Ontology Development and Usage for Protozoan Parasite Research

John A. Miller and Alok Dhamanaskar

Collaborators:
Michael E. Cotterell, Chaitanya Guttula, Yung Long Li and Jessica C. Kissinger, University of Georgia.
Tania Tudorache, Natasha Noy, Mark Musen, Stanford University

Integrative Tools for Protozoan Parasite Research
R01 GM093132
Ontology Development and Usage for Protozoan Parasite Research: 
Aim 1

1. Support development of forms for isolate collections, phenotypes, and other metadata using a semantic framework based on OBO Foundry-compliant ontologies

   1. Ontology for Biomedical Investigations (OBI)
   2. Ontology for Parasite Lifecycle (OPL)
   3. Collaborative ontology development with Web Protégé
Ontology Development and Usage for Protozoan Parasite Research: Aim 2

   1. Enrichment of the Ontology for Biomedical Investigations (OBI)
      1. **OBI**: purl.obolibrary.org/obo/obi.owl
      2. **OBI-enriched-WS**: obi-webservice.googlecode.com
      3. **OBI-WS**: OBI Extended Web Services Ontology
   2. Creation of Web Services / Processes for Protozoan Parasite Research
      1. **Phylip** SOAP Web Service
      2. Re-engineer REST Web Services for **EuPathDB**
      3. **Workflow**: Sequence and Phylogenetic Analysis of Proteins
   3. Semantic Annotation of Existing Web Services Using Ontologies
      1. Using Semantic Annotations for WSDL (**SAWSDL**) for **SOAP Services**
Ontology Development and Usage for Protozoan Parasite Research: Aim 3

3. Development of Web Services, Workflows and Data Integration Tools
   1. Provision of Web Service Support for Galaxy
      1. Tool 1: Web Service Extension for Galaxy: Galaxy WS-Extensions
   2. Improved Automation through Semantics
      1. Tool 2: Service Annotation: RadiantWeb
      2. Tool 3: Service Discovery: LuminaWeb
      3. Tool 4: Service Suggestion Engine: SSE

Example from 2.2.3: Development of a Workflow
I have heard that jBPM is very powerful, Taverna is very popular, but Galaxy is very easy to use.
Find Services / Tools for My Workflow

Seems like a lot of services to look through! I wonder if I missed any?

Total Services = 483 with maybe 1000s of operations

Popular Web service Registries
Tools available in Galaxy (461) +
Find Input-Output Compatible Services

I selected 2 of those operations, entered them in my workflow designer, but when I tested, it didn’t work.

Require Data Mediation

Type Mismatch

Input output Matching Problem
Connect the Newly Added Service into the Workflow

Wish I could figure out which service operation(s) might fill the gap in my workflow.

Why can’t the computer make a suggestion?

Workflow design / service composition
Find Services / Tools

Issues related to working with Web Services and Tools:

- Web services are developed by different contributors
- Not developed to work with one another
- No standard naming conventions
- Text descriptions are inherently ambiguous
Find Services / Tools, contd.

We have

- formalized a methodology for enriching an Ontology with terms to support Web service annotation,
- Enriched the OBI ontology, by generating ~130 new terms to support annotation of 61 operations from 12 popular Web Services.

**Ontologies**

- provide a rich modeling framework,
- enable reuse of domain knowledge,
- facilitate community agreement,
- facilitates reasoning to ensure consistency.

WSDL for WUBLAST Web service

```
<WSDL>
<operation name="runBlast" />
<input name="seq" />
```

Ontology
Semantic Mark-up of Description Documents: What does it take?

RadiantWeb Annotation Tool

WSDL Viewer

Ontology Viewer

WU-BLAST stands for Washington University Basic Local Alignment Search Tool. The emphasis of this tool is to find regions of sequence similarity or homology quickly, with minimum loss of sensitivity. This will yield functional and evolutionary clues about the structure and function of your novel sequence. Dr. Warren Gish at Washington University released this first “gapped” version of BLAST allowing for gapped alignments and statistics.

Identified Operations in the WSDL (portType : JDispatcherService)

- Run
  - Inputs (runRequest)
    - Email
      - email address
    - Title
    - Program
    - BLAST-algorithm

A textual entity that is used as a unique identifier of an email box, that is located on a particular email server, and which is used by a person or an organization for purposes of email delivery or sending (i.e., email)

Class

ObjectProperty

Legend

- Approve Suggested Term
- Reject Suggested Term
- Remove annotation
- Suggested Term
- Pre-existing annotation
- Approved Term
- Operation
- Messages (Input/Output)
- Parameters
Find Services / Tools, contd.

