Cell Behavior Ontology Workshop

National Institutes of Health Campus, Bethesda, MD
May 4, 2009

Building Ontologies
Best practices, pitfalls and positives

Olivier Bodenreider

Lister Hill National Center
for Biomedical Communications
Bethesda, Maryland - USA

Outline

- If ontology is the solution, what is the problem?
  - Think *use cases*
- Don’t try this at home!
  - *Ontologies for dummies* hasn’t been written yet
- Where to start?

If ontology is the solution,
what is the problem?

Uses of biomedical ontologies

- Knowledge management
  - Annotating data and resources
  - Accessing biomedical information
  - Mapping across biomedical ontologies
- Data integration, exchange and semantic interoperability
- Decision support
  - Data selection and aggregation
  - Decision support
  - NLP applications
  - Knowledge discovery

[Bodenreider, YBMI 2008]

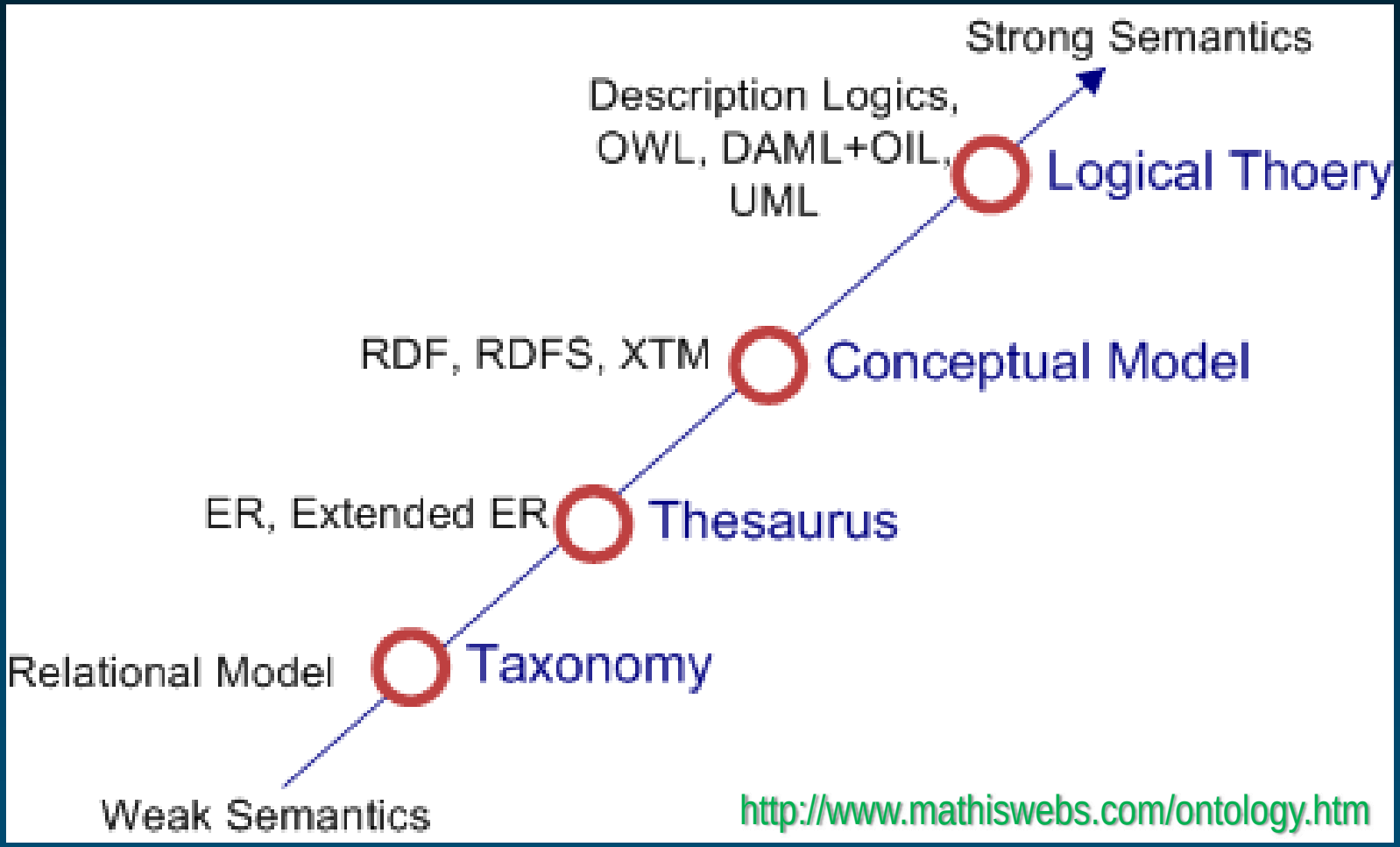
Properties of biomedical ontologies

- Knowledge management
  - Annotating data and resources
  - Accessing biomedical information
  - Mapping across biomedical ontologies
- Data integration, exchange and semantic interoperability
- Decision support
  - Data selection and aggregation
  - Decision support
  - NLP applications
  - Knowledge discovery

