

The National Center for Biomedical Ontology

Daniel Rubin, MD, MS
Stanford Medical Informatics

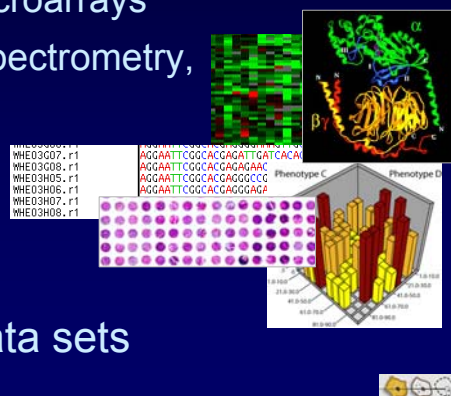
Stanford – Berkeley
Mayo – Victoria – Buffalo
UCSF – Oregon – Cambridge

<http://www.bioontology.org>



The biomedical data explosion

- Explosion in online biomedical data
 - Genomics (genetic sequences, SNPs)
 - Gene expression microarrays
 - Proteomics (mass spectrometry, protein arrays)
 - Tissue arrays, ICH
- Need for people & machines to make sense of massive data sets



Copyright © Daniel L. Rubin 2006

Biomedical researchers use ontologies

- *Controlled vocabulary* for science
- *Representation of biomedical knowledge*, shared by humans and computers
- Terms for *annotating experimental data*
- *Knowledge source* for biomedical applications
 - Decision support
 - Natural language-processing
 - Data integration

Copyright © Daniel L. Rubin 2006




Ontology development is fragmented

- Many different groups/consortia create ontologies—efforts are uncoordinated
- Many different ontologies, overlapping content and variable quality
- Ontologies are not interoperable
- Data integration efforts are laborious
- Barriers to accessing and effectively using numerous existing ontologies

Copyright © Daniel L. Rubin 2006





THE NATIONAL CENTER FOR
BIOMEDICAL ONTOLOGY

- Consortium of informaticians, biologists, clinicians, and ontologists, funded by the NIH Roadmap
- Ontology research and services
 - Ontology access, alignment, and management
 - Ontology-based annotation of large data sets
 - Enhance quality of ontology development
 - Collaboration with diverse biomedical projects

