing out problems with definitions and requesting changes from ontology authors. Notes are RSS-enabled to provide timely updates when new notes are added. The content of the notes can be exported and used in context with ontology editing software to incorporate changes to the ontology.

Reviews

BioPortal also supports peer reviews of ontologies. When a user is evaluating an ontology for use in a project, a key piece of information is what other projects have used the ontology and the suitability of the ontology for the tasks of the project. Users can submit descriptions of their ontology-based projects to BioPortal and link these descriptions to BioPortal ontologies. They can provide comments on the ontology along several different dimensions, such as degree of formality, documentation and support, usability, domain coverage, quality of content.

Mappings

BioPortal also represents mappings between terms in different ontologies. Users can browse the mappings, create new point-to-point mappings, upload mappings created with other tools, and download the mappings. They can add notes on existing mappings and carry out discussions about the mappings. 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, BioPortal contains more than 30 000 mappings created by biomedical researchers in different contexts. Mappings among ontologies constitute a key component that enables the use of the ontologies for data and information integration. In the future, we plan to include the ability to access the mapping data via the BioPortal Web services.

Open biomedical repository

The Open Biomedical Resources (OBR) component automatically indexes the metadata for important biomedical data sets available online (e.g. entries in GEO, ArrayExpress, ClinicalTrials.gov), and links the underlying data sets to the terms in the ontologies in BioPortal and UMLS. The annotation of these resources takes advantage of the semantic relationships in the ontology, including subsumption relationships among ontology terms and mappings among ontologies (3). OBR allows biomedical investigators to use the terms in the BioPortal ontologies to enhance their ability to search for relevant online data in a manner that is not possible with conventional keyword search strategies. These resources can be accessed through the Resources tab when users navigate the ontology (Figure 1), via a dedicated search mechanism or programmatically through Web services.

User interface enhancements

We have added functionality to the BioPortal user interface to facilitate the human computer interaction. Users can filter ontologies by group, such as all OBO Foundry ontologies (4) or caBIG ontologies (e.g. http://bioportal.bioontology.org/?filter=OBOFoundry). Users can also filter by domain or relevant organism. The full listing of ontologies available from BioPortal can be viewed at http://bioportal.bioontology.org/ontologies. RSS feeds provide alerts of changed content, notes, mappings, new ontology versions, and new ontology submissions.

Architecture

Ontologies in BioPortal may be represented in OWL, RDF, OBO format or the Protégé frame language. BioPortal uses the Mayo Clinic's LexGrid system (http://informatics.mayo.edu/LexGrid) to store ontologies in OBO format and to access standard biomedical terminologies, such as UMLS. Protégé (http://protege.stanford.edu) serves as the backend for OWL, RDF and Protégé frames ontologies.

BioPortal adopts a layered architecture approach (Figure 2), which decouples the logic and domain object models between each layer. The Presentation Tier delivers the BioPortal user interface, which currently uses Ruby-on-Rails technology. The Interface Tier consists of RESTful Web services that present all BioPortal capabilities to the upper tiers (e.g., browse, search, ontology concept display, visualization, etc.). The Presentation Tier is driven solely by the RESTful Web services. The Business Logic Tier provides API access to the ontologies and the resource index.

Web services

The BioPortal Web services are RESTful services. REST architecture consists of four verbs (GET, POST, PUT and DELETE) and nouns, which are URIs for the resources being accessed. Similar to SOAP services, RESTful Web services are language and platform independent. However, RESTful Web services have additional

