Ontological realism as a strategy for integrating ontologies

Ontology Summit February 7, 2013

Barry Smith http://ontology.buffalo.edu/smith

- Scientific data is stored in databases
- There are very few constraints on the creation of new databases
- Scientific data becomes ever more siloed
- Proposal to counteract this silo-formation: create a common non-redundant suite of ontologies covering all scientific domains to annotate scientific data
- Question: What strategy shall we use to build this suite of ontologies

More precisely:

How shall we *build* a common ontology = an ontology that will integrate well with ontologies built for neighboring domains?

This question arises in science, but it arises also in other domains, such as commerce, government and the military

The Semantic Web provides only a small part of the solution

- html demonstrated the power of the Web to allow sharing of information
- can we use semantic technology to create a Web 2.0 which would allow algorithmic reasoning with online information based on XML, RDF and above all OWL (Web Ontology Language)?
- can we use RDF and OWL to break down silos, and create useful integration of on-line data and information



The more Semantic Technology is successful, they more it fails to solve the problem of silos

Indeed it leads to the creation of multiple, new, **semantic silos** – because multiple ontologies are being created in ad hoc ways

| O BioPortal | Browse | Search | Mappings | Recommender | Annotator | Resource Index |
|---|--------|--------|----------|-------------|-----------|----------------|
| Term Search Search for a term in multiple ontologies 2 | | | | | | |
| obesity | | | advance | ed options | | |

Top 143 results in 37 ontologies

Obesity - SNOMED Clinical Terms

http://purl.bioontology.org/ontology/SNOMEDCT/414916001 details - visualize - 26 more from this ontology

Obesity - MedDRA

http://purl.bioontology.org/ontology/MDR/10029883 <u>details</u> - <u>visualize</u> - <u>7 more from this ontology</u>

Obesity - NCI Thesaurus

http://ncicb.nci.nih.gov/xml/owl/EVS/Thesaurus.owl#Obesity Having a high amount of body fat (body mass index [BMI] of 30 or more). details - visualize - 5 more from this ontology

Obesity - Common Terminology Criteria for Adverse Events

http://ncicb.nci.nih.gov/xml/owl/EVS/ctcae.owl#Obesity details - visualize

Obesity - Medical Subject Headings

http://purl.bioontology.org/ontology/MSH/D009765

A status with BODY WEIGHT that is grossly above the acceptable or desirable weight, usually due to accumulation of excess FATS in the body. The standards may vary with age, sex, genetic or cultural background. ...

details - visualize - 3 more from this ontology

obesity - Experimental Factor Ontology

http://www.ebi.ac.uk/efo/EFO_0001073

A status with BODY WEIGHT that is grossly above the acceptable or desirable weight, usually due to accumulation of excess FATS in the body. The standards may vary with age, sex, genetic or cultural background. ...

details - visualize - 1 more from this ontology

Obesity - Human Phenotype Ontology

http://purl.obolibrary.org/obo/HP_0001513

Body weight that is grossly above the acceptable or desirable weight, usually due to accumulation of excess fats in the body. The standards may vary with age, sex, genetic or cultural background. In the body ... <u>details</u> - <u>visualize</u> - <u>1 more from this ontology</u>

Obesity - Cell line ontology

http://purl.obolibrary.org/obo/DOID_9970 details - visualize - 1 more from this ontology

Obesity - Read Codes, Clinical Terms Version 3 (CTV3)

http://purl.bioontology.org/ontology/RCD/C380. details - visualize - 14 more from this ontology

OBESITY - Online Mendelian Inheritance in Man

http://purl.bioontology.org/ontology/OMIM/601665 details - visualize - 24 more from this ontology

obesity - NIFSTD http://purl.obolibrary.org/obo/DOID_9970 details - visualize - 3 more from this ontology

OBESITY - COSTART http://purl.bioontology.org/ontology/CST/OBESITY details - visualize

obesity - Neuro Behavior Ontology

http://purl.obolibrary.org/obo/NBO_0000242 <u>details</u> - <u>visualize</u>

Obesity - ICD10 http://purl.bioontology.org/ontology/ICD10/E66 details - visualize - 2 more from this ontology

