If we build it, will they come?
Social engineering of new technology to disseminate biomedical ontologies

Mark A. Musen and the BioPortal Team
Stanford University
Thanks to a ton of people!

• Benjamin Dai
• Misha Dorf
• Nick Griffith
• Suzanna Lewis
• Dilvan Moreira
• Michael Montegut

• Chris Mungall
• Natasha Noy
• Kaustubh Supekar
• Nicole Washington
• Daniel Rubin
• Nigam Shah
<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>724</td>
<td>Unspecified disorders of the back</td>
</tr>
<tr>
<td>724.0</td>
<td>Spinal stenosis, other than cervical</td>
</tr>
<tr>
<td>724.00</td>
<td>Spinal stenosis, unspecified region</td>
</tr>
<tr>
<td>724.01</td>
<td>Spinal stenosis, thoracic region</td>
</tr>
<tr>
<td>724.02</td>
<td>Spinal stenosis, lumbar region</td>
</tr>
<tr>
<td>724.09</td>
<td>Spinal stenosis, other</td>
</tr>
<tr>
<td>724.1</td>
<td>Pain in thoracic spine</td>
</tr>
<tr>
<td>724.2</td>
<td>Lumbago</td>
</tr>
<tr>
<td>724.3</td>
<td>Sciatica</td>
</tr>
<tr>
<td>724.4</td>
<td>Thoracic or lumbosacral neuritis</td>
</tr>
<tr>
<td>724.5</td>
<td>Backache, unspecified</td>
</tr>
<tr>
<td>724.6</td>
<td>Disorders of sacrum</td>
</tr>
<tr>
<td>724.7</td>
<td>Disorders of coccyx</td>
</tr>
<tr>
<td>724.70</td>
<td>Unspecified disorder of coccyx</td>
</tr>
<tr>
<td>724.71</td>
<td>Hypermobility of coccyx</td>
</tr>
<tr>
<td>724.71</td>
<td>Coccygodynia</td>
</tr>
<tr>
<td>724.8</td>
<td>Other symptoms referable to back</td>
</tr>
<tr>
<td>724.9</td>
<td>Other unspecified back disorders</td>
</tr>
</tbody>
</table>
The NCI Thesaurus in Protégé-OWL
GO:0003673: Gene Ontology (92932)
  □ GO:0008150: biological process (56952)
    + GO:0007610: behavior (566)
    . GO:0000004: biological process unknown (6152)
  □ GO:0007154: cell communication (11916)
    + GO:0007155: cell adhesion (830)
    . GO:0030260: cell invasion (0)
    + GO:0008037: cell recognition (210)
  □ GO:0007267: cell-cell signaling (1318)
    + GO:0045168: cell-cell signaling involved in cell fate commitment (0)
  □ GO:0030072: peptide hormone secretion (6)
    . GO:0030252: growth hormone secretion (2)
    . GO:0030073: insulin secretion (4)
    . GO:0030103: vasopressin secretion (2)
    + GO:0019226: transmission of nerve impulse (688)
  □ GO:0030383: host-pathogen interaction (12)
Goals of Biomedical Ontologies

• To provide a classification of biomedical entities
• To annotate data to enable summarization and comparison across databases
• To provide for semantic data integration
• To drive NLP systems
• To simplify the engineering of complex software systems
• To provide a formal specification of biomedical knowledge
# Open Biomedical Ontologies library

