

Jie Zheng^{1*}, Zuoshuang Xiang^{2*}, Christian J. Stoeckert Jr¹, Yongqun He²
¹ Department of Genetics, University of Pennsylvania, Philadelphia, PA 19104, USA

² Department of Microbiology and Immunology, University of Michigan, Ann Arbor, MI 48109, USA

ABSTRACT

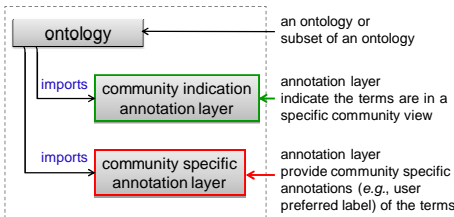
Reference ontologies are often very large and complex. When applied to a specific application, generally a subset of one reference ontology is needed. Moreover, the labels of ontology terms that were given in the perspective of ontology developers might not be preferred labels to the end users. Therefore, it is desirable to have a community view of a reference ontology that is a subset of the ontology including the terms needed for a particular application or community with user-preferred labels. Ontodog is a web-based system to support generation of ontology community views. Ontodog allows users to provide terms of interest in a source ontology and customized annotation information, such as user-preferred label. With these inputs, Ontodog can extract a subset of the source ontology containing all the terms of interest and generate user specified annotations in RDF/XML format (i.e., OWL files) which can be used to build an ontology community view. Currently over 100 ontologies including all OBO Foundry ontologies are available in Ontodog to generate views for a specific application or community. We demonstrate the application of Ontodog in generating ontology community views using the Ontology for Biomedical Investigations (OBI) [1] as the source ontology.

Ontodog website: <http://ontodog.hegroup.org/>

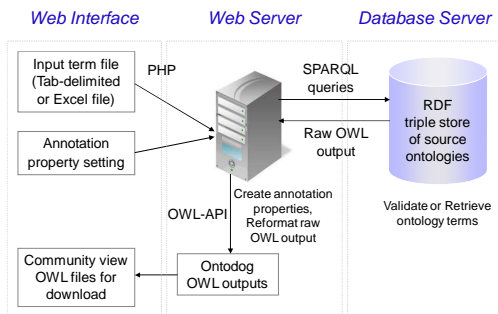
Ontology Community View

- a subset of the whole ontology or tagged subset of terms in the whole ontology to meet users' specific need
- contains user specified annotations (e.g., user preferred label) where needed

Ontology Community View



System Architecture



Ontodog workflow

1. The input data is processed using PHP.
2. SPARQL queries are issued against an RDF triple store to validate whether terms exist in the source ontology or retrieve terms from the source ontology
3. OWL-API is used to create annotation properties and reformat the files in RDF/XML format as output files.
4. Then the RDF/XML format output files are provided to the users for download.

References:

1. Brinkman RR, et al (2010) Modeling biomedical experimental processes with OBI. *J. Biomed. Semantics*. 1(Suppl. 1), S7.
2. Xiang Z, et al (2010) OntoFox: web-based support for ontology reuse. *BMC Res Notes*. 3:175.

Funding: NIH grants R01AI081062, P41HG003619, and R01GM93132

Features and Usage

The screenshot shows the Ontodog web interface. At the top, the logo and URL 'http://ontodog.hegroup.org/' are displayed. Below is the 'Ontodog data input' section with three main steps:

- (1) Provide Ontodog input term file:** A file upload field for a Microsoft Excel or tab-delimited text file. A 'Browse...' button is available. A note indicates this is required.
- (2) Select one source ontology:** A dropdown menu showing 'Ontology for biomedical investigations (OBI)'. A note indicates this is required. Below the dropdown, there is a field for 'Or enter your own source ontology and SPARQL endpoint: Example'.
- (3) Settings for Ontodog output files:** A section with three options, each with a 'help' link and a note: 'Please check at least one file'.
 - Output File 1: inSubSet annotation ontology:** Includes a field for the URI of the owl file (e.g., http://purl.obolibrary.org/obo/FGED_inSubset.owl), a field for the Annotation Property URI (default: <http://www.geneontology.org/formats/oboInOwl#inSubset>), and a field for the Annotation Value (e.g., SLIM, FGED, IEDB ...).
 - Output File 2: User preferred annotation ontology:** Includes a field for the URI of the owl file (e.g., http://purl.obolibrary.org/obo/FGED_label.owl), a field for the Annotation Property URI (one child term of IAO: alternative term, e.g., http://purl.obolibrary.org/obo/OBI_9991119), and a field for the label of the Annotation Property if not defined. A note indicates this is required for output file 1.
 - Output File 3: Subset of source ontology containing all view terms and related terms/axioms:** Includes a field for the URI of the owl file (e.g., http://purl.obolibrary.org/obo/obi_FGED.owl), a checkbox for 'Include inferred hierarchy?', and a checkbox for 'Include all individuals of a class in the subset?'.

Buttons for 'Get OWL (RDF/XML) Output Files' and 'Reset' are at the bottom.

Input term file (tab-delimited or Excel file, template file can be generated by Ontodog)

| A | B | C | D | E |
|---|--|-------------------------------------|------------------------------|-------------------------------------|
| http://purl.obolibrary.org/obo/OBI_0001182 | negative binding datum | <input checked="" type="checkbox"/> | User-preferred label | Include all children |
| http://purl.obolibrary.org/obo/OBI_0001183 | epitope protection from tumor challenge experiment | <input type="checkbox"/> | | |
| http://purl.obolibrary.org/obo/OBI_0001184 | epitope protection experiment | <input type="checkbox"/> | | |
| http://purl.obolibrary.org/obo/OBI_0001185 | selectively maintained organism | <input checked="" type="checkbox"/> | strain, cultivar, or ecotype | <input checked="" type="checkbox"/> |

Annotations: indicate terms are chosen for inclusion in the view. indicate all children of the chosen term will be included in the view.

Use Cases: OBI FGED View, OBI Core

OBI FGED view: a view generated for the Functional Genomics Data (FGED) community, a simplified set of OBI terms with FGED community friendly labels.

The screenshot shows the OBI FGED view. On the left, a tree view shows the hierarchy of terms: 'organism' (Archaea, Bacteria, Eukaryota, Viruses), 'genetically modified organism', 'host of immune response', 'infectious agent', and 'selectively maintained organism'. The main area shows a list of terms with their annotations. For example, 'FGED alternative term' has the annotation '*strain, cultivar, or ecotype@en'. 'FGED preferred term' has the annotation '*selectively maintained organism@en'. 'FGED term editor' has the annotation '*Bjoern Peters, Helen Parkinson, Philipps'. 'FGED definition' has the annotation '*an organism that is bred to have some characteristics with similarly bred organ'. 'FGED inSubset' has the annotation '*FGED@en'. 'FGED label' has the annotation '*selectively maintained organism@en'.

OBI core: a view contains all OBI core terms with labels in different languages

The screenshot shows the OBI core view. It displays a list of terms with their annotations in multiple languages. For example, 'alternative term' has the annotation '*measuring@en'. 'definition source' has the annotation '*OBI branch derived@en'. 'definition' has the annotation '*A planned process with the objective to produce it'. 'label' has the annotation '*assay@en'. The view also shows terms in Chinese and German, such as '*测量@ch' and '*messung@de'.

Summary

- A web system that generates ontology community views with customized annotations
- Easy to use with minimal ontology knowledge and no installation required
- Currently only supports ontologies in OWL format

* Authors contribute equivalently