Controlled vocabularies

Ontological resources

Ontology “spectrum”



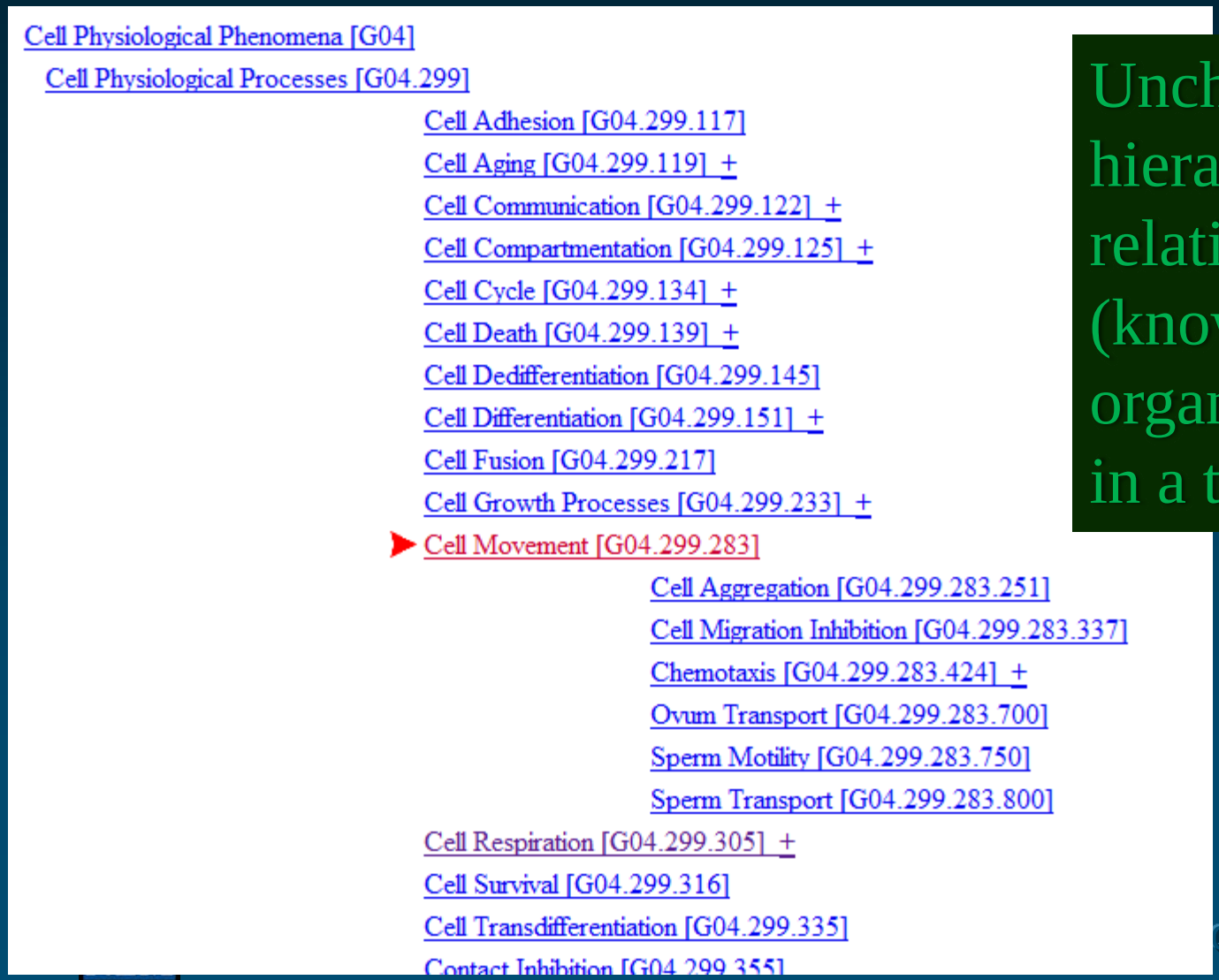
Cell movement in MeSH

MeSH Heading	Cell Movement
Tree Number	G04.299.283
Tree Number	G07.700.560.500.180
Annotation	not for microorganisms; do not confuse with CYTOKINESIS or CYTOPLASMIC STREAMING
Scope Note	The movement of cells from one location to another. Distinguish from CYTOKINESIS which is the process of dividing the CYTOPLASM of a cell.
Entry Term	Cell Migration
Entry Term	Locomotion, Cell
Entry Term	Migration, Cell
Entry Term	Motility, Cell
Entry Term	Movement, Cell
Allowable Qualifiers	DE ES GE IM PH RE
Previous Indexing	Cytology (1966-1967)
Previous Indexing	specific cell or tissue/cytology (1966-1967)
History Note	70(68)
Date of Entry	19990101
Unique ID	D002465

