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
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
Ontology “spectrum”

http://www.mathiswebs.com/ontology.htm
# Cell movement in MeSH

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

Cell movement in MeSH

- Cell Physiological Phenomena [G04]
  - Cell Physiological Processes [G04.299]
    - Cell Adhesion [G04.299.117]
    - Cell Aging [G04.299.119] +
    - Cell Communication [G04.299.122] +
    - Cell Compartmentation [G04.299.125] +
    - Cell Cycle [G04.299.134] +
    - Cell Death [G04.299.139] +
    - Cell Dedifferentiation [G04.299.145]
    - Cell Differentiation [G04.299.151] +
    - Cell Fusion [G04.299.217]
    - Cell Growth Processes [G04.299.233] +
    - Cell Movement [G04.299.283]
      - Cell Aggregation [G04.299.283.251]
      - Cell Migration Inhibition [G04.299.283.337]
      - Chemotaxis [G04.299.283.424] +
      - Ovum Transport [G04.299.283.700]
      - Sperm Motility [G04.299.283.750]
      - Sperm Transport [G04.299.283.800]
    - Cell Respiration [G04.299.305] +
    - Cell Survival [G04.299.316]
    - Cell Transdifferentiation [G04.299.335]
    - Contact Inhibition [G04.299.355]
Cell movement in GO

**cell motion**

**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
Cell movement in GO

Term Lineage

Switch to viewing term parents, siblings and children

Filter tree view

Filter Gene Product Counts

Data source

Species

- View Options

Tree view

Set filters

Remove all filters

Actions...

Last action: Reset the tree
Graphical View
View in tree browser
Download...
OBO
RDF-XML
GraphViz dot

Back to top

is a 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?
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 an 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/albook/html/ArtInt_313.html

Formal ontology (philosophy)
Dependent continuants, existential quantification and rigidity

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

[Suntisrivaraporn, AIME 2007]
Dependent continuants, existential quantification and **rigidity**

**Formal ontology (philosophy)**

**OntoClean**, http://www.loa-cnr.it/Papers/GuarinoWeltyOntoCleanv3.pdf
Instances, classes, collectives

“\textit{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

- Semantic wikis
  - Intermediate representations
  - Collaborative development

Too complex for most biologists

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

http://bioportal.bioontology.org/

Search all ontologies

- cell movement

Selected Ontologies (105):
- All Ontologies

Matching Concepts

<table>
<thead>
<tr>
<th>Concept Name</th>
<th>Ontology</th>
<th>Found In</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell Movement</td>
<td>NCI Thesaurus</td>
<td>Preferred Name</td>
</tr>
<tr>
<td>cell movement</td>
<td>Gene Ontology</td>
<td>Synonym</td>
</tr>
<tr>
<td>Cell Movement Process</td>
<td>NCI Thesaurus</td>
<td>Preferred Name</td>
</tr>
<tr>
<td>Cell-to-cell movement</td>
<td>Event (INOH pathway ontology)</td>
<td>Preferred Name</td>
</tr>
<tr>
<td>cell-to-cell movement of Hedgehog</td>
<td>Event (INOH pathway ontology)</td>
<td>Preferred Name</td>
</tr>
<tr>
<td>negative regulation of cell motion</td>
<td>Gene Ontology</td>
<td>Synonym</td>
</tr>
</tbody>
</table>
Shop around Ontology repositories


Lister Hill National Center for Biomedical Communications
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

- **Infectious Disease Ontology**
  - OBO Foundry approach
  - [http://www.infectiousdiseaseontology.org/Home.html](http://www.infectiousdiseaseontology.org/Home.html)

- **International Classification of Diseases (ICD11 revisions)**
  - Semantic wiki approach + Protégé background

- **Neuroscience Information Framework**
  - [http://neuinfo.org/](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)
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.  

- Smith B. *Introduction to Biomedical Ontologies – A training course in eight lectures (video)*  
  [http://ontology.buffalo.edu/smith/Ontology_Course.html](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/](http://icbo.buffalo.edu/)