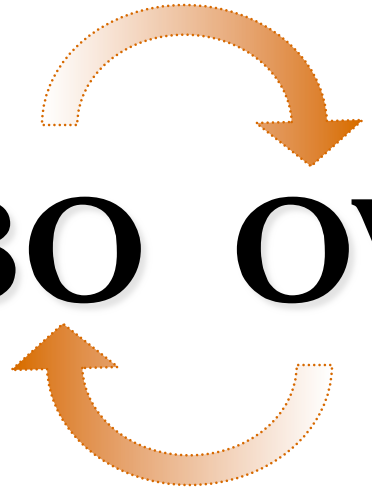
THE UNIVERSITY OF TEXAS AT AUSTIN



OBO OWL

Round Trip Transformation, (starting from OBO)

Prof. Daniel P. Miranker, presenting for
Syed Hamid Tirmizi
Department of Computer Science
The University of Texas at Austin
hamid@cs.utexas.edu

OBO & OWL: The Common Standard Mapping

The result of a standards process

- but without a parent body and associated formal process

Contributors:

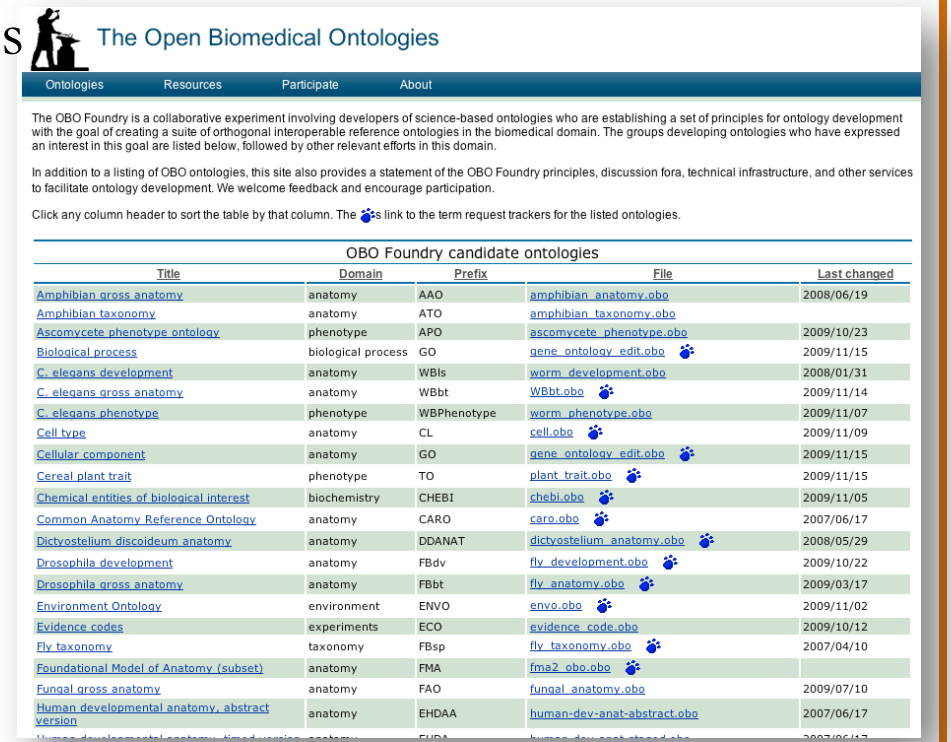
- Syed Hamid Tirmizi (Univ. of Texas at Austin)
- Stuart Aitken (Univ. of Edinburgh)
- Dilvan Moreira (Univ. of São Paulo)
- Chris Mungall (Lawrence Berkeley National Lab.)
- Juan Sequeda (Univ. of Texas at Austin)
- Nigam H. Shah (Stanford University)
- Daniel P. Miranker (Univ. of Texas at Austin)

Goal: Interoperability Between Two Worlds

OBO: Open Biomedical Ontology Language

- Began as the infrastructure for the Gene Ontology (GO)
  - Likely the world's best known and most used ontology
  - Operational Semantics
- ~100 biomedical ontologies of various kinds

OBO Foundry
http://www.obofoundry.org/



The Open Biomedical Ontologies

Ontologies Resources Participate About

The OBO Foundry is a collaborative experiment involving developers of science-based ontologies who are establishing a set of principles for ontology development with the goal of creating a suite of orthogonal interoperable reference ontologies in the biomedical domain. The groups developing ontologies who have expressed an interest in this goal are listed below, followed by other relevant efforts in this domain.