Controlled vocabulary

http://www.nlm.nih.gov/mesh/2009/mesh_browser/MBrowser.html

Cell movement in MeSH



Uncharacterized hierarchical relations (knowledge organization in a thesaurus)

Cell movement in GO

cell motion

Term information ⬇ Term lineage ⬇ External references ⬇ 2568 gene product associations ➡

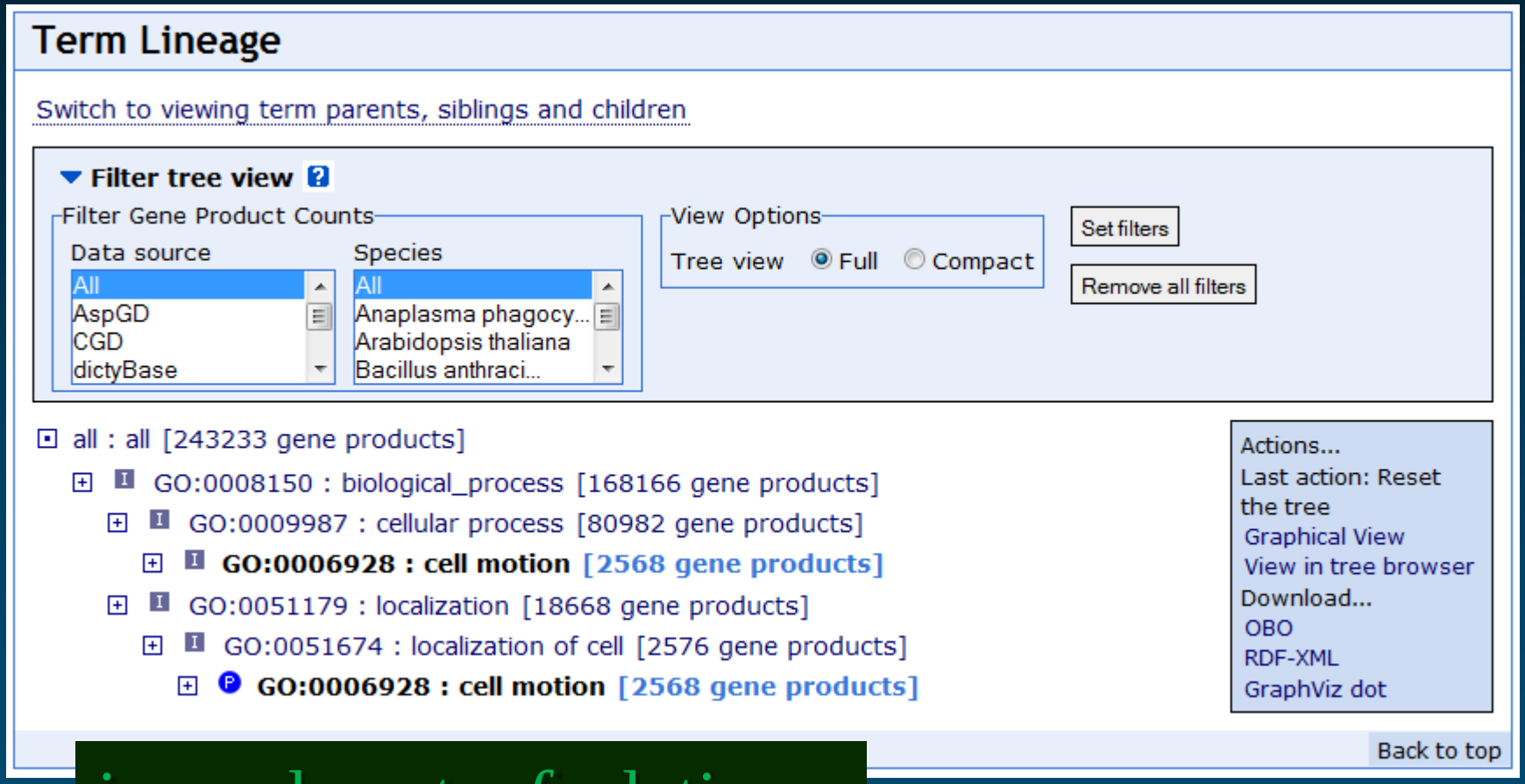
Term Information

Accession	GO:0006928
Ontology	**biological process**
Synonyms	**related**: cell locomotion **exact**: cell movement
Definition	Any process involved in the controlled movement of a cell. [source: GOC:dgh, GOC:dph, GOC:jl, GOC:mlg]
Comment	None
Subset	goslim_goa goslim_pir Unavailable
Community	There have been **0** comments for this term. If you would like to view or participate in the community annotation, please continue to the GONUTS page.

http://amigo.geneontology.org/cgi-bin/amigo/search.cgi

Back to top

Cell movement in GO



isa and *part_of* relations

Thesaurus vs. Ontology

- Define *use cases*
  - Typical situations in which the resource to be created is expected to contribute to the solution
    - Resource annotation (controlled vocabulary)
      - Lexical aspects (e.g., synonyms, variants) are important
    - Resource classification/organization (thesaurus)
    - Inference based on the attributes of biological entities (ontology)
  - Competency questions
  - Minimal ontological commitment (just enough information for the task at hand)

Don’t try this at home!

