Ontologies for Web Service Annotations

OBI & EDAM

Dr. Jessica Kissinger
Department Of Genetics
University Of Georgia
Outline

- Brief Introduction To OBI (Ontology For Biomedical Investigation)
  - Our Role of Enriching OBI for Web Service Annotations
  - EDAM
  - Similarities Between EDAM and OBI, working towards the same GOAL
  - Our work on OBI so far
  - Comparison of OBI and EDAM on the basis of definitions, relations, etc.
  - Methodology we follow
- Our Related Projects
  - Galaxy Extensions
  - Suggestion Engine
The Ontology for Biomedical Investigations (OBI) project is developing an integrated ontology for the description of biological and clinical investigations. This includes a set of 'universal' terms, that are applicable across various biological and technological domains, and domain-specific terms relevant only to a given domain.

This ontology will support the consistent annotation of biomedical investigations, regardless of the particular field of study.

The ontology will represent the design of an investigation, the protocols and instrumentation used, the material used and the data generated.

Currently OBI is being built under the Basic Formal Ontology (BFO) (more on it in the next slide..!) and is a member of the OBO Library making it interoperable with OBO compliant ontologies.
BFO has grown out of a philosophical orientation which overlaps with narrow focus on the task of providing a genuine upper ontology that can be used in support of domain ontologies developed for scientific research (as for example in biomedicine within the framework of the OBO Foundry).

Thus BFO does not contain physical, chemical, biological or other terms which would properly fall within the special sciences domains.
Basic Formal Ontology (BFO) – The Upper level Ontology for OBI

The graphs are only for visualization purposes and are not an EXACT representation of the hierarchy in the Ontology, some intermediate levels have been intentionally skipped for simplicity.

entity

continuant

dependent
continuant

independent
continuant

occurrent

processual_entity

Represents entities which exists in full at any time in which it exists at all and has no temporal parts. *(data...!!)*

Represent entities that exists in time by occurring or happening, has temporal parts. *(operations...!!)*
Outline

- Brief Introduction To OBI (Ontology For Biomedical Investigation)
- **Our Role of Enriching OBI for Web Service Annotations**
- EDAM
  - Similarities Between EDAM and OBI, working towards the same GOAL
- Our work on OBI so far
- Comparison of OBI and EDAM on the basis of definitions, relations, etc.
- Methodology we follow
- Our Related Projects
  - Galaxy Extensions
  - Suggestion Engine
What are we doing…?

- **Enriching OBI for supporting annotation of Bioinformatics Web Services.**

- **Enriching OBI makes sense since**

  - OBI is a domain specific ontology already having a lot of background terms pertaining to bioinformatics domain
  - With BFO as an upper level ontology, provides interoperability with other OBO ontologies in this domain
  - OBI provides a well established existing structure and relationships making it easier to add further concepts to support Web service annotations and enrich it.
Enriching OBI with

- Terms to annotate inputs and outputs along with their data format or type wherever required

- Terms to annotate an operation’s functionality, referred to as “objective specification”

- Organizational terms and additional terms that complete the hierarchy, e.g. Scoring matrix was added as a common parent for protein scoring matrix and DNA scoring matrix,

We attempt to model every single term in the WSDL: input, output, operation

More on this in the “Methodology to Enrich OBI”
Outline

- Brief Introduction To OBI (Ontology For Biomedical Investigation)
- Our Role of Enriching OBI for Web Service Annotations
- **EDAM (EMBRACE Data and Methods)**
  - Similarities Between EDAM and OBI, working towards the same GOAL
  - Our work on OBI so far
  - Comparison of OBI and EDAM on the basis of definitions, relations, etc.
- Methodology we follow
- Our Related Projects
  - Galaxy Extensions
  - Suggestion Engine
EDAM (EMBRACE Data and Methods)

Of course we need not tell about it….

- EDAM is an ontology for **bioinformatics tools** and **data**. EDAM provides a controlled vocabulary for **description in semantic terms** of things such as:
  - **Web services** e.g. **WSDL** files
  - **Standalone tools**
  - **Web servers**
  - **Databases**
  - **Ontologies**
  - **Data objects, Data syntax and file formats**

So basically the **semantic annotations Of Bioinformatics Web services, tools, databases and other ontologies** which is what we are moving towards with **OBI with a focus on Web services and tools**.
Outline

- Brief Introduction To OBI (Ontology For Biomedical Investigation)
- Our Role of Enriching OBI for Web Service Annotations
- EDAM
- **Similarities Between EDAM and OBI, working towards the same GOAL**
  - Our work on OBI so far
  - Comparison of OBI and EDAM on the basis of definitions, relations, etc.
- Methodology we follow
- Our Related Projects
  - Galaxy Extensions
  - Suggestion Engine
The Overlap....!!

How EDAM distinguishes different aspects of a web services to annotate, at a higher level.

Let's go over each of these aspects.

The graphs are only for visualization purposes and are not an exact representation of the hierarchy in the Ontology, some intermediate levels have been intentionally skipped for the sake of simplicity.
Data and Format: Basically representing inputs and outputs of a web service/web service operation.
Operation: function or process performed by a tool, for example a WS operation.
The Overlap....!!