In addition to a listing of OBO ontologies, this site also provides a statement of the OBO Foundry principles, discussion fora, technical infrastructure, and other services to facilitate ontology development. We welcome feedback and encourage participation.

Click any column header to sort the table by that column. The s link to the term request trackers for the listed ontologies.

OBO Foundry candidate ontologies

Title	Domain	Prefix	File	Last changed
Amphibian gross anatomy	anatomy	AAO	amphibian_anatomy.obo	2008/06/19
Amphibian taxonomy	anatomy	ATO	amphibian_taxonomy.obo	
Ascomycete phenotype ontology	phenotype	APO	ascomycete_phenotype.obo	2009/10/23
Biological process	biological process	GO	gene_ontology_edit.obo	2009/11/15
C. elegans development	anatomy	WBls	worm_development.obo	2008/01/31
C. elegans gross anatomy	anatomy	WBbt	WBbt.obo	2009/11/14
C. elegans phenotype	phenotype	WBPhenotype	worm_phenotype.obo	2009/11/07
Cell type	anatomy	CL	cell.obo	2009/11/09
Cellular component	anatomy	GO	gene_ontology_edit.obo	2009/11/15
Cereal plant trait	phenotype	TO	plant_trait.obo	2009/11/15
Chemical entities of biological interest	biochemistry	CHEBI	chebi.obo	2009/11/05
Common Anatomy Reference Ontology	anatomy	CARO	caro.obo	2007/06/17
Dictyostelium discoideum anatomy	anatomy	DDANAT	dictyostelium_anatomy.obo	2008/05/29
Drosophila development	anatomy	FBdv	fly_development.obo	2009/10/22
Drosophila gross anatomy	anatomy	FBbt	fly_anatomy.obo	2009/03/17
Environment Ontology	environment	ENVO	envo.obo	2009/11/02
Evidence codes	experiments	ECO	evidence_code.obo	2009/10/12
Fly taxonomy	taxonomy	FBsp	fly_taxonomy.obo	2007/04/10
Foundational Model of Anatomy (subset)	anatomy	FMA	fma2_obo.obo	
Fungal gross anatomy	anatomy	FAO	fungal_anatomy.obo	2009/07/10
Human developmental anatomy, abstract version	anatomy	EHDAA	human-dev-anat-abstract.obo	2007/06/17

US National Institutes of Health, (NIH), sponsors

- OBO Biomedical ontologies:
  - Model Organisms (Yeast, Drosophila, Mouse, Human…)
    - Anatomy
    - Development
    - Phenotypes
  - Health
    - Human disease nomenclature
    - Including vocabularies that are part of U.S. gov't health care system.
  - Standards
    - Relations Ontology, a form of upper ontology
- Biomedical ontologies often viewed as an ends.

OWL: Web Ontology Language for the Semantic Web

- Semantic Web, a means
  - Inference
    - Formal semantics
    - GUIDs
- Expansive support
  - Piecemeal
    - Academic research groups, various funding sources
    - Companies, various sizes, various funding sources
  - Shepherded by the W3C
- Some important biomedical ontologies are in OWL
  - The NIH, National Cancer Institute (NCI) Thesaurus
  - BioPAX (Biological Pathways Exchange)

The Problem: OBO or the Semantic Web!

- In the Miranker lab: The Morphster project
- Morphster: Image driven ontology editing
  - Productivity tool for systematic biologists
  - Embodies knowledge capture, data integration, workflow
    - Single taxon descriptions
    - Morphological phylogenetic study
- Morphster has to do both:
  - Biodiversity data, (GBIF), OWL adherents
  - Anatomy and phenotypes, OBO

An Informally Gathered Standards Process

- Each contributor has their own motivating story.
- Each contributor initially developed a mapping system.