<table>
<thead>
<tr>
<th>Domain</th>
<th>Prefix</th>
<th>Ontology</th>
<th>Defs file</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arabidopsis gross anatomy</td>
<td>TAIR</td>
<td>arabidopsis anatomy.ontology</td>
<td>arabidopsis anatomy.definitions</td>
</tr>
<tr>
<td>Arabidopsis development</td>
<td>TAIR</td>
<td>arabidopsis development.ontology</td>
<td>arabidopsis development.definitions</td>
</tr>
<tr>
<td>Cell type</td>
<td>CL</td>
<td>cell.obo</td>
<td>included in cell.obo</td>
</tr>
<tr>
<td>Cereal plant gross anatomy</td>
<td>GRO</td>
<td>anatomy gr ont</td>
<td>anatomy gr def</td>
</tr>
<tr>
<td>Cereal plant development</td>
<td>GRO</td>
<td>temporal gr ont</td>
<td>temporal gr def</td>
</tr>
<tr>
<td>Cereal plant trait ontology</td>
<td>TO</td>
<td>trait ontology</td>
<td>trait definitions</td>
</tr>
<tr>
<td>Chemical entities of biological interest</td>
<td>CHEBI</td>
<td>ontology.obo</td>
<td>included in ontology.obo</td>
</tr>
<tr>
<td>Protein covalent bond</td>
<td>CV</td>
<td>[none]</td>
<td>[none]</td>
</tr>
<tr>
<td>Protein-protein Interaction</td>
<td>MI</td>
<td>psi-mi.dag</td>
<td>psi-mi.def</td>
</tr>
<tr>
<td>Maize gross anatomy</td>
<td>ZEA</td>
<td>Zea mays anatomy ontology.txt</td>
<td>Zea mays anatomy ontology definitions.txt</td>
</tr>
<tr>
<td>Dictyostelium anatomy</td>
<td>DDANAT</td>
<td>anatomy.obo</td>
<td>anatomy.definitions</td>
</tr>
<tr>
<td>Drosophila gross anatomy</td>
<td>FBbt</td>
<td>fly anatomy.obo</td>
<td>fly anatomy.definitions</td>
</tr>
<tr>
<td>Habronattus courtship</td>
<td></td>
<td>protege source</td>
<td>included in protege source</td>
</tr>
<tr>
<td>Loggerhead nesting</td>
<td></td>
<td>protege source</td>
<td>included in protege source</td>
</tr>
<tr>
<td>Human anatomy and development</td>
<td>EV</td>
<td>ontologies</td>
<td>[none]</td>
</tr>
<tr>
<td>Microarray experimental conditions</td>
<td></td>
<td>MGEDOntology.daml</td>
<td>included in MGEDOntology.daml</td>
</tr>
<tr>
<td>Physical-chemical methods and properties</td>
<td>FIX</td>
<td>fixontology</td>
<td>[none]</td>
</tr>
<tr>
<td>Fungal gross anatomy</td>
<td>FAO</td>
<td>fungal anatomy.obo</td>
<td>fungal anatomy.definitions</td>
</tr>
<tr>
<td>Molecular function</td>
<td>GO</td>
<td>gene_ontology.obo</td>
<td>included in gene_ontology.obo</td>
</tr>
<tr>
<td>Biological process</td>
<td>GO</td>
<td>gene_ontology.obo</td>
<td>included in gene_ontology.obo</td>
</tr>
<tr>
<td>Cellular component</td>
<td>GO</td>
<td>gene_ontology.obo</td>
<td>included in gene_ontology.obo</td>
</tr>
</tbody>
</table>
OLS - Ontology Lookup Service

Enter Ontology Term

Search Ontology: Arabidopsis Development [TAIR]

Term Name: (Include obsolete terms ✓)  Term ID:

stem

Additional information:

Enter a partial search term. As you are typing, you will see suggested terms that match what are entering in the form. If you select one from the pull-down list, its corresponding ID will be displayed in the form. If you see "... and more" in the list of suggested values, you can select this value to be redirected to a page where all possible values are listed. As an example, enter mitoc in the Term Name box while the Gene Ontology ontology is selected.

For better search results, do not type punctuation or symbols. For example, if you are looking for 4′-(L-tryptophan), try typing 4 L tryp.

You can browse an ontology by clicking on the "browse" button next to the ontology selector. To view the complete ontology, do not select a term name. If a term name has been selected, it will be the root from which the ontology will be browsed.

Simple Term ID Search:

Term ID:  Search

Enter a complete term ID (example: GO:0008150) and click on the 'Search' button to quickly obtain all pertinent information for this term. Searches are case-sensitive, so ensure that the proper ontology prefix is used (GO:, rather than go: or Go:).
In biology, lots of ontology developers are almost hobbyists

- Nearly always, ontologies are created to address pressing practical needs
- The people who have the most insight into professional knowledge of a given biomedical domain may have little appreciation for metaphysics, principles of knowledge representation, or computational logic
- There simply aren’t enough good ontologists to go around
Issues in assuring ontology quality

• Unlike the case with journal submissions, it makes no sense for ontologies to be peer-reviewed by just a handful of experts
• Open, community-based review of ontologies may be haphazard and chaotic
• Top–down solutions may offer rigid review criteria at the expense of scalability
• There is a pressing need for empirical evaluation of methods for ontology evaluation
A Curated Approach for Quality Assurance

• A proposal to create a family of interoperable “gold standard” biomedical reference ontologies

• Formulated by Barry Smith and members of the GO Consortium

• A Good Housekeeping Seal of Approval for biomedical ontologies
For an ontology to be accepted as one of the Open Biomedical Ontologies in the Foundry, the following criteria must be met (further principles will be added over time):

Version as of 24 April 2006

1. The ontology must be **open** and available to be used by all without any constraint other than (a) its origin must be acknowledged and (b) it is not to be altered and subsequently redistributed under the original name or with the same identifiers.