Ontologies vs. cars

What is the difference between an ontology and a car?

Dependent continuants, existential quantification and rigidity

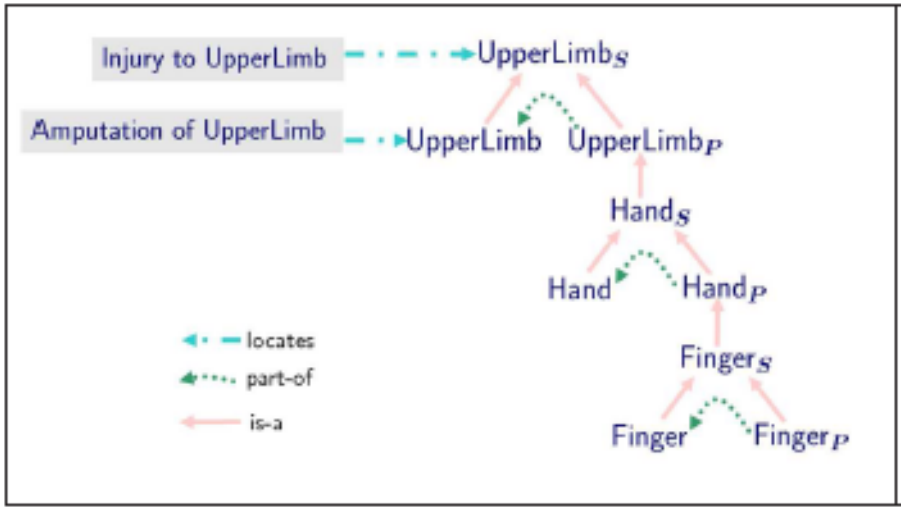
A **dependent continuant** is a quality or a realizable entity. A **quality** is something that all objects of a particular type have for all of the time they exist, for example, the mass of a bag of sugar, the shape of a hand, the fragility of a cup, the beauty of a view, the brightness of a light, and the smell of the ocean. While these can change, the bag of sugar always has a mass and the hand always has a shape. This is contrasted with a **realizable entity** where the value does not need to exist, the existence can change though time, for example, the role of being a teacher or the disposition of metal to conduct electricity. A realizable entity is either

- a **function** that specifies the purpose of a object, for example, the function of a cup may be to hold coffee, the function of the heart is to pump blood.
- a **role** specifies a goal that is not essential to the object's design, but can be carried out, for example, the role of being a judge, the role of delivering coffee.
- a **disposition** is something that can happen to an object, for example, the disposition of a cup to break if dropped, the disposition of vegetables to rot if not refrigerated, the disposition of matches to light if they are not wet.

BFO, presented in http://www.cs.ubc.ca/~poole/aibook/html/ArtInt_313.html

Formal ontology (philosophy)