Topic: As what a web service does/achieves or implements (If it is an algorithm); EDAMs ‘Topic’ covers additional areas too.
As far as terms under Resource are considered;

There is no close match as it mainly deals with databases and ontologies.
At a coarse level we can say that both the ontologies would model all the important aspects of a web service/tool that need to be annotated viz: Operations, inputs and outputs.
Outline

- Brief Introduction To OBI (Ontology For Biomedical Investigation)
- Our Role of Enriching OBI for Web Service Annotations
- EDAM
- Similarities Between EDAM and OBI, working towards the same GOAL
- Our work on enriching OBI so far
- Comparison of OBI and EDAM on the basis of definitions, relations, etc.
- Methodology we follow
- Our Related Projects
  - Galaxy Extensions
  - Suggestion Engine
Our work on enriching OBI so far…

- We have started our task of enriching OBI by considering important and commonly used Bioinformatics Web Services one by one and modeling terms required for them.

- Web Services considered/covered so far include
  - WUBlast, NCBI Blast
  - ClustalW, T-coffee
  - Signal Peptide, WsFetchbatch (In Progress)

The documentation produced during intermediate stages like Excel sheets and Concept Maps can be found at http://mango.ctegd.uga.edu/jkissingLab/SWS/Wsannotation/index.html

More on this in Methodology Section
Outline

- Brief Introduction To OBI (Ontology For Biomedical Investigation)
- Our Role of Enriching OBI for Web Service Annotations
- EDAM
- Similarities Between EDAM and OBI, working towards the same GOAL
- Our work on enriching OBI so far
- **Comparison of OBI and EDAM on the basis of definitions, relations, etc.**
- Methodology we follow
- Our Related Projects
  - Galaxy Extensions
  - Suggestion Engine
EDAM does not provide all the terms that would be required for a COMPLETE annotation of a WSDL file e.g. that of WUBlast, NCBIBlast and ClustalW.

Neither does it claim to do so

http://edamontology.sourceforge.net/ Says

“The expectation is for EDAM to be used alongside other ontologies for annotation where possible and desirable. For example, an operation that predicts specific features of a molecular sequence could be annotated with terms from SO (Sequence Ontology) for the features.”

Our aim is to either add or import terms to OBI such that it can single handedly supports annotation of majority of important Bioinformatics Web services.
Following are few of the terms that would be required for annotating Blast Web services but we could not find any suitable concepts for these in the EDAM ontology.

- stats (pairwise alignment statistics),
- sensitivity (pairwise alignment sensitivity),
- topcombon (number of top combinations),
- filter (low complexity sequence filter).

To see all the terms that were found missing with respect to Blast and ClustalW Web Services refer to Excel Sheet “Compare.xlsx”
An example of OBI providing more Comprehensive definitions

- **Word size / alignment word size**
  - EDAM: Size of a sequence word
  - OBI: A data item that specifies the alignment word size used to find matches between the sequences. Increase for speed (max = 2 for proteins; 4 for DNA), decrease for sensitivity.

**The terms we are projecting as being added to OBI are actually proposed to OBI and pending approval.**
As far as BLAST and ClustalW Web Services were concerned EDAM

- Lacked terms for Complete annotation of the WSDL file (inputs, outputs and operations)
- EDAM has not provided any terms for annotation of supplementary operations of a web service like GetResults, GetResultTypes etc.. Which too can be of importance in some cases, specially GetResults.

For details on missing terms for these two web services in consideration, refer to Excel Sheet “Compare.xlsx”
Relations Defined/available for use

**EDAM**
- is_a
- concerns
- has_input
- has_output
- has_attribute
- is_source_of
- is_identifier_of
- is_format_of

**OBI**
- is_a
- is_about
- has_input
- has_output
- achieves_planned_objective
- executes <-> is_executed_in
- has_part <-> part of
- is_encoded_in

*OBI has a lot more relations defined but only the above ones are used for Web Service Annotations*

- **EDAM**: has_input, has_attribute, is_format_of and has_output are undefined in a lot of places making their existence void.
Relations Between terms in EDAM

Sequence Database Search Corresponding to a run WUBlast Operation
run Blast Operations and its relationships with neighboring terms (OBI)
OBO format and OWL

- EDAM is in OBO format where as OBI in OWL

- Many restrictions that can be specified using OWL cannot be specified using OBO format, but we wont go into details of it.
Outline

- Brief Introduction To OBI (Ontology For Biomedical Investigation)
- Our Role of Enriching OBI for Web Service Annotations
- EDAM
- Similarities Between EDAM and OBI, working towards the same GOAL
- Our work on enriching OBI so far
- Comparison of OBI and EDAM on the basis of definitions, relations, etc.

**Methodology we follow**

- Our Related Projects
  - Galaxy Extensions
  - Suggestion Engine
Methodology we follow

As mentioned earlier we have started with important and commonly used Bioinformatics Web Services and modeling terms required for their annotation.

