

PRO Science Commons Projects-Current issues

Alan Ruttenberg
Science Commons

Overall Technical approach

- From OBO Foundry: Carefully model biology to enable integration of data sources. “Audit trail to reality”. All classes are defined in Foundry or Foundry candidate ontologies.
- From Web: Assign all biological entities URIs (lots already provided by OBO). Be on the Semantic Web.
- Translate to RDF and managed in a triple store or provide technology that enables SPARQL query against traditional data stores.
- From OWL: Add triples inferred by reasoner to increase expressiveness of queries with even simple query engine
- From software engineering: Provide data via SPARQL first (API). Build tools on top of that.

Projects

- PONS - Program for ontologies of neural structures
- IDO - Infectious Disease ontology
- FMA – Foundational model of anatomy
- Semantic Resources Projects (in aid of SCF)
- OBI – Ontology of Biomedical Investigations
- CRUX – Management and repository for Parkinson disease grant supported research

Target OBO Ontologies for PONS Work

- Gene Ontology (GO) *
 - Subcellular structures added to GO CC
- Cell Ontology (CL)
 - Target for neuron types
- Foundational Model of Anatomy (FMA)
 - Target for cross-species Anatomical Structures
- Relation Ontology (RO)
 - Target for connectivity and spatial relations
- Chemical Entities of Biological Interest (CheBI) *
 - Target for some (non-gene-product/small molecule) neurotransmitters
- Phenotype Ontology (PATO) *
 - Target for morphology description
- Ontology for Biomedical Investigations (OBI)
 - Target for methods
- Protein Ontology (PRO) *
 - Target for expressed proteins/ receptors/neurotransmitters

* Current foundry ontologies

Deployment target: The Semantic Web

Adds to Web standards and practices encouraging

- Unambiguous names for things, classes, and relationships
- Well organized and documented in ontologies
- With data expressed using uniform knowledge representation languages
- To enable computationally assisted exploitation of information
- That can be easily integrated from different sources
- Both within and across public and organizational boundaries

What the browser sees at http://purl.obolibrary.org/obo/OBI_0000225



OBI_0000225: Investigation Site

http://purl.obofoundry.org/obo/OBI_0000225 Google

phd | lool | Leech | hakia | Paris | BeeSpace Navigator (v3) | w3c | data | games | ont | Omics! Omics! | uri | To Investigate

 

Ontology for Biomedical Investigations (OBI) | [OBI Wiki](#) | [obi.owl](#) | [Issue tracker](#) | [OBO Foundry](#)

Investigation Site

Class: http://purl.obofoundry.org/obo/OBI_0000225

(and role
(disjoint-with nutrient-role study-personnel-role patient-role regulatory-role drug-role study-participant-role vector-role reference-role))

definition: Investigation site is a role borne by a site realized in an investigation which is located at the site

curation status: metadata-incomplete

preferred term: Investigation site

example of usage: A field, a laboratory, a medical institute, a pharmaceutical company

definition source: source pending

editor note: solution2: site is related to trial used located_in relation – site can bear the role

editor note: solution1: is a physical location, should maybe go under processual context, and then be used in conjunction with the located relation

editor note: site is a material (building) having the role site

definition editor: Jennifer Fosterl

Subject of: location_of, participates_in, is_proxy_for, proper_part_of, derives_from, relationship, transformed_into, has_proper_part, is_output_of, is_realized_as, part_of, has_integral_part, derived_into, is_input_of, ObsoleteProperty, agent_in, has_part, integral_part_of, contains, transformation_of, has_improper_part, improper_part_of, located_in, contained_in, adjacent_to

Object of: location_of, is_proxy_for, proper_part_of, derives_from, relationship, transformed_into, has_proper_part, part_of, has_participant, has_integral_part, derived_into, has_output, ObsoleteProperty, has_part, has_input, integral_part_of, contains, transformation_of, is_realization_of, has_improper_part, improper_part_of, located_in, contained_in, has_role, adjacent_to, has_agent

What's really at http://purl.obolibrary.org/obo/OBI_0000225

Don't read this (but machines love it!)