Dependent continuants, existential quantification and rigidity



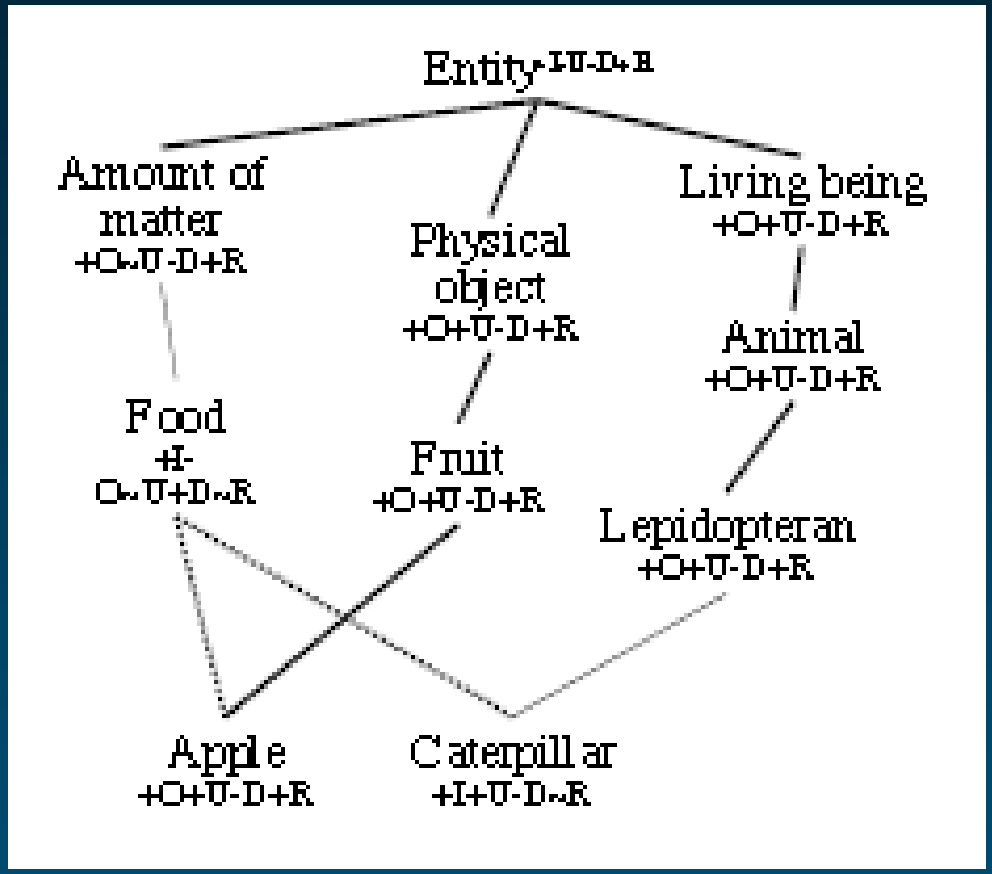
$$
\begin{aligned}
\text{UpperLimb} &\sqsubseteq \text{UpperLimb}_S \\
\text{UpperLimb}_P &\sqsubseteq \text{UpperLimb}_S \sqcap \\
&\quad \exists \text{part-of}.\text{UpperLimb} \\
\text{Hand}_S &\sqsubseteq \text{UpperLimb}_P \\
\text{Hand} &\sqsubseteq \text{Hand}_S \\
\text{Hand}_P &\sqsubseteq \text{Hand}_S \sqcap \exists \text{part-of}.\text{Hand} \\
\text{Finger}_S &\sqsubseteq \text{Hand}_P \\
\text{Finger} &\sqsubseteq \text{Finger}_S \\
\text{Finger}_P &\sqsubseteq \text{Finger}_S \sqcap \exists \text{part-of}.\text{Finger}
\end{aligned}
$$

Fig. 1. Complete SEP-triplets in SNOMED CT.

[Suntisrivaraporn, AIME 2007]

Knowledge representation

Dependent continuants, existential quantification and rigidity



Formal ontology (philosophy)

OntoClean, http://www.loa-cnr.it/Papers/GuarinoWeltyOntoCleanv3.pdf

Instances, classes, collectives

- *"Lmo-2 interacts with Elf-2"*
  - One individual Lmo-2 molecule interacts with one individual Elf-2 molecule.
  - A collective of Lmo-2 molecules interacts with one individual Elf-2 molecule.
  - One individual Lmo-2 molecule interacts with a collective of Elf-2 molecules.
  - A collective of Lmo-2 molecules interacts with a collective of Elf-2 molecules.
  - […]