Copyright © Daniel L. Rubin 2006

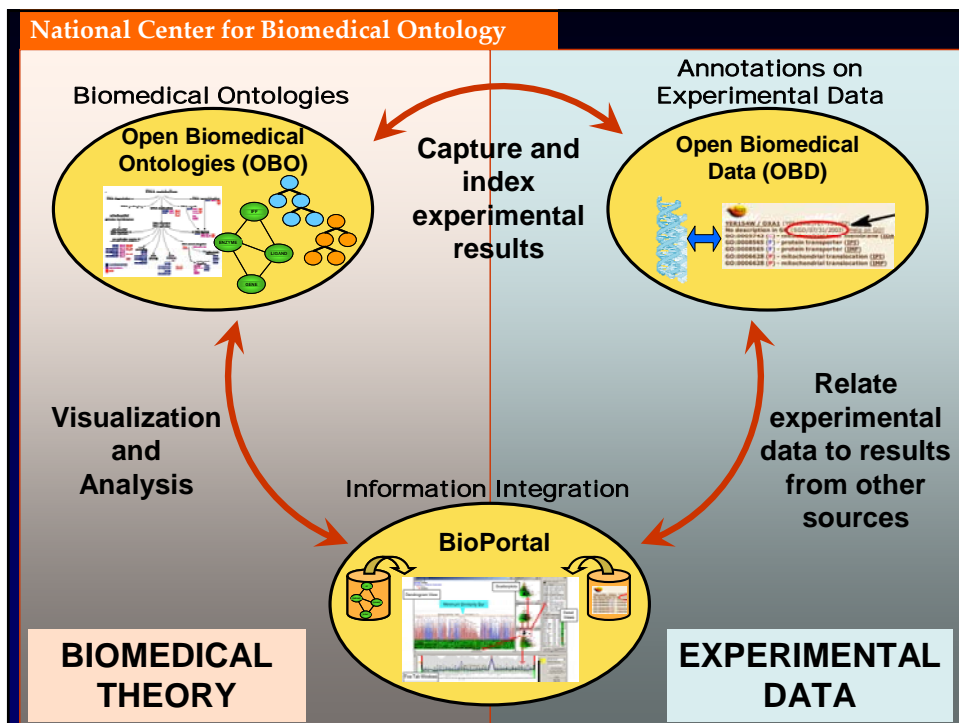


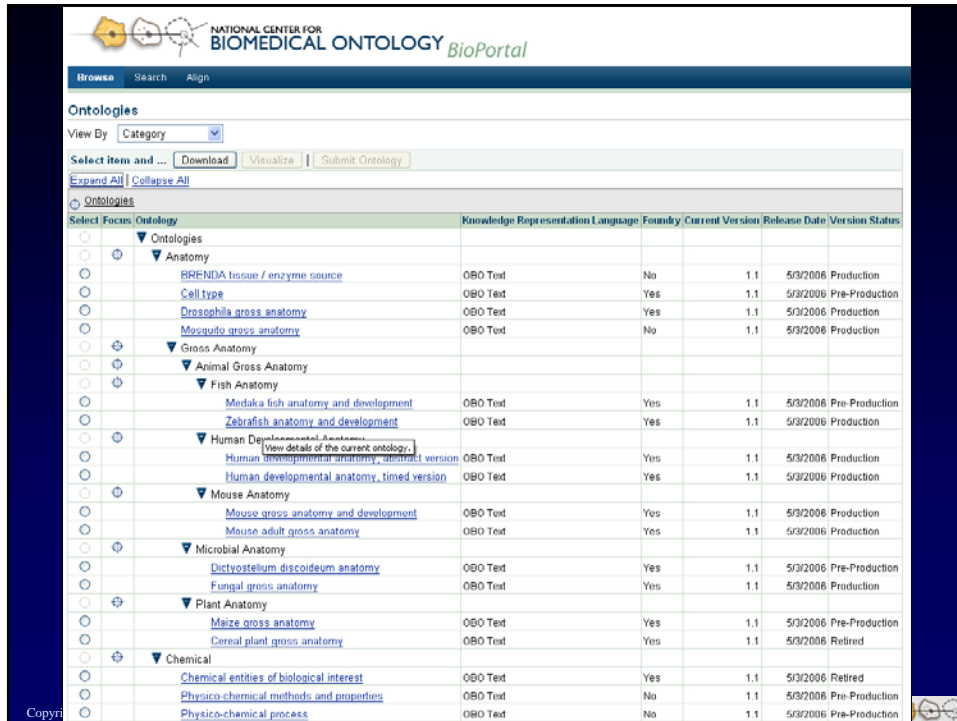


THE NATIONAL CENTER FOR
BIOMEDICAL ONTOLOGY

- **Stanford:** Tools for ontology search, alignment, versioning, and peer review
- **Lawrence Berkeley Labs:** Tools to use ontologies for data annotation
- **Mayo Clinic:** Tools for access to large controlled terminologies
- **Univ. of Victoria:** Tools for ontology visualization
- **Univ. at Buffalo:** Dissemination of best practices for ontology engineering
- **Univ. of Cambridge, Univ. of Oregon, UCSF:** Driving biomedical projects

Copyright © Daniel L. Rubin 2006





The screenshot shows the BioPortal interface with a list of ontologies. The table below represents the data visible in the screenshot.

Select	Focus	Ontology	Knowledge Representation Language	Foundry	Current Version	Release Date	Version Status
<input type="radio"/>	▼	Ontologies					
<input type="radio"/>	⊕	Anatomy					
<input type="radio"/>		BRINDA tissue / enzyme source	OBO Text	No	1.1	5/3/2006	Production
<input type="radio"/>		Cell type	OBO Text	Yes	1.1	5/3/2006	Pre-Production
<input type="radio"/>		Drosophila gross anatomy	OBO Text	Yes	1.1	5/3/2006	Production
<input type="radio"/>		Mosquito gross anatomy	OBO Text	No	1.1	5/3/2006	Production
<input type="radio"/>	⊕	Gross Anatomy					
<input type="radio"/>	⊕	Animal Gross Anatomy					
<input type="radio"/>	⊕	Fish Anatomy					
<input type="radio"/>		Medaka fish anatomy and development	OBO Text	Yes	1.1	5/3/2006	Pre-Production
<input type="radio"/>		Zebrafish anatomy and development	OBO Text	Yes	1.1	5/3/2006	Production
<input type="radio"/>	⊕	Human Developmental Anatomy					
<input type="radio"/>		View details of the current ontology					
<input type="radio"/>		Human developmental anatomy, abstract version	OBO Text	Yes	1.1	5/3/2006	Production
<input type="radio"/>		Human developmental anatomy, timed version	OBO Text	Yes	1.1	5/3/2006	Production
<input type="radio"/>	⊕	Mouse Anatomy					
<input type="radio"/>		Mouse gross anatomy and development	OBO Text	Yes	1.1	5/3/2006	Production
<input type="radio"/>		Mouse adult gross anatomy	OBO Text	Yes	1.1	5/3/2006	Production
<input type="radio"/>	⊕	Microbial Anatomy					
<input type="radio"/>		Dictyostelium discoidium anatomy	OBO Text	Yes	1.1	5/3/2006	Pre-Production
<input type="radio"/>		Fungal gross anatomy	OBO Text	Yes	1.1	5/3/2006	Production
<input type="radio"/>	⊕	Plant Anatomy					
<input type="radio"/>		Maize gross anatomy	OBO Text	Yes	1.1	5/3/2006	Pre-Production
<input type="radio"/>		Cereal plant gross anatomy	OBO Text	Yes	1.1	5/3/2006	Retired
<input type="radio"/>	⊕	Chemical					
<input type="radio"/>		Chemical entities of biological interest	OBO Text	Yes	1.1	5/3/2006	Retired
<input type="radio"/>		Physico-chemical methods and properties	OBO Text	No	1.1	5/3/2006	Pre-Production
<input type="radio"/>		Physico-chemical process	OBO Text	No	1.1	5/3/2006	Pre-Production