Step 1: We begin by analyzing the WSDL file or available documentation of the web service in consideration.

Step 2: We try to populate the Excel Template that we have created and go through the steps specified in the sheet. Refer to: Excel "Template for Annotations"

Step 3: Once the terms are satisfactorily defined in the sheet we move to creating a concept map or Ontology Analysis Diagram for the Web service. All the Concept maps follow the generic model that we have created. (On next slide)
Generic Concept Map / Ontology Analysis Diagram

remote execution of software objective
is_a OBI: objective specification

achieves_planned_objective

has_part

web service
is_a OBI: planned process

operation 1
is_a OBI: planned process

achieves_planned_objective

has_part

operation 2
is_a OBI: planned process

achieves_planned_objective

has_part

operation n
is_a OBI: planned process

achieves_planned_objective

has_part

OBI: objective specification

IAO: information content entity

IAO: information content entity

IAO: information content entity

IAO: information content entity
Continued…

Step 4: We create a specific detailed Concept map for the web service in consideration, such that all the possible relationships of the newly coined terms with other terms are documented in the diagram. This can happen only after a series of discussions within our team making sure we get them right. (refer to Concept Map for ClustalW/Blast)

Step 5: In most cases making of Concept map would give a clear idea of the super class of the term if not, this is the last thing to be found out which demands a sound understanding of the OBI hierarchy as well as the new concept coined.

Step 6: Once this is done the term is ready to be added to our copy of OBI residing on Web Protégé server and send it in for approval from OBI.

Some of the steps can be revisited or can be done iteratively depending the complexity of the Web service.
is there a need to have two different approaches trying to achieve the same goal...??
Outline

- Brief Introduction To OBI (Ontology For Biomedical Investigation)
- Our Role of Enriching OBI for Web Service Annotations
- EDAM
  - Similarities Between EDAM and OBI, working towards the same GOAL
  - Our work on enriching OBI so far
  - Comparison of OBI and EDAM on the basis of definitions, relations, etc.
- Methodology we follow
- **Our Related Projects**
  - Galaxy Extensions
  - Suggestion Engine
Web Service Extensions for Galaxy

- Galaxy is an easy-to-use, open-source, scalable framework for tools and data integration.

- It comes with a bunch of tools integrated into the system for data analysis and also a workflow designer so that a series of analysis tasks can be executed as a workflow.

- We have developed a plugin/tool to be integrated into Galaxy which would let users add SOAP/REST web services as tools to the Galaxy.

- Which means that any bioinformatics tool/application available as a web service can be added as a tool into Galaxy system with few clicks.

- This would make it possible to construct Web Service Workflows (/compositions) as now each web service operation would be a tool which can be dragged onto the canvas just like any other tool.
Three Step Process to add a Web Service – **Step 1**

*Select a WSDL/WADL file of the Web Service to add*

![Image of Galaxy interface with WSDL/WADL file selection](image)

**Input the WSDL URL:**

http://www.ebi.ac.uk/Tools/services/soap/clustalw2?wsdl

*The output data file will contain the operations of the Web service*
Three Step Process to add a Web Service – **Step 2**

- Select the desired operation to add from the set of operations provided by selected Web Service.
Three Step Process to add a Web Service – **Step 3**

**Final Step** Reload Galaxy for the newly added tool to take effect.

---

**Step 3**

Execute step 3 by clicking on Execute button:

- Refresh Galaxy

To use the newly added tool go to Tools on the left side of this screen.

If you want to use the tool once use the tool under Web service tool.

If you want to use the tool in a workflow use the tool under Web Service Workflow tool.

If the tool is not visible click on “Galaxy” on the top left corner of this window to refresh the page.

**NOTE:** Step 3 merely refreshes Galaxy to recognize the newly added tool. After this step, your tool will be registered at the galaxy depending on the usage.
Web Service (operation) added as a tool in Galaxy
A web services Composition/Workflow in Galaxy using the web services (operations) added using our Extensions.
Outline

- Brief Introduction To OBI (Ontology For Biomedical Investigation)
- Our Role of Enriching OBI for Web Service Annotations
- EDAM
- Similarities Between EDAM and OBI, working towards the same GOAL
- Our work on enriching OBI so far
- Comparison of OBI and EDAM on the basis of definitions, relations, etc.
- Methodology we follow

**Our Related Projects**
- Galaxy Extensions
- **Suggestion Engine**
Web Service Composition using Service Suggestions / Suggestion Engine

- Follows a semi-automatic Web service composition approach.

- It ranks all available candidate Web service operations and suggests service operations to a human designer during the process of Web service composition. The ranking scores are based on data mediation, functionality and formal service specifications.

- The suggestion engine basically takes into account the semantic annotations of the web services to reach the final scores.
Preliminary evaluations of suggestion engine support that annotated web services give better results than un-annotated ones for service suggestions.

We are working on towards coupling the suggestion engine with galaxy such that it can provide service suggestions during the process of web service composition in Galaxy.
How can we Collaborate?
Similar approach and similar kind of relationships for runClustalW operation.