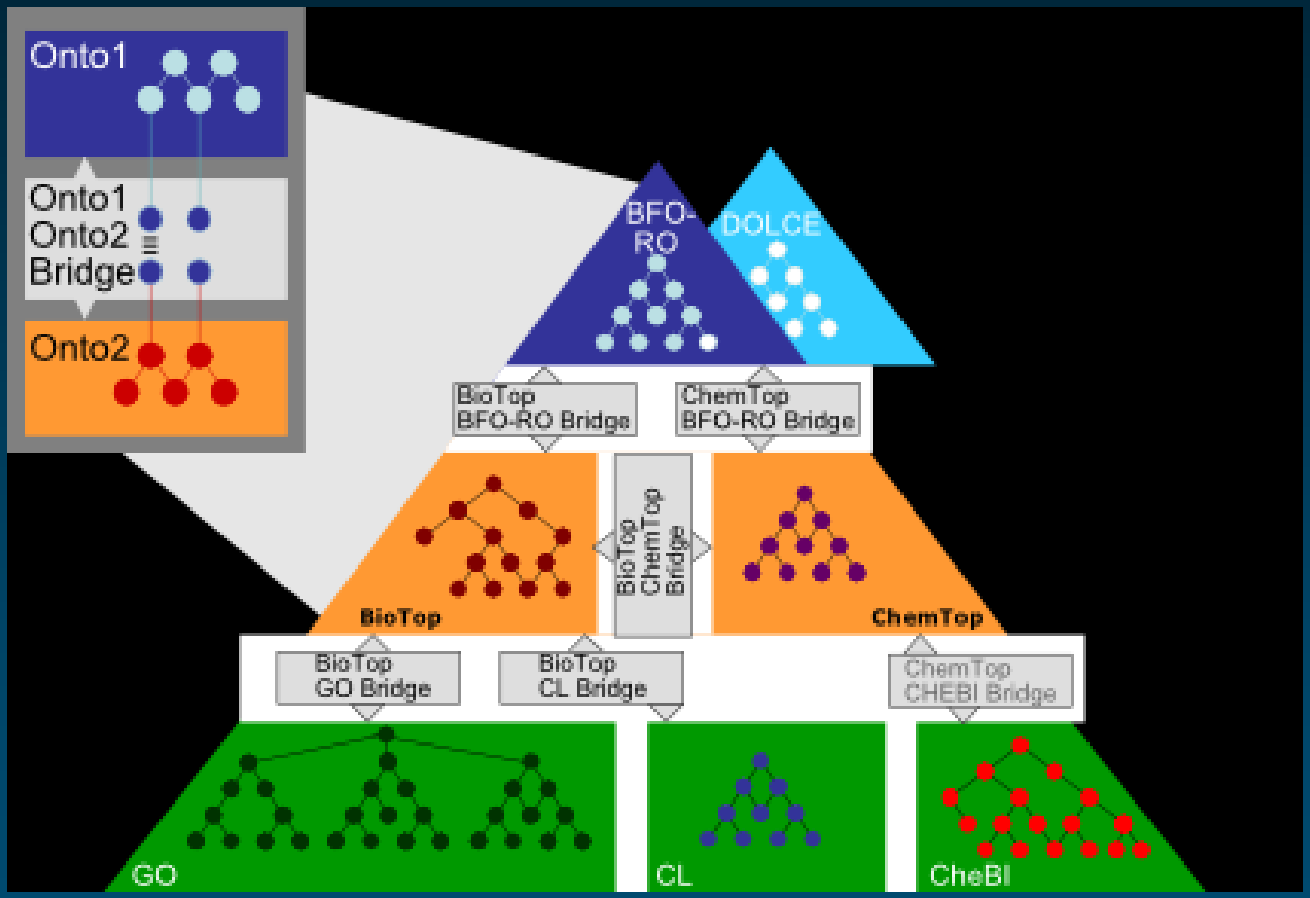
[Schulz and Landsen, Applied ontology 2009]

Top-level ontologies

- Basic Formal Ontology (BFO)
- DOLCE
- BioTop

- Ground domain ontologies into sound philosophical foundations
- Difficult to understand for “folks form the trenches”