The Process

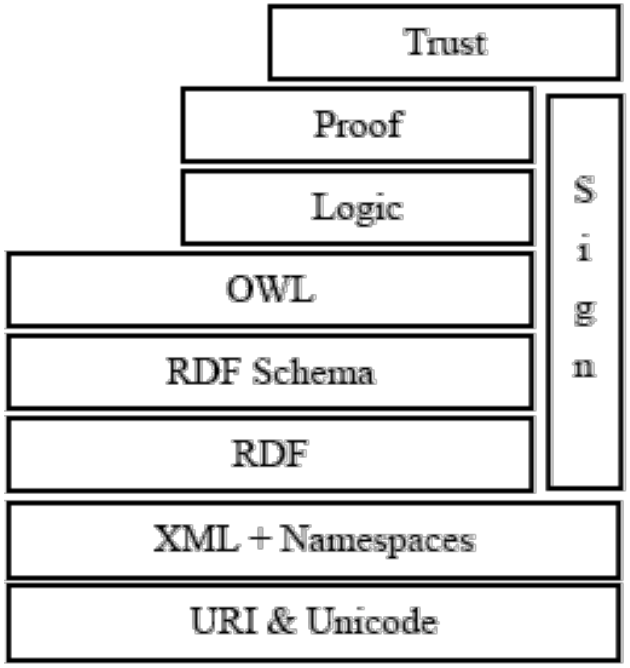
- At the start of 2007,
  - An invitation was broadcast to all interested parties.
  - OBO constructs were enumerated.
  - A shared Google spreadsheet was created,
    - One tab for each contributors mapping
    - One tab for the consensus.
- A wiki page was created for discussion on the mapping
- Artifacts of this process can be found at
  - http://www.bioontology.org/wiki/index.php/OboInOwl:Main_Page

Semantic Web as a Guide

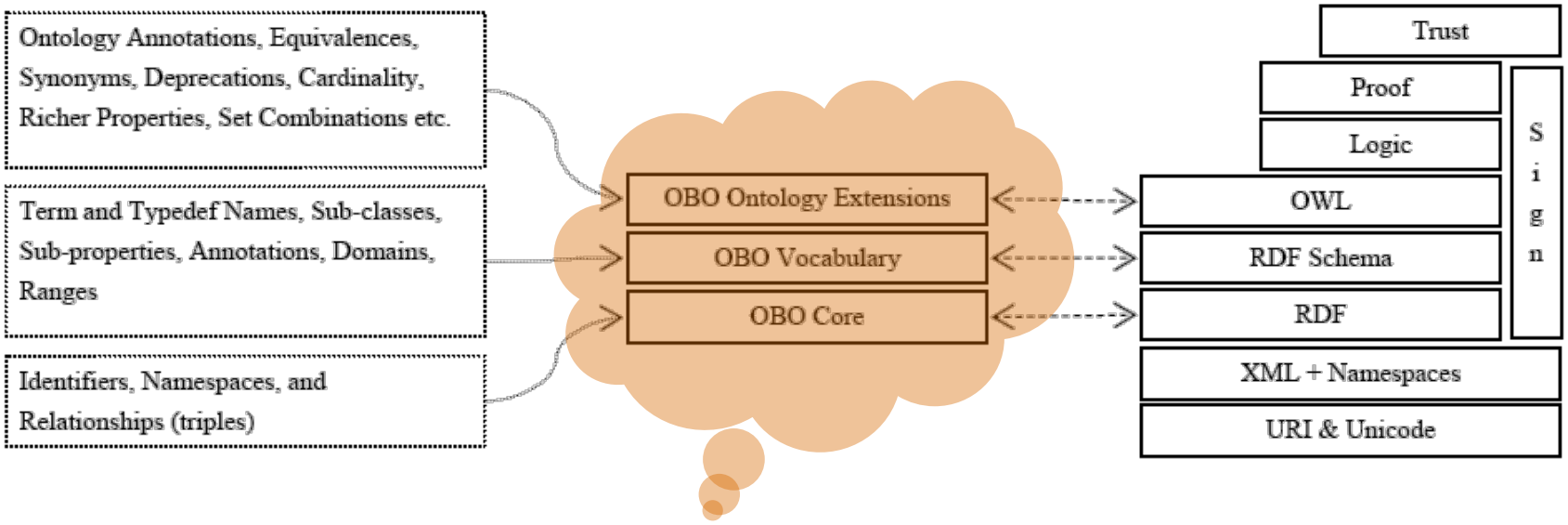
- Thesis:

The organization of the Semantic Web hierarchy (layer cake) transcends the Semantic Web

- Thus, the Semantic Web hierarchy itself can be leveraged to study other systems.



Emerged an Organizing Principle: Two Layer Cakes

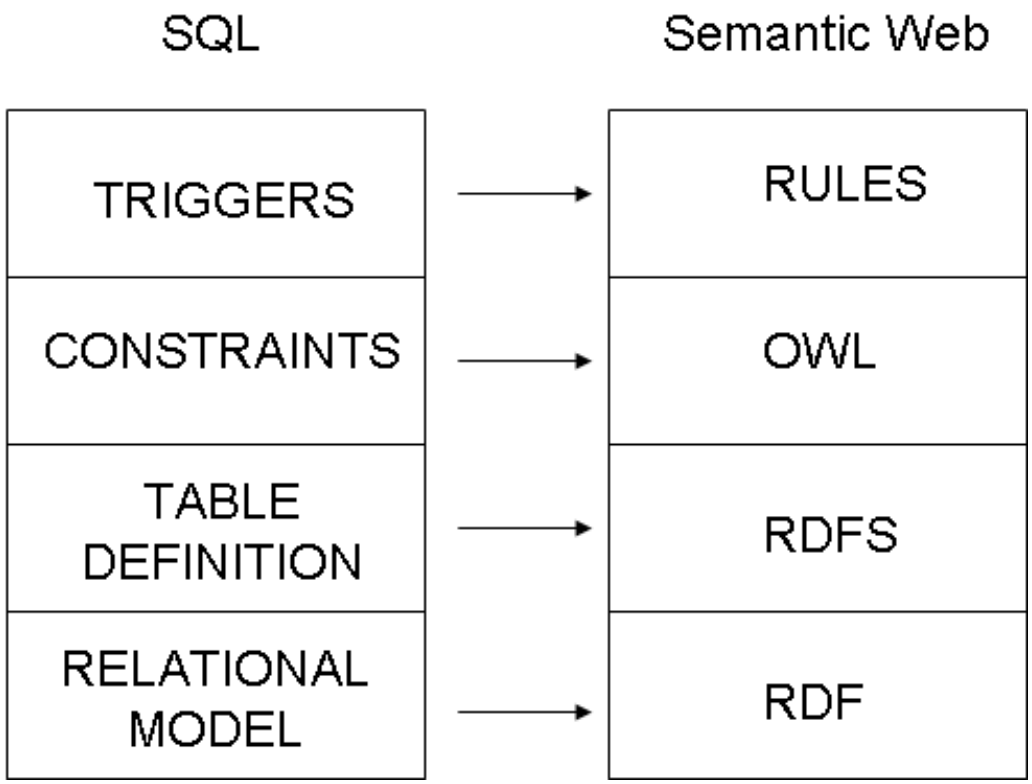


Examples of Elements | OBO Layer Cake | Semantic Web Layer Cake

- Two layer cakes prove useful for identifying:
  - What is the same
  - What is different

about the two systems

Miranker lab has done this for SQL as well



Basis for Ultrawrap: SPARQL endpoint for relational databases

A DR2 like system, except

- completely automatic
- leverages SQL optimizer

Mapping Examples *(What is the same.)*

OBO	OWL
[Typedef] id: part_of name: part of is_transitive: true	<owl:TransitiveProperty rdf:about="…#part_of"> <rdfs:label>part of</rdfs:label> </owl:TransitiveProperty>
Example A Simple transformations: name, transitivity	
[Term] id: ZFA:0000434 name: skeletal system is_a: ZFA:0001439	<owl:Class rdf:about="...#ZFA_0000434"> <rdfs:label>skeletal system</rdfs:label> <rdfs:subClassOf rdf:resource="...#ZFA_0001439"/> </owl:Class>
Example B Transformation of 'is-a'	