   The OBO ontologies are for sharing and are resources for the entire community. For this reason, they must be available to all without any constraint or license on their use or redistribution. However, it is proper that their original source is always credited and that after any external alterations, they must never be redistributed under the same name or with the same identifiers.

2. The ontology is in, or can be expressed in, a **common shared syntax**. This may be either the OBO syntax, extensions of this syntax, or OWL.

   The reason for this is that the same tools can then be usefully applied. This facilitates shared software implementations. This criterion is not met in all of the ontologies currently listed, but we are working with the ontology developers to have them available in a common OBO syntax.

3. The ontologies possess a **unique identifier space** within the OBO Foundry.

   The source of concepts from any ontology can be immediately identified by the prefix of the identifier of each concept. It is, therefore, important that this prefix be unique.

4. The ontology provider has procedures for identifying distinct successive **versions**.

5. The ontology has a clearly specified and clearly **delineated content**.

   The ontology must be orthogonal to other ontologies already lodged within OBO.

   The major reason for this principle is to allow two different ontologies, for example anatomy and process, to be combined through additional relationships. These relationships could then be used to constrain when terms could be jointly applied to describe complementary (but distinguishable) perspectives on the same biological or medical entity.

   As a corollary to this, we would strive for community acceptance of a single ontology for one domain, rather than encouraging rivalry between ontologies.

6. The ontologies include **textual definitions** for all terms.

   Many biological and medical terms may be ambiguous, so concepts should be defined so that their precise meaning within the context of a particular ontology is clear to a human reader.

7. The ontology uses relations which are unambiguously defined following the pattern of definitions laid down in the **OBO Relation Ontology**.

8. The ontology is **well documented**.

9. The ontology has a plurality of independent **users**.

10. The ontology will be developed **collaboratively** with other OBO Foundry members.
OBO Foundry must address lots of questions

• Can the top–down approach scale? How many ontologies can be managed by a small panel of curators?
• Who gets to reject an ontology on the basis of form or content? What is the appeals process? How do we know whom to believe?
• Who will curate the curators?
The National Center for Biomedical Ontology

- One of three National Centers for Biomedical Computing launched by NIH in 2005
- Collaboration of Stanford, Berkeley, Mayo, Buffalo, Victoria, UCSF, Oregon, and Cambridge
- Primary goal is to make ontologies accessible and usable
- Research will develop technologies for ontology dissemination, indexing, alignment, and peer review
New Pathways to Discovery

- Building Blocks, Biological Pathways, and Networks
- Molecular Libraries and Imaging
- Structural Biology
- Bioinformatics and Computational Biology
- Nanomedicine

Research Teams of the Future

- High-Risk Research
  - NIH Director’s Pioneer Award
- Interdisciplinary Research
- Public-Private Partnerships

Re-engineering the Clinical Research Enterprise

- Re-engineering the Clinical Research Enterprise Initiatives
  - Clinical Research Networks and NECTAR
  - Clinical Outcomes Assessment
  - Clinical Research Training
  - Clinical Research Policy Analysis and Coordination
  - Translational Research

What’s New

- Press Release: NIH Launches Major Program to Transform Clinical and Translational Science
- RFA: Planning Grants for Institutional Clinical and Translational Science Awards
- RFA: Institutional Clinical and Translational Science Award
- Program: Institutional Clinical and Translational Science Award Program Information
- Meeting: Interdisciplinary Research Centers Workshop
- Press Release: 2005 NIH Director’s Pioneer Award Recipients Announced
- Press Release: NIH Roadmap Continues to Move Forward on All Fronts
- Meeting Summary: BAA Roadmap Steering Committee, May 2005
- What’s New – Archives
NCBO will offer

- Technology for uploading, browsing, and using biomedical ontologies
- Methods to make the online “publication” of ontologies more like that of journal articles
- Tools to enable the biomedical community to put ontologies to work on a daily basis
Goals for BioPortal

• Web accessible repository of ontologies for the biomedical community
  – Archived locally
  – Anywhere in cyberspace

• Support for ontology
  – Peer review
  – Annotation (marginalia)
  – Versioning
  – Alignment
  – Search
http://bioportal.bioontology.org