A pyramid of ontologies



BioTop, http://www.imbi.uni-freiburg.de/biotop/

Power tools for ontologies

- Ontology editors
  - Protégé
  - OBO-Edit

Too complex for most biologists

- Semantic wikis
  - Intermediate representations
  - Collaborative development

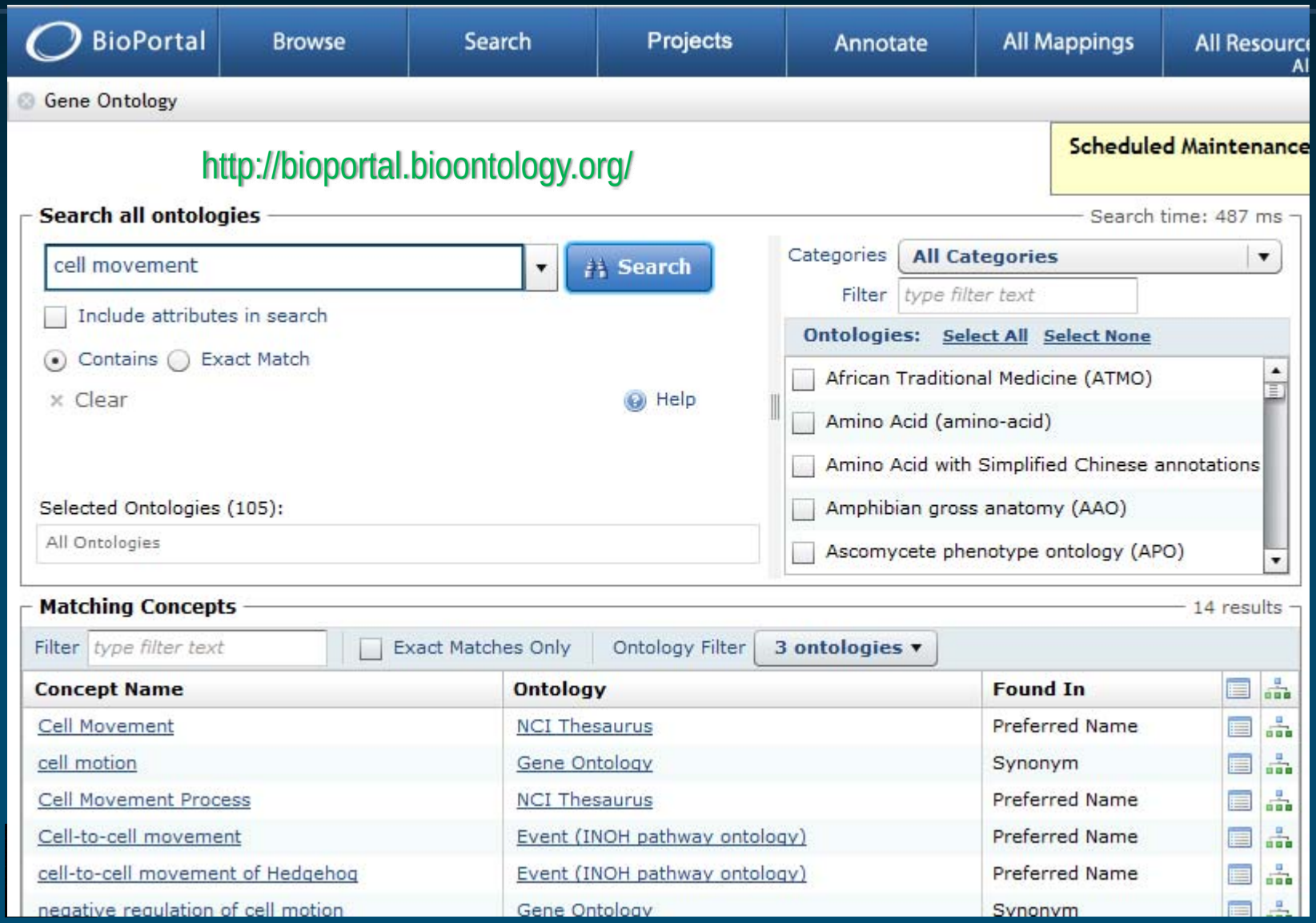
Simplified representations

Where to start?

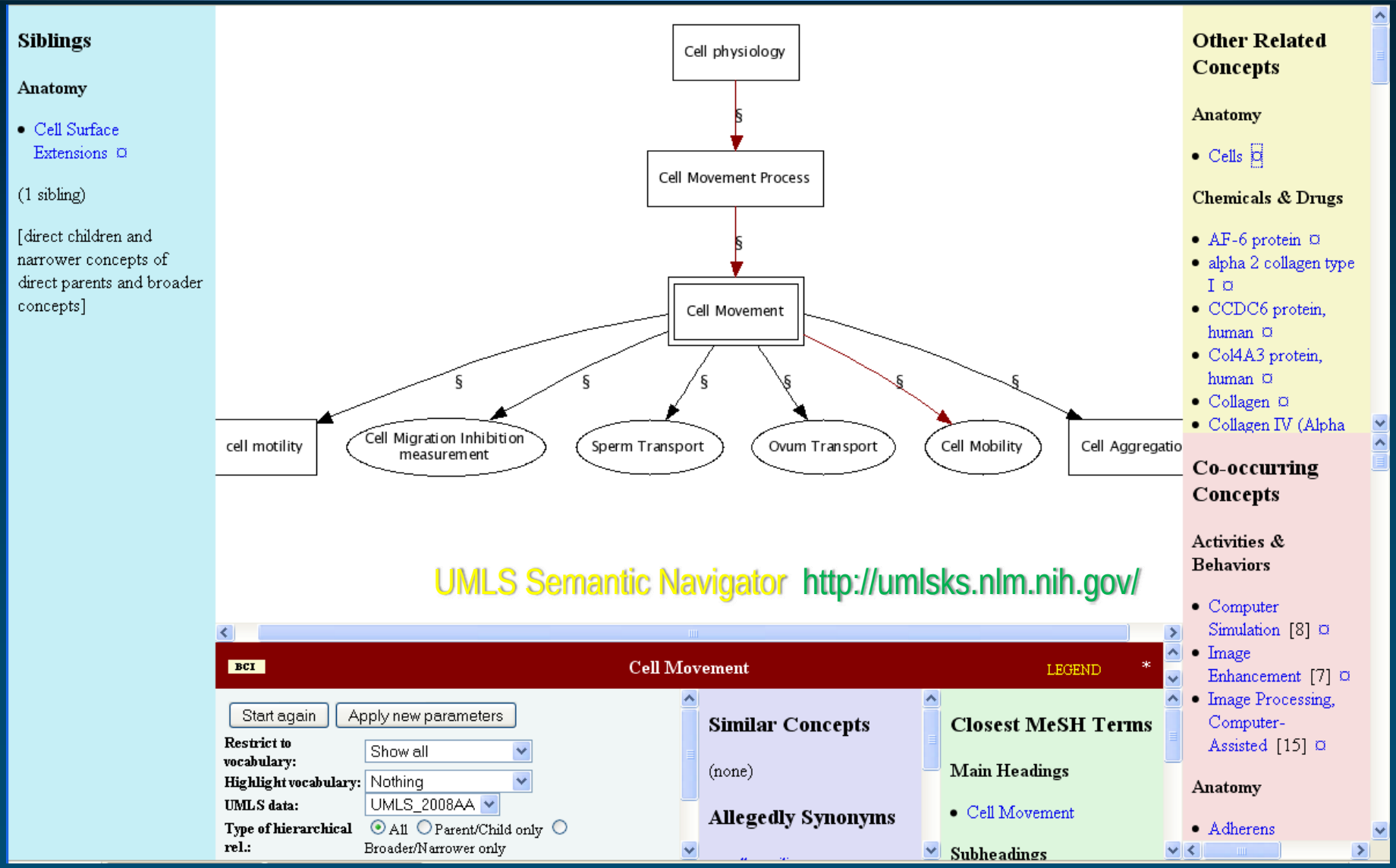
Collect entities from the universe of discourse