Mapping Examples

[Term] id: ZFA:0001439 name: anatomical system relationship: part_of ZFA:0001094	<owl:Class rdf:about= "…#ZFA_0001439"> <rdfs:label>anatomical system</rdfs:label> <rdfs:subClassOf><owl:Restriction> <owl:onProperty rdf:resource = "…#part_of" /> <owl:someValuesFrom rdf:resource = "…#ZFA_0001094" /> </owl:Restriction></rdfs:subClassOf> </owl:Class>
Example C Transformation of a relationship	
[Term] id: ZFA:0000437 name: stomach is_obsolete: true	<owl:Class rdf:about="&oboInOwl;ObsoleteClass"/> <owl:Class rdf:about="...#ZFA_0000437"> <rdfs:label>stomach</rdfs:label> <rdfs:subClassOf rdf:resource="&oboInOwl;ObsoleteClass"/> </owl:Class>
Example D Transformation of obsolete term	

Differences

Required in OWL, missing from OBO

1. **Globally unique identifiers**
   - OBO has a local identification scheme for its concepts
   - OWL classes and properties need global IDs
   - Special consideration required to complete roundtrip

Missing from OWL, part of OBO

2. **Synonyms**
   - Various kinds (possibly) emerging from biomedical domain
   - Lack of semantics and documentation creates problems

3. **Subsets**
   - OWL does not have an exact match to this OBO construct

Mapping OBO IDs to URIs - I

- Any string can be an OBO identifier
- Preferred ID syntax: <IDSPACE>:<LOCALID>
- 'idspace' tag can be added to an OBO ontology header to make GUID possible
  - E.g. "idspace: GO http://www.go.org/owl#"
  - Read: GO is an ID space that refers to the given URI
  - Example: **GO:0000001** maps to **http://www.go.org/owl#GO_0000001**

Mapping OBO IDs to URIs - II

- ID space is *not* defined in the ontology header
  - Each ontology also has a default base URI
  - **<default_base_URI>**: http://www.bioontology.org/…#

- ID is of the form: <IDSPACE>:<LOCALID>
  - Example: **SO:0000001** maps to
    <default_base_URI>SO_0000001

- ID is of the form: <LOCALID>
  - Example: **ABC001** maps to
    <default_base_URI>UNDEFINED_ABC001

Implementation

- Java implementation is a part of official Gene Ontology source
  - http://sourceforge.net/projects/geneontology/
  - Also in tools like OBO-Edit and Morphster

- Web service available for online conversion

  http://www.cs.utexas.edu/~hamid/oboowl.html

  http://www.youtube.com/watch?v=GYnFMq0W_8g

- Already converted OBO Foundry ontologies

  http://www.berkeleybop.org/ontologies/

Towards Formal Semantics for OBO

- OWL is formally defined, while OBO has operational semantics

- A semantics document for OBO can be mechanically derived using the mapping and the semantics for corresponding OWL elements

Editing OBO Ontologies in OWL

- OWL has a larger construct set than OBO
- To make roundtrips possible, any editing in OWL must be limited to a defined set of constructs
  - i.e. the constructs used in the mappings
- We call this subset *OWL-Bio*

Interconnecting OBO and the Semantic Web

- Roundtrip transformations on arbitrary ontology
- Both OBO and OWL are moving targets.

- OWL-Bio, raise awareness

"A problem clearly stated is a problem half solved"

Dorothea Brande

Thank you

Any questions?