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="http://ashby.csail.mit.edu/cgi-bin/obiterm.xsl?ref=OBI_0000225"?>
<owl:Class rdf:about="&obo;OBI_0000225">
  <obo:OBI_0000274 rdf:datatype="&xsd:string">Jennifer Fostel</obo:OBI_0000274>
  <obo:OBI_0000275 rdf:datatype="&xsd:string">site is a material (building) having the role site</obo:OBI_0000275>
  <obo:OBI_0000275 rdf:datatype="&xsd:string">solution1: is a physical location, should maybe go under processual
    context, and then be used in conjunction with the located relation</obo:OBI_0000275>
  <obo:OBI_0000275 rdf:datatype="&xsd:string">solution2: site is related to trial used located_in relation –
    site can bear the role</obo:OBI_0000275>
  <obo:OBI_0000281 rdf:resource="&obo;OBI_0000320"/>
  <obo:OBI_0000287 rdf:datatype="&xsd:string">A field, a laboratory, a medical institute, a pharmaceutical company</obo:OBI_0000287>
  <obo:OBI_0000288 rdf:datatype="&xsd:string">Investigation site</obo:OBI_0000288>
  <obo:OBI_0000291 rdf:datatype="&xsd:string">Investigation site is a role born by a site or material realized in an investigation which is
    located at the investigation site</obo:OBI_0000291>
  <dc:creator rdf:resource="&obo;obi/project"/>
  <rdfs:isDefinedBy rdf:resource="&obo;obi.owl"/>
  <rdfs:label rdf:datatype="&xsd:string">investigation site role</rdfs:label>
  <rdfs:subClassOf rdf:resource="http://www.ifomis.org/bfo/1.1/snap#Role"/>
</owl:Class><doap:Project rdf:about="&obo;obi/project">
  <doap:browse rdf:resource="http://obi.svn.sourceforge.net/viewvc/obi"/>
  <doap:bug-database rdf:resource="http://sourceforge.net/tracker/?group_id=177891&atid=886178"/>
  <doap:homepage rdf:resource="http://obi.sourceforge.net"/>
  <doap:mailing-list rdf:resource="mailto:obi-devel@groups.google.com"/>
  <doap:release rdf:resource="&obo;obi/version-2008-03-10"/>
  <doap:repository rdf:resource="http://obi.svn.sourceforge.net/svnroot/obi"/>
  <doap:wiki rdf:resource="https://wiki.cbil.upenn.edu/obiwiki/index.php?title=HomePage"/>
</doap:Project>
<doap:Version rdf:about="&obo;obi/version-2008-03-10">
  <doap:file-release rdf:resource="http://purl.org/obo/2008-03-10/obi.owl"/>
  <doap:file-release rdf:resource="http://purl.org/obo/obi.owl"/>
</doap:Version>
</rdf:RDF>
```

Current issues with PRO representation

- Are PFAM domains protein parts or protein families?
 - The current use of `has_part` for PFAM domains is not sanctioned
 - Proposal: Generate 3 ontologies (or RABOPs)
 - 1) Information artifacts and processes (hmms, alignments)
 - 2) Protein parts (domains)
 - 3) Protein families (proteins with a given part)

Current issues with PRO representation

- Are GO Molecular functions functions?
 - An important cadre of GO ontologist don't consider MF terms to denote functions – they say they are processes
 - But: For every function there is a processes. In addition functions say *how* participants participate in those processes.
 - Proposal:
 - 1) We need functions. Lobby the GO to get clear on this matter
 - 2) Use them as functions and win by majority

Current issues with PRO representation

- PRO is too many files and representations
 - PRO.obo
 - PAF “Annotations”
 - Mappings
- Proposal
 - No “Annotations” or “Mappings” – just assertions
 - One document that holds all assertions
 - Subsets programmatically extracted from the whole

Current issues with PRO representation

- Uniprot coordination
 - “The UniProtKB/Swiss-Prot policy is to describe all the protein products encoded by one gene in a given species in a single entry.” <http://www.uniprot.org/faq/30>
 - In ontologies and RABOPs what identifier are we to use to represent these?
 - Potential for repeated duplicative efforts (modifications, complexes) and corresponding difficult user choices on identifier use.

Current issues with PRO representation

- Positions
 - How to stably deal with with naming positions describing where modifications or bindings happen or parts are located
 - Across genome builds
 - Across species
- Proposal:
 - PRO includes distinct entities for these