- From experts
  - Workshops
- From textual corpora
  - Manually
  - Automatic term extraction
- From existing terminologies and ontologies

Shop around Ontology repositories



Shop around Ontology repositories



Link to/Borrow from existing ontologies

- Advantages
  - Avoid reinventing the wheel
  - Benefit from the experience of specialists of a given subdomain
- Disadvantages
  - Borrow ontological commitment from these ontologies
  - Might align (or not) with the ontological commitment in your ontology

Decide on standards and tools

- With the help of *experienced ontologists*
- For knowledge representation
  - e.g., OWL
- For editing ontologies
  - e.g., Protégé
- For ontological commitment
  - e.g., top-level ontology, relation ontology

Guidelines for ontology development

- OBO Foundry principles
  - *Openness*
  - *Common shared syntax* (OBO or OWL)
  - *Unique identifier space* within the OBO Foundry
  - *Versioning mechanism*
  - Clearly specified and clearly *delineated content*
  - Textual *definitions* for all terms
  - Use relations from the *OBO Relation Ontology*
  - The ontology is *well documented*
  - Plurality of independent *users*
  - *Collaborative development* with other OBO Foundry members

[Smith et al., Nature Biotechnology 2007]

http://www.obofoundry.org/crit.shtml

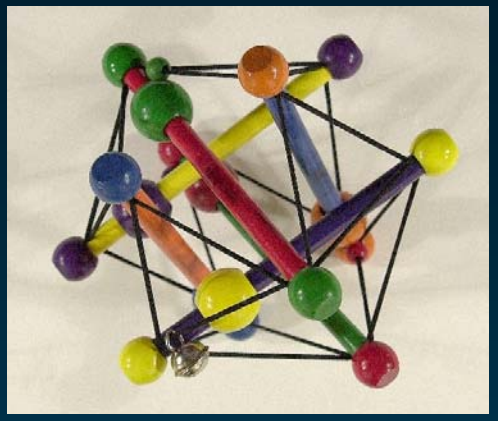
Learn from others Three recent projects

- BiomedGT
  - National Cancer Institute
  - Semantic wiki approach
  - http://biomedgt.nci.nih.gov/wiki/index.php/Main_Page
- Infectious Disease Ontology
  - OBO Foundry approach
  - http://www.infectiousdiseaseontology.org/Home.html
- International Classification of Diseases (ICD11 revisions)
  - Semantic wiki approach + Protégé background
  - http://www.who.int/classifications/icd/ICDRevision/en/index.html
- Neuroscience Information Framework
  - http://neuinfo.org/

Take home points

Take home points

- Start by defining use cases, not ontologies
  - Define and measure success
- Let the biologists be biologists
  - Seek the assistance of ontologists when dealing with
    - Top-level ontology
    - Formalism
    - Complex ontology editors
- Follow experience/guidelines, not gurus
  - NIF, BiomedGT, ICD11 revisions
  - Ontology Foundry
- Think prospectively
  - Maintenance
  - Funding (beyond short term)



Medical Ontology Research

Contact: olivier@nlm.nih.gov
Web: mor.nlm.nih.gov

NATIONAL LIBRARY OF MEDICINE

Olivier Bodenreider

Lister Hill National Center
for Biomedical Communications
Bethesda, Maryland - USA

Additional references

- Bodenreider O. *Biomedical ontologies in practice* Short course at the University of Utah, Department of Biomedical Informatics, Salt Lake City, Utah, June 9-11, 2008.
  http://mor.nlm.nih.gov/pubs/pres/080609-11_bioontologies_in_practice.pdf
- Smith B. *Introduction to Biomedical Ontologies – A training course in eight lectures (video)*
  http://ontology.buffalo.edu/smith/Ontology_Course.html

- International Conference on Biomedical Ontology University at Buffalo, NY · July 24-26, 2009
  http://icbo.buffalo.edu/