Discovery of operations over annotated Web services: **LuminaWeb**

WSDL for WUBLAST Web service
The Service Suggestion Engine (SSE) is a semi-automatic workflow composition system.

What it does:
- Facilitates the construction and extension of workflows by providing suggestions to the user for the next step.
- It is capable of doing Forward Suggestions, Backward Suggestions and Bi-directional Suggestions.
Sample Use Case for SSE

• Find out more information about a protein sequence and its evolutionary relationships to other protein sequences.

• The user might be aware that he wants to
  • first search a database for similar sequences,
  • then perform multiple sequence alignment and
  • finally perform phylogenetic analysis to construct phylogenetic trees.
Workflow Construction
How NCBO Enables this Work

• Editor: **Protégé**
  - Has been vital for the Ontology Development aspect of our project.
  - Collaborative Protégé made it possible for all the members of our team to work on the same copy of the Ontology, while maintaining the change log.

• Editor: **Web-Protégé**
  - Recently we have moved to Web Protégé for the ease of use.

• Ontology Registry: **NCBO BioPortal**
  - Consulted before proposing terms to add to OBI
  - Made it easy to determine whether terms are already present in existing ontologies
Communities where We Can Increase NCBO Visibility

- **Galaxy**
  - A popular Web-based tool that allows users to integrate data from a large number of resources.
  - [http://galaxy.psu.edu](http://galaxy.psu.edu)

- **GenomeSpace**
  - A platform enabling one to work with multiple genomic analysis tools, seamlessly.

- **Web services**
  - IEEE SERVICES: IEEE World Congress on Services
  - [http://www.servicescongress.org/2012/](http://www.servicescongress.org/2012/)

- **Ontologies**
  - International Conference on Biomedical Ontology (ICBO)
  - Formal Ontology in Information Systems (FOIS)
Suggestions & Complaints for Core NCBO Team

• Suggestions already addressed by the NCBO team:
  • Web Protégé
    • Support for **Adding and Editing Axioms**
      e.g., Adding restrictions on properties for concepts
  • Documentation for **Layout Configuration** to the Wiki

• Additional Suggestions
  • Web Protégé
    • Support for **Reasoning**
## Milestones and Plans

<table>
<thead>
<tr>
<th>Artifact</th>
<th>Current Version</th>
<th>Next Version</th>
<th>Improvements</th>
</tr>
</thead>
<tbody>
<tr>
<td>OBI-WS</td>
<td>March 2012</td>
<td>Jul 2012</td>
<td>Addition of terms for annotation of Galaxy Tools and EuPathDB Web services</td>
</tr>
<tr>
<td>OBI-enriched-WS</td>
<td>March 2012</td>
<td>Jul 2012</td>
<td></td>
</tr>
<tr>
<td>WSDL_XSD Parser</td>
<td>Apr 2011</td>
<td>Apr 2012</td>
<td>Add documentation and release</td>
</tr>
<tr>
<td>ConceptSim Metrics</td>
<td>Nov 2011</td>
<td>Jul 2012</td>
<td>Generalize to work with multiple ontologies</td>
</tr>
<tr>
<td>Galaxy WS-Extensions</td>
<td>Jul 2011</td>
<td>Apr 2012</td>
<td>Add support for complex types for outputs</td>
</tr>
<tr>
<td>RadiantWeb</td>
<td>Dec 2011</td>
<td>Apr 2012</td>
<td>Add support for WADL Web services</td>
</tr>
<tr>
<td>LuminaWeb</td>
<td>NA</td>
<td>Jul 2012</td>
<td>Developing discovery tool</td>
</tr>
<tr>
<td>SSE</td>
<td>Mar 2012</td>
<td>Jul 2012</td>
<td>Improved use of Semantics, work with multiple Ontologies and data mediation</td>
</tr>
<tr>
<td>Phylip WS</td>
<td>NA</td>
<td>Jul 2012</td>
<td>Developing phylogenetic service</td>
</tr>
<tr>
<td>EuPathDB-WS</td>
<td>NA</td>
<td>Jul 2012</td>
<td>Re-engineering EuPathDB</td>
</tr>
</tbody>
</table>
Questions