### Ontologies

**List View**  **Category View**

<table>
<thead>
<tr>
<th>Name</th>
<th>Format</th>
<th>Current Version</th>
<th>Content Location</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amino Acid</td>
<td>OWL Full</td>
<td>1.1</td>
<td>NCBO Library</td>
<td></td>
</tr>
<tr>
<td>Animal natural history and life history</td>
<td>Protégé</td>
<td>Unknown Remote</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arabidopsis development</td>
<td>OBO</td>
<td>1.1</td>
<td>NCBO Library</td>
<td></td>
</tr>
<tr>
<td>Basic -Vertebrate</td>
<td>OWL Full</td>
<td>1.1</td>
<td>NCBO Library</td>
<td></td>
</tr>
<tr>
<td>Biological imaging methods</td>
<td>OBO</td>
<td>1.1</td>
<td>NCBO Library</td>
<td></td>
</tr>
<tr>
<td>BRENDAs tissue / enzyme source</td>
<td>OBO</td>
<td>1.96</td>
<td>NCBO Library</td>
<td></td>
</tr>
<tr>
<td>C. elegans development</td>
<td>OBO</td>
<td>1.1</td>
<td>NCBO Library</td>
<td></td>
</tr>
<tr>
<td>C. elegans gross anatomy</td>
<td>OBO</td>
<td>Unknown Remote</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein modification</td>
<td>OBO</td>
<td>1.1 / 4</td>
<td>NCBO Library</td>
<td></td>
</tr>
<tr>
<td>Protein Ontology</td>
<td>OWL Full</td>
<td>2.0</td>
<td>NCBO Library</td>
<td></td>
</tr>
<tr>
<td>Protein-protein interaction</td>
<td>OBO</td>
<td>1.68</td>
<td>NCBO Library</td>
<td></td>
</tr>
<tr>
<td>Proteomics data and process provenance</td>
<td>OWL Full</td>
<td>1.1</td>
<td>NCBO Library</td>
<td></td>
</tr>
<tr>
<td>RadLex</td>
<td>Protégé</td>
<td>1.1</td>
<td>NCBO Library</td>
<td></td>
</tr>
<tr>
<td>Sample processing and separation techniques</td>
<td>OBO</td>
<td>Unknown Remote</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sequence types and features</td>
<td>OBO</td>
<td>1.29</td>
<td>NCBO Library</td>
<td></td>
</tr>
<tr>
<td>Systems Biology</td>
<td>OBO</td>
<td>Unknown Remote</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zebradish anatomy and development</td>
<td>OBO</td>
<td>1.9</td>
<td>NCBO Library</td>
<td></td>
</tr>
</tbody>
</table>
Browsing/Visualizing Ontologies

Zebradish anatomy and development

Tree View
Tree view constructed based on is_a hierarchy
- Stages
- zebrafish anatomical entity
  - anatomical set
  - anatomical structure
    - acellular anatomical structure
    - anatomical cluster
  - cardinal organism part
  - cell
    - dopaminergic neuron
    - epidermal cell
    - granulocyte
  - embryonic structure
  - extraembryonic structure
  - organ
  - organ system
    - portion of tissue
  - whole organism
- unspecified

Class Details

<table>
<thead>
<tr>
<th>General</th>
<th>Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class Name: cell</td>
<td>Database References: ZFIN:ZDB-ANAT-060816-76</td>
</tr>
<tr>
<td>Id: CL:0000000</td>
<td></td>
</tr>
</tbody>
</table>

Graph View

Graph Type: Local Neighborhood

Local Neighborhood view
Zebrash fish anatomy and development

Tree View

Tree view constructed based on is_a hierarchy

- Stages
  - zebrafish anatomical entity
    - anatomical set
      - acellular anatomical structure
      - anatomical cluster
      - cardinal organism part
    - cell
      - dopaminergic neuron
      - epidermal cell
      - granulocyte
    - embryonic structure
    - extraembryonic structure
    - organ
    - organ system
    - portion of tissue
    - whole organism