BMJ Publishing Group Ex. 1041
BMJ Publishing Group v. Evidium

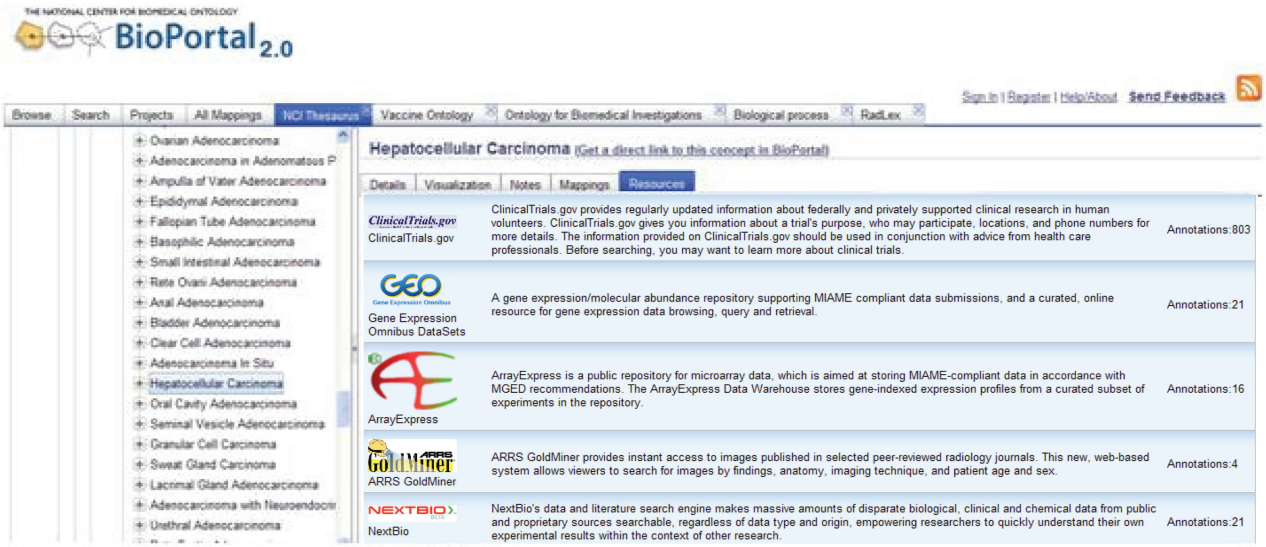

Figure 1. BioPortal resources tab displaying the ontology-indexed resources for the term Hepatocellular Carcinoma.

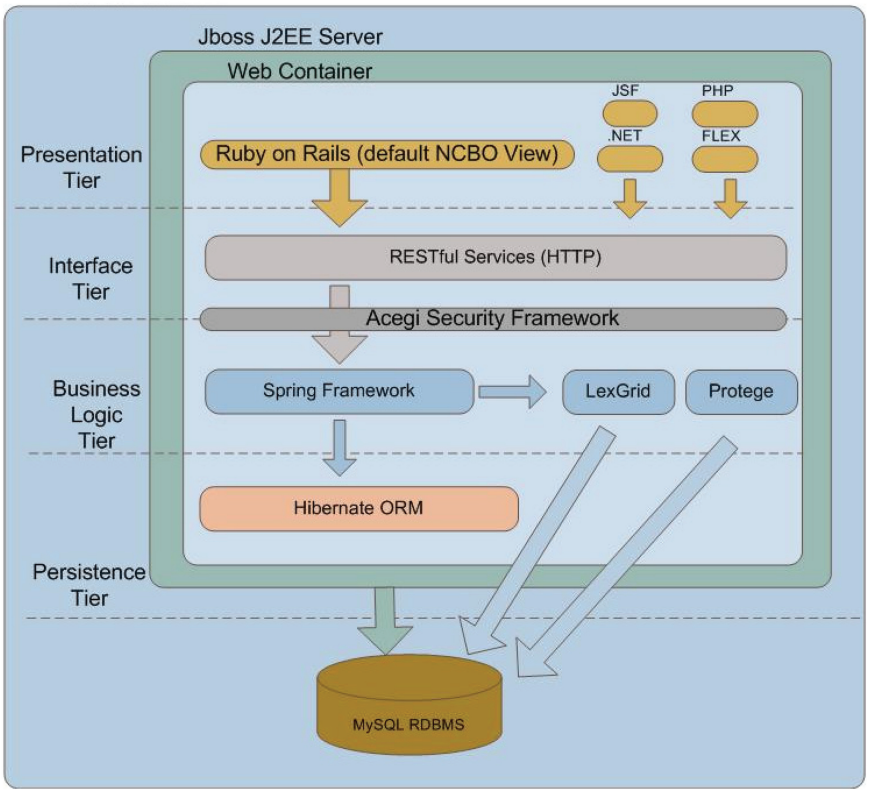

Figure 2. BioPortal architecture.

advantages in that they are light weight, simple and easier to integrate.