OBESITY - WHO Adverse Reaction Terminology

http://purl.bioontology.org/ontology/WHO/0397

details - visualize

Obesity - PMA 2010

http://www.bioontology.org/pma.owl#PMA_936 excessively high accumulation of body fat or adipose tissue in relation to lean body mass; the amount of body fat (or adiposity) includes concern for both the distribution of fat throughout the body and the ... <u>details</u> - <u>visualize</u> - <u>1 more from this ontology</u>

Obesity - Galen http://www.co-ode.org/ontologies/galen#Obesity details - visualize

<u>obesity - eVOC (Expressed Sequence Annotation for Humans)</u> http://purl.obolibrary.org/obo/EV_0600016 <u>details - visualize - 1 more from this ontology</u>

Obesity - Bone Dysplasia Ontology

http://purl.org/obo/owl/HP#HP_0001513 details - visualize - 1 more from this ontology

Obesity - MedlinePlus Health Topics

http://purl.bioontology.org/ontology/MEDLINEPLUS/T61 <u>details</u> - <u>visualize</u> - <u>2 more from this ontology</u>

obesity - Bleeding History Phenotype

http://www.semanticweb.org/ontologies/2010/10/BPO.owl#obesity <u>details</u> - <u>visualize</u> - <u>1 more from this ontology</u>

Obesity - NIF Dysfunction

http://ontology.neuinfo.org/NIF/Dysfunction/NIF-Dysfunction.owl#nlx_dys_20090302 An eating-related disorder in which excess body fat has accumulated to such an extent that health may be negatively affected. It is commonly defined as a body mass index (weight divided by height squared) of ... <u>details</u> - <u>visualize</u> - <u>1 more from this ontology</u>

Obesity - HOM ElixhauserScores http://purl.bioontology.org/ontology/HOM_ElixhauserScores/Class_23 details - visualize

obesity - Physician Data Query

http://purl.bioontology.org/ontology/PDQ/CDR0000467910

details - visualize

There are many ways to create ontologies



How do we build ontologies so that they converge? (The problem of 'knowledge representation')

– III New York State♪ R ywye has ah sh uqwu ja ? r of Excellence in ormatics & Life Sciences.

uw has!

What is going on here ?

There are many ways to create ontologies



Evidence-based ontology development

- Q: What is to serve as constraint?
- A1: Authority (I tell you what to do)
- A2: Homesteading (Founder effect)
- A3: Best candidate terminology But what does 'best' mean?
- A4: Voting ? But then on what grounds should people vote?

A5: Reality, as revealed, incrementally, by which results of experimentation become incorporated into textbook science

Three Levels to Keep Straight

- Level 1: the entities in reality, both instances and universals
- Level 2: cognitive representations of this reality, e.g. on the part of scientists ...
- Level 3: publicly accessible concretizations of these cognitive representations in textual and graphical artifacts



What is going on here ?

The realist approach provides the basis for coordination (consistency) at the level of specific domains – reflecting the consistency of textbook science

But integration requires more than consistency – it requires also a common realist methodology for ontology development

Smith and Ceusters, "Ontological Realism as a Methodology for Coordinated Evolution of Scientific Ontologies", *Applied Ontology*, 5 (2010), 139–188.

Benefits of common methodology

- No need to reinvent the wheel for each new domain
- Can profit from storehouse of lessons learned
- Can more easily reuse what is made by others
- Can more easily reuse training
- Can more easily inspect and criticize results of others' work
- Leads to innovations (e.g. Mireot, Ontofox) in strategies for combining ontologies

Candidate Upper Level Ontologies

- Domain Ontology for Linguistic and Cognitive Engineering (DOLCE)
- Suggested Upper Merged Ontology (SUMO)
- Upper Cyc Ontology
- Basic Formal Ontology

– all reflections of recognized need for semantic standardization via upper level ontology