unspecified

Class Details

General
- Class Name: cell
- Id: CL:0000000

Attributes
- Database References: ZFIN:ZDB-ANAT:060816-76

Graph View

Graph Type: Hierarchy To Root

Hierarchy-to-root view
## Search Results

**Ontology** Protein-protein interaction

### Class Name (11) Attributes (83)

<table>
<thead>
<tr>
<th>Class Name</th>
<th>Id</th>
<th>Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>3d repertoire</td>
<td>Ml:0731</td>
<td>Definition: The aim of 3D Repertoire is to determine the structures of all amenable complexes in a cell at medium or toponomic and dynamic analyses of protein complexes in a cell. Complex models, EM pictures, expression and purification database connected to the PDB repository. RELATED SYNONYM: &quot;3D Repertoire&quot;</td>
</tr>
<tr>
<td>agonist</td>
<td>Ml:0626</td>
<td>Definition: Description of an activator that acts on an external cell receptor or other upstream molecule to stimulate or more of the interactions.</td>
</tr>
<tr>
<td>nucleic acid conjugation</td>
<td>Ml:0715</td>
<td>Definition: Bacterial conjugation is the transfer of genetic material between bacteria through cell-to-cell contact. Bacterial equivalent of sexual reproduction or mating. It is not actually sexual, as it does not involve the fusing of a conjugative plasmid from a donor cell to a recipient EXACT SYNONYM: &quot;nuc conjugation&quot;</td>
</tr>
</tbody>
</table>
BioPortal’s impact in the community

• National Cancer Institute
  – Deploying BioPortal locally to evaluate its use as a method for visualizing and navigating enterprise terminologies and ontologies

• Biomedical Informatics Research Network (BIRN)
  – Adopting BioPortal for disseminating and visualizing BIRNLex terminology

• Radiological Society of North America
  – Using BioPortal for graphical visualization of RadLex
BioPortal will allow NCBO to experiment with new models for

- Dissemination of knowledge on the Web
- Integration and alignment of online content
- Knowledge visualization and cognitive support
- Peer review of online content
The NCI Thesaurus in Protégé-OWL
Ontologies are not like journal articles

• It is difficult to judge methodological soundness simply by inspection

• We may wish to use an ontology even though some portions
  – Are not well designed
  – Make distinctions that are different from those that we might want
Ontologies are not like journal articles

- The utility of ontologies
  - Depends on the task
  - May be highly subjective
- The expertise and biases of reviewers may vary widely with respect to different portions of an ontology
- Users should want the opinions of more than 2–3 hand-selected reviewers
- Peer review needs to scale to the entire user community
Community-Based Annotation as Peer Review

• Makes ontology evaluation a democratic process
• Assumes users’ application of ontologies will lead to insights not achievable by inspection alone
• Assumes end-users will be motivated to comment on and engage in dialog about ontologies in the repository
Comment: Class appropriateness

Bill Bug at 11/20/07 14:57

Having a disjunction in a class name is an odd thing to have

Reply

Explanation: Re: Class appropriateness

MaryAnn Martone at 11/20/07 20:47
An ontology of “marginal notes”
The Da Vinci Code
by Dan Brown

List Price: $24.95
Price: $14.97 & Eligible for FREE Super Saver Shipping on orders over $25. See details
You Save: $9.98 (40%)
Availability: Usually ships within 24 hours from Amazon.com

Want it delivered Monday, February 28? Order it in the next 20 hours and 14 minutes, and choose One-Day Shipping at checkout. See details

417 used & new from $6.95
Edition: Hardcover

See 1 customer image
Share your own customer images

Unbelievable Book, February 16, 2005
Reviewer: Mohamed Abdulkhalil (Kingdom of Bahrain) - See all my reviews

There is no question that everybody should read this book. It is very entertaining and full of very peculiar facts (assuming that they are true). The writer skillfully turns religious history (highly sensitive and mostly boring subject to read) into a page turning thriller. I highly recommend it.

I have a general advise though, make sure that you read it on a weekend, as you will not be able to put it down. I read it on a business trip with near disastrous consequences.

Don't Take It as Gospel, November 9, 2003
Reviewer: Leslie Strang Akers (Riverside, CA) - See all my reviews

In the beginning I was intrigued by the premise set down in THE DA VINCI CODE, but my initial interest turned first to annoyance and then by the time I got to the info on Disney was laughing so hard at the absurdity of the whole novel. First of all, this is a work of fiction, so let's deal with that part. Far from being the taut, fast-paced thriller that the potential reader is lead to believe it is, TDVC is turgid, jerky, and filled with cliches. The characters are characterless and stupid, merely cardboard for the author to push around like pawns on a chessboard. Langford, a Harvard professor, can't distinguish between backwards English and a Semitic language. Sophie, a French police cryptologist, doesn't have the brains to figure out that an armor truck from a Swiss bank might be lo-jacked. These are only two of the many idiotic things the main characters aren't intelligent enough to figure out. The characters ponder clues ad nauseum, which turns a 300-page book into 454 pages. I don't know if the author is writing down to his audience, or if he really thinks that gifted people are idiot savants. Whatever it is, it's exasperating.
The Da Vinci Code
by Dan Brown

List Price: $24.95
Price: $14.97 & Eligible for FREE Super Saver Shipping on orders over $25. See details
You Save: $9.98 (40%)
Availability: Usually ships within 24 hours from Amazon.com

Want it delivered Monday, February 28? Order it in the next 20 hours and 14 minutes, and choose One-Day Shipping at checkout. See details

417 used & new from $6.95
Edition: Hardcover

Review: Unbelievable Book, February 16, 2005
Reviewer: Mohamed Abdulmalik (Kingdom of Bahrain) - See all my reviews

There is no question that everybody should read this book. It is very entertaining and full of very peculiar facts (assuming that they are true). The writer skillfully turns religious history (highly sensitive and mostly boring subject to read) into a page turning thriller. I highly recommend it.