BioPortal Web services include services to get ontology metadata, to get ontology content, to download an ontology, and to search ontologies. Each ontology in BioPortal is indexed with a stable ontology identifier and is the same for all versions of the same ontology, ontology versions are indicated by a version identifier, which changes from one version to the next. BioPortal Web services provide a list of the latest versions of all BioPortal ontologies, enable callers to find an ontology identifier based on a version identifier, find all version identifiers for a specific ontology, and list ontology categories. Web services to get ontology content include services to get all root concepts,

BMJ Publishing Group Ex. 1041
BMJ Publishing Group v. Evidium

get children or parents of a specific concept, and get details of a concept. Users can request ontology content based on the ontology identifier, in which case BioPortal provides a response based on the latest available version of the ontology, or they can use a specific version identifier. In the search service, users can specify whether to search the whole repository or a specific set of ontologies, whether to use exact or approximate matching, and whether to search only in concept names and synonyms or in property values as well. A detailed list of Web services is available at http://bioontology.org/wiki/index.php/BioPortal_REST_services. Additional Web services include a hierarchy service to provide all paths to the root and all paths to the leaf node terms. These services can be used to develop applications using ontologies without the need to create a local store of the ontology. A few examples of projects accessing ontology content from BioPortal include ISAcreator (http://isatab.sourceforge.net/isacreator.html), which enables users to structure experimental metadata, BioLit (5), which provides metadata describing the content of full text articles from PubMed Central as ontology terms, and the Microsoft Word Add-in for Ontology Recognition (http://ucsdbiolit.codeplex.com/), which enables authors to include semantic content in documents at the time of writing.

We are also developing Web services to retrieve the content of the annotated resources. A few example services include get all resources, get all annotations for a given concept (BioPortal or UMLS), and get all annotations for a given resource identifier from GEO, Reactome, or OMIM for example. Ontology-indexed content can be accessed and then displayed on a site of interest. For example, ontology-indexed annotations from GEO using the Rat Strain ontology and anatomy ontology can be generated and then the data displayed on a third party Web site such as the Rat Genome Database by retrieving the data via the Web services.

DISCUSSION

BioPortal is a repository for ontologies developed in various formats that allows multiple mechanisms for content updates and that provides access via Web services. 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 a general infrastructure to support community-based access, peer-review, mapping and annotation of ontology content. BioPortal not only provides investigators, clinicians and developers 'one-stop shopping' to programmatically access biomedical ontologies, but also integrates data from various biomedical resources. The BioPortal technology is open-source and is domain-independent. Thus, other communities can reuse the software to maintain their own ontology repositories.

Currently, the Marine Metadata Interoperability Project (http://marinemetadata.org/aboutmmi) and National Cancer Institute (http://bioportal.nci.nih.gov/ncbo/faces/index.xhtml) are using BioPortal as their ontology repository.

FUNDING

National Center for Biomedical Ontology, under roadmap-initiative from the National Institutes of Health [grant U54 HG004028]. Funding for open access charge: National Institutes of Health [grant U54 HG004028].

Conflict of interest statement. None declared.

REFERENCES

1. Cote,R.G., Jones,P., Martens,L., Apweiler,R. and Hermjakob,H. (2008) The Ontology Lookup Service: more data and better tools for controlled vocabulary queries. *Nucleic Acids Res.*, **36**, W372–W376.
2. Cote,R.G., Jones,P., Apweiler,R. and Hermjakob,H. (2006) The Ontology Lookup Service, a lightweight cross-platform tool for controlled vocabulary queries. *BMC Bioinformatics*, **7**, 97.
3. Shah,N.H., Jonquet,C., Chiang,A.P., Butte,A.J., Chen,R. and Musen,M.A. (2009) Ontology-driven indexing of public datasets for translational bioinformatics. *BMC Bioinformatics*, **10(Suppl. 2)**, S1.
4. Smith,B., Ashburner,M., Rosse,C., Bard,J., Bug,W., Ceusters,W., Goldberg,L.J., Eilbeck,K., Ireland,A., Mungall,C.J. *et al.* (2007) The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat. Biotechnol.*, **25**, 1251–1255.
5. Fink,J.L., Kushch,S., Williams,P.R. and Bourne,P.E. (2008) BioLit: integrating biological literature with databases. *Nucleic Acids Res.*, **36**, W385–W389.

BMJ Publishing Group Ex. 1041
BMJ Publishing Group v. Evidium