Why choose BFO

Very small

- Designed to support the consistent representation of different domains of reality in support of scientific research
- integration of data via ontologies presupposes consistent ontologies
- Associated with aggressive program of projectbased testing, feedback and training

Main reason to use BFO

BFO has the largest body of scientist users (compare: This telephone network has the largest number of subscribers)

Snowballing network effects:

data annotated using BFO-conformant ontologies becomes more valuable

numbers of people with expertise in building BFO-conformant ontologies increases

Basic Formal Ontology

- No fictions
- No non-existents
- No 'possible worlds'
- BFO is designed to support scientific research
- Science is distinct from story-telling
- There is no science of unicorns (though they may be a branch of psychiatry which deals with unicorn-obsessions or unicorn-delusions)

Basic Formal Ontology

- a true upper level ontology
- no interference with domain ontologies
- no interference with issues of cognition
- a small subset of DOLCE but with a clearer treatment of instances, universals, relations and qualities, time

| RELATION TO TIME | CONTINUANT | | | | OCCURRENT |
|-----------------------------------|---------------------------------------|--|----------------------------------|-----------------------|----------------------------|
| GRANULARITY | INDEPENDENT | | DEPEN | | |
| ORGAN AND ORGANISM | Örganism (NCBI Taxonomy) | Anatomical Entity (FMA, CARO) | Organ Function (FMP, CPRO) | Phenotypic Quality | Biological Process (GO) |
| CELL AND CELLULAR COMPONENT | Cell (CL) | Cellular Component (FMA, GO) | Cellular Function (GO) | (Paro) | |
| MOLECULE | Molecule (ChEBI, SO, RnaO, PrO) | | Molecular Function (GO) | | Molecular Process (GO) |

The Open Biomedical Ontologies (OBO) Foundry

| RELA TIME | TION TO | OCCURRENT | | | |
|-----------------------------------|--|--|-----------------------------------|-----------------------|-----------------------------------|
| GRANULARI | INDEPENDENT FY | | DEPEN | | |
| ORGAN AND ORGANISM | Organism (NCBI Taxonomy) | Anatomical Entity (FMA, CARO) | Organ Function (FMP, CPRO) | Phenotypic Quality | Organism-Level Process (GO) |
| CELL AND CELLULAR COMPONENT | Cell (CL) | Cellular Component (FMA, GO) | Cellular Function (GO) | (ParO) | Cellular Process (GO) |
| MOLECULE | Molecule (ChEBI, SO, RNAO, PRO) | | Molecular Function (GO) | | Molecular Process (GO) |

rationale of OBO Foundry coverage (homesteading principle)

Users of BFO (Consortia)

| NIF Standard | Neuroscience Information Framework |
|--------------------|---|
| eagle-I Ontologies | used by VIVO and CTSAconnect |
| IDO Consortium | Infectious Disease Ontology |
| CROP | Common Reference Ontologies for Plants |

The OBO Foundry

- GO Gene Ontology
- CL Cell Ontology
- ChEBI Chemical Ontology
- PRO Protein Ontology
- OBI Ontology for Biomedical Investigations
- OGMS Ontology for General Medical Science
- PATO Phenotype (Quality) Ontology
- IDO Infectious Disease Ontology
- PO Plant Ontology
- SO Sequence Ontology
- FMA Foundational Model of Anatomy
- CARO Common Anatomy Reference Ontology
- EnvO Environment Ontology
- Disease Ontology

http://www.ifomis.org/bfo/users

Neurological Disease Ontologies (ND)

Interdisciplinary Prostate Ontology (IPO)

- Nanoparticle Ontology (NPO): Ontology for Cancer
 - Nanotechnology Research
- Neural Electromagnetic Ontologies (NEMO)
- ChemAxiom Ontology for Chemistry
- Ontology for Risks Against Patient Safety (RAPS/REMINE) (EU FP7)

Petrochemicla Ontology

US Army (I2WD)

Example: The Cell Ontology

SUBCLASS EXPLORER

For Project: 🔶 DC_CL

Asserted Hierarchy