I have a general advise though, make sure that you read it on a weekend, as you will not be able to put it down. I read it on a business trip with near disastrous consequences.

Review: Don't Take It as Gospel, November 9, 2003
Reviewer: Leslie Strange Akers (Riverside, CA) - See all my reviews

In the beginning I was intrigued by the premise set down in THE DA VINCI CODE, but my initial interest turned first to annoyance and then by the time I got to the info on Disney was laughing so hard at the absurdity of the whole novel. First of all, this is a work of fiction, so let's deal with that part. Far from being the taut, fast-paced thriller that the potent reader is lead to believe it is, TDVC is turgid, jerky, and filled with cliches. The characters are characterless and stupid, merely cardboard for the author to push around like pawns on a chessboard. Langford, a Harvard professor, can't distinguish between backwards English and a Semitic language. Sophie, a French police cryptologist, doesn't have the brains to figure out that an armor truck from a Swiss bank might be lo-jacked. These are only two of the many idiotic things the main characters aren't intelligent enough to figure out. The characters ponder clues ad nauseam, which turns a 300-page book into 454 pages. I don't know if the author is writing down to his audience, or if he really thinks that gifted peop are idiot savants. Whatever it is, it's exasperating.
The MGED Ontology
by EMBL "The primary purpose of the MGED Ontology is to provide standard" (more)

List Price: $0
Price: $0 & Eligible for FREE Access
You Save: $0
Availability: Usually available 24/7 on cBIO.org
Edition: Pragmatic

2 of 3 people found the following review helpful:

★★★★★ A Great resource, Aug 11, 2004
Reviewer: Catherine Ball (Stanford, CA USA) - See all my reviews

MGED Ontology aims to facilitate the sharing of microarray data generated by functional genomics and proteomics experiments....
Was this review helpful to you?  Yes  No  (Report this)

1 of 1 people found the following review helpful:

★★★★★ Needs considerable improvement, November 9, 2003
Reviewer: Barry Smith (Buffalo, NY) - See all my reviews

MGED ontology is indeed an essential part of any solution to the problems of Microarray analysis - but only if it is understood in the right sort of way. Ontological engineering, should in every case go hand in hand with a sound ontological theory....
Open ratings for ontologies

• Any user can
  – rate an ontology
  – add a “marginal note”

• Ontology evaluation becomes a community-based initiative

• A web of trust can enable users to filter comments or ratings to avoid “noise”
Possible Review Criteria

• What is the level of user support?
• What documentation is available?
• What is the granularity of the ontology content in specific areas?
• How well does the ontology cover a particular domain?
• In what applications has the ontology been used successfully? Where has it failed?
Users can make proposals for changes
The Ideal World

" The same language
" No overlap in coverage
" No new versions
" A single extension tree
" Small reusable modules
The “Bad” News: The Real World

- “The same language
- “No overlap in coverage
- “No new versions
- “A single extension tree
- “Small reusable modules
PROMPT: Dealing with the Messy World