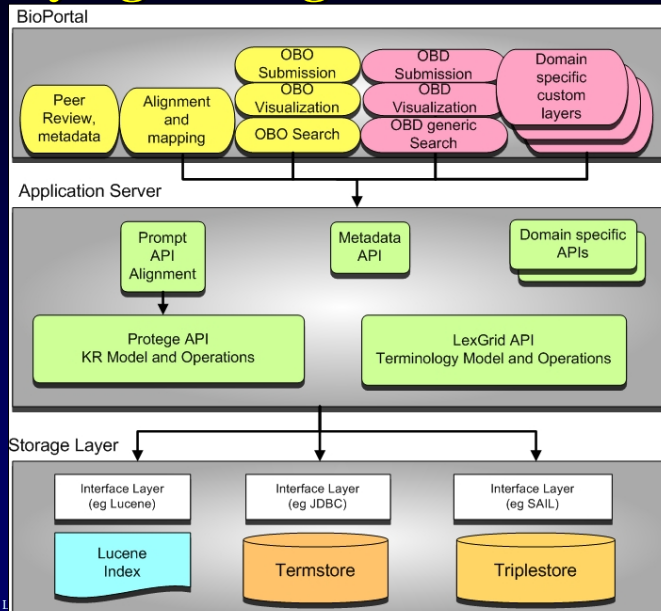
Core technologies for BioPortal

- **Protégé:**
 - Ontology visualization
 - Ontology alignment and version diff
- **LexGrid:**
 - Defines common information model for terminology content
 - Access to controlled terminologies in many different formats
 - Ontology content indexing and search
- **Tiered Web app/services architecture**

Copyright © Daniel L. Rubin 2006



BioPortal Architecture: Unifying ontologies and annotations



Copyright © Daniel L.

Biomedical ontology challenges

- Find ontologies or terms of interest
- Visualize and navigate ontologies and annotated data
- Support distributed, collaborative ontology development
- Enable community-based evaluation of ontology quality
- Use ontologies to annotate data, and use annotations to make discoveries

Copyright © Daniel L. Rubin 2006

Biomedical ontology challenges

- Find ontologies or terms of interest
- Visualize and navigate ontologies and annotated data
- Support distributed, collaborative ontology development
- Enable community-based evaluation of ontology quality
- Use ontologies to annotate data, and use annotations to make discoveries

Copyright © Daniel L. Rubin 2006



LexGrid: Ontology indexing and search

- **Terminological services**
 - Search for ontology terms
 - Use homophone, exact and partial search
 - Map free-text to ontologies
- **Ontology indexes and services**
 - Lucene index on ontology terms, definitions, and synonyms
 - Global identifiers for terms
 - Ontology version information

Copyright © Daniel L. Rubin 2006



Biomedical ontology challenges

- Find ontologies or terms of interest
- Visualize and navigate ontologies and annotated data
- Support distributed, collaborative ontology development
- Enable community-based evaluation of ontology quality
- Use ontologies to annotate data, and use annotations to make discoveries

Copyright © Daniel L. Rubin 2006



Degree of Interest modeling

- Creates a user profile to identify **relevant information**
- Developed by **monitoring** the user activities (e.g. navigation actions, editing and annotations)
- Permits model-based **highlighting or filtering** of “interesting” entities in the ontology
- Based on Degree of Interest Trees (Stuart Card) and Mylar (Mik Kersten)

Copyright © Daniel L. Rubin 2006

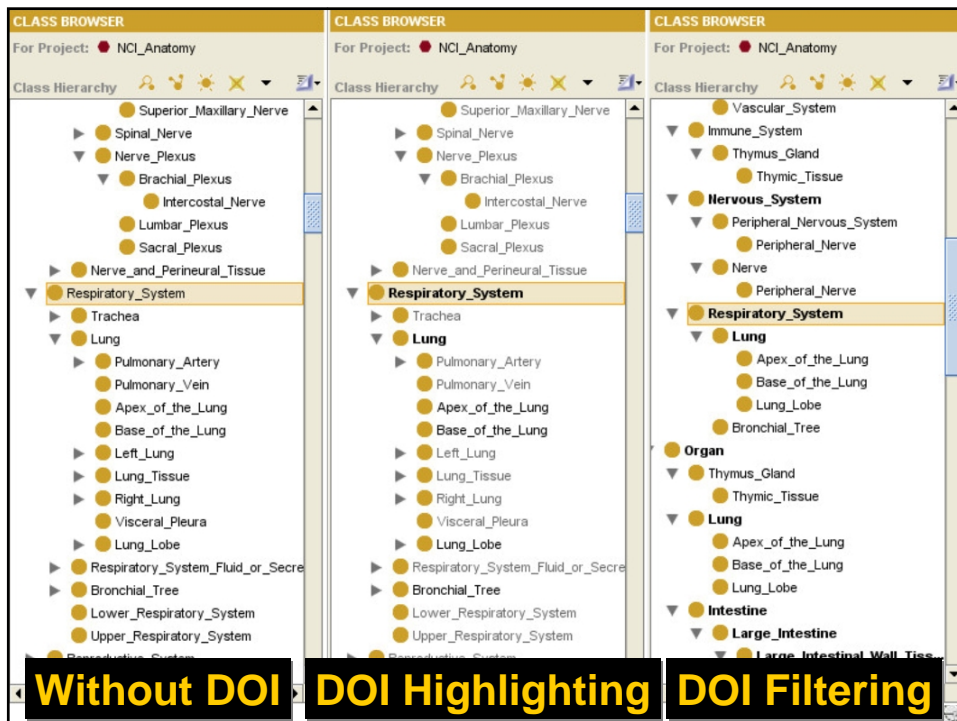


DIaMOND

- Degree of Interest Modeling for Ontology Navigation and Development
- Integrates Mylar degree of interest model (DOI) for Eclipse with Protégé
- Uses the DOI to provide adaptive visualizations of the ontology

Work by Tricia d'Entremont


Copyright © Daniel L. Rubin 2006



Pictorial-guided ontology navigation

- Users often interested in ontology subset pertinent to
 - **Biological scale** (organ/tissue/cell/molecule)
 - **Image regions** (locations, components)
- Strategy: browse ontology views driven by the biological scale of the image
- Accomplished by annotating multi-scale images using ontologies to describe their contents
- Also enables image retrieval driven by ontology

Copyright © Daniel L. Rubin 2006



- Users often interested in ontology subset pertinent to
 - **Biological scale** (organ/tissue/cell/molecule)
 - **Image regions** (locations, components)
- Strategy: browse ontology views driven by the biological scale of the image
- Accomplished by annotating multi-scale images using ontologies to describe their contents
- Also enables image retrieval driven by ontology

Copyright © Daniel L. Rubin 2006

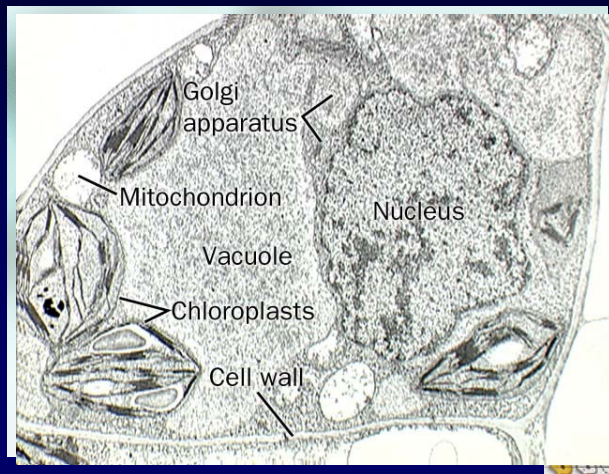
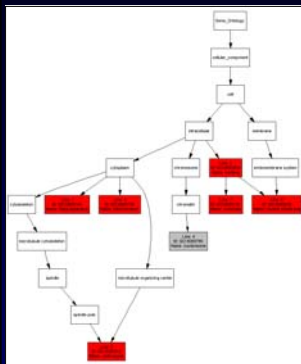