- Find similarities and differences between ontologies
- Compare versions of ontologies
- Extract meaningful portions of ontologies
- Integrate in an ontology-editing environment
Reason for selected suggestion
frames have identical names
Users can view mappings uploaded from PROMPT in BioPortal
Users can push changes to RSS feeds
51
BioPortal will support specialized views on the repository.
<table>
<thead>
<tr>
<th>Resources</th>
<th>Description</th>
<th>Elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>PubMed</td>
<td>PubMed is a service of the U.S. National Library of Medicine that includes over 17 million citations from MEDLINE and other life science journals for biomedical articles back to the 1950s. PubMed includes links to full text articles and other related resources.</td>
<td>29</td>
</tr>
<tr>
<td>ArrayExpress</td>
<td>ArrayExpress is a public repository for microarray data, which is aimed at storing MIAME-compliant data in accordance with MGED recommendations. The ArrayExpress Data Warehouse stores gene-indexed expression profiles from a curated subset of experiments in the repository.</td>
<td>8</td>
</tr>
<tr>
<td>ClinicalTrials.gov</td>
<td>ClinicalTrials.gov provides regularly updated information about federally and privately supported clinical research in human volunteers. ClinicalTrials.gov gives you information about a trial's purpose, who may participate, locations, and phone numbers for more details. The information provided on ClinicalTrials.gov should be used in conjunction with advice from health care professionals. Before searching, you may want to learn more about clinical trials.</td>
<td>206</td>
</tr>
<tr>
<td>Gene Expression Omnibus</td>
<td>A gene expression/molecular abundance repository supporting MIAME compliant data submissions, and a curated, online resource for gene expression data browsing, query and retrieval.</td>
<td>7</td>
</tr>
<tr>
<td>ARRS GoldMiner</td>
<td>ARRS GoldMiner provides instant access to images published in selected peer-reviewed radiology journals. This new, web-based system allows viewers to search for images by findings, anatomy, imaging technique, and patient age and sex.</td>
<td>2</td>
</tr>
</tbody>
</table>
### BioPortal

**PubMed**

PubMed is a service of the U.S. National Library of Medicine that includes over 17 million citations from MEDLINE and other life science journals for biomedical articles back to the 1950s. PubMed includes links to full text articles and other related resources.

**ArrayExpress**

ArrayExpress is a public repository for microarray data, which is aimed at storing MIAME-compliant data in accordance with MGED recommendations. The ArrayExpress Data Warehouse stores gene-indexed expression profiles from a curated subset of experiments in the repository.

<table>
<thead>
<tr>
<th>Element ID</th>
<th>Annotation Context</th>
<th>Element Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>E-GEOID-4731</td>
<td>description</td>
<td>View Element</td>
</tr>
<tr>
<td>E-GEOID-5230</td>
<td>title</td>
<td>View Element</td>
</tr>
<tr>
<td>E-GEOID-5230</td>
<td>description</td>
<td>View Element</td>
</tr>
<tr>
<td>E-MEXP-199</td>
<td>title</td>
<td>View Element</td>
</tr>
<tr>
<td>E-MEXP-199</td>
<td>description</td>
<td>View Element</td>
</tr>
<tr>
<td>E-MEXP-84</td>
<td>description</td>
<td>View Element</td>
</tr>
<tr>
<td>E-EMDB-2975</td>
<td>description</td>
<td>View Element</td>
</tr>
<tr>
<td>E-TABM-36</td>
<td>description</td>
<td>View Element</td>
</tr>
</tbody>
</table>

**ClinicalTrials.gov**

ClinicalTrials.gov provides regularly updated information about federally and privately supported clinical research in human volunteers. ClinicalTrials.gov gives you information about a trial's purpose, who may participate, locations, and phone numbers for more details. The information provided on ClinicalTrials.gov should be used in conjunction with advice from health care professionals. Before searching, you may want to learn more about clinical trials.

**Gene Expression Omnibus**

A gene expression/molecular abundance repository supporting MIAME compliant data submissions, and a curated, online resource for gene expression data browsing, querying and retrieval.
(Generated description): Experiment with 12 hybridizations, using 12 samples of species [Mus musculus], using 12 arrays of array design [Affymetrix GeneChip® Mouse Genome 430A 2.0 [Mouse430A_2]], producing 12 raw data files and 12 transformed and/or normalized data files.

(Submitter's description 1): We studied the molecular mechanisms of hepatocellular carcinoma (HCC) initiation and promotion using the Mdr2-knockout (Mdr2-KO) mice at pre-cancerous stages of liver disease. These mice lack the liver-specific P-glycoprotein responsible for phosphatidylcholine transport across the canalicular membrane. Portal inflammation ensues at an early age followed by the development of HCC between the ages of 12 and 15 months. Liver tissue samples of Mdr2-KO and control Mdr2-heterozygotes mice aged 3 and 12 months, were subjected to histological, biochemical and gene expression profiling analysis using Affymetrix Mouse Genome Array. The RNA samples from Mdr2-KO and control heterozygous mice aged 3 and 12M (3 males in each experimental group) were subjected to genome scale gene expression profiling with Affymetrix Mouse Array. The gene expression values were extracted with the help of MAS 5.0 software, and analyzed by cluster analysis, and by fold change filtering.
<table>
<thead>
<tr>
<th>Resource</th>
<th>Description</th>
<th>Elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>PubMed</td>
<td>PubMed is a service of the U.S. National Library of Medicine that includes over 17 million citations from MEDLINE and other life science journals for biomedical articles back to the 1950s. PubMed includes links to full text articles and other related resources.</td>
<td>29</td>
</tr>
<tr>
<td>ArrayExpress</td>
<td>ArrayExpress is a public repository for microarray data, which is aimed at storing MIAME-compliant data in accordance with MGED recommendations. The ArrayExpress Data Warehouse stores gene-indexed expression profiles from a curated subset of experiments in the repository.</td>
<td>8</td>
</tr>
<tr>
<td>ClinicalTrials.gov</td>
<td>ClinicalTrials.gov provides regularly updated information about federally and privately supported clinical research in human volunteers. ClinicalTrials.gov gives you information about a trial's purpose, who may participate, locations, and phone numbers for more details. The information provided on ClinicalTrials.gov should be used in conjunction with advice from health care professionals. Before searching, you may want to learn more about clinical trials.</td>
<td>206</td>
</tr>
<tr>
<td>Gene Expression Omnibus</td>
<td>A gene expression/molecular abundance repository supporting MIAME compliant data submissions, and a curated, online resource for gene expression data browsing, query and retrieval.</td>
<td>7</td>
</tr>
<tr>
<td>ARRS Gold Miner</td>
<td>ARRS GoldMiner provides instant access to images published in selected peer-reviewed radiology journals. This new, web-based system allows viewers to search for images by findings, anatomy, imaging technique, and patient age and sex.</td>
<td>2</td>
</tr>
</tbody>
</table>
172 closure annotations and 171 useful.

Examples:
Cancer, concept (DOID:162) in ontology Human disease
Skin Neoplasms, concept (DOID:3165) in ontology Human disease

Melanoma progression
Analysis of tissue specimens representing benign nevus, atypical nevus, melanoma in situ, vertical growth phase (VGP) melanoma, and metastatic growth phase (MGP) melanoma. Results identify expression signatures that distinguish benign and atypical nevi and melanomas in situ from VGPS and MGPs.

23 direct annotations
(4 title, 19 description)
Example:
Melanoma, concept (DOID:1909) in ontology Human disease.
Biportal search for “melanoma”

362 matches in all Biportal ontologies.
Example:
Melanoma, concept (DOID:1909) in ontology Human disease.

Onrez index results display:
- 227 PubMed elements,
- 3 ArrayExpress elements,
- 969 ClinicalTrials.gov elements,
- 10 ARRS GoldMiner elements,
- 3 Gene Expression Omnibus elements.

Example:
Melanoma progression element (GDS1989)
PubMed is a service of the U.S. National Library of Medicine that includes over 17 million citations from MEDLINE and other life science journals for biomedical articles back to the 1950s. PubMed includes links to full text articles and other related resources.

ArrayExpress is a public repository for microarray data, which is aimed at storing MIAME-compliant data in accordance with MGED recommendations. The ArrayExpress Data Warehouse stores gene-indexed expression profiles from a curated subset of experiments in the repository.

ClinicalTrials.gov provides regularly updated information about federally and privately supported clinical research in human volunteers. ClinicalTrials.gov gives you information about a trial's purpose, who may participate, locations, and phone numbers for more details. The information provided on ClinicalTrials.gov should be used in conjunction with advice from health care professionals. Before searching, you may want to learn more about clinical trials.

Gene Expression Omnibus

ARRS GoldMiner provides instant access to images published in selected peer-reviewed radiology journals. This new, web-based system allows viewers to search for images by findings, anatomy, imaging technique, and patient age and sex.
If we build it, will they come?
A problem in both technology and sociology

- How can we identify communities of likely early adopters?
- How will we know when we will have sufficient functionality to entice early adopters to adopt?
- How can we measure the affects of our technology on the way that science gets done?
- How can we engage in participatory design of technology that potential users cannot even imagine?
BioPortal User Group

- CTSAs
- Immunology
- Imaging
- RadLex
- W3C HCLS SIG
- BioPAX

- CVRGrid
- caBIG
- HL7
- MODs
- GO Consortium
- BIRN
BioPortal can build an online community of users who

• Develop, upload, and apply ontologies
• Map ontologies to one another
• Comment on ontologies via “marginal notes” to give feedback
  – To the ontology developers
  – To one another
• Make proposals for specific changes to ontologies
• Stay informed about ontology changes and proposed changes via active feeds
Goals for the NCBO

• Providing technology for ontology archiving, access, browsing, visualization, peer review, mapping, versioning
• Making most biomedical ontologies accessible via a common portal
• Educating the community about principles of ontology development and use
• Serving as a generalizable model for the formalization of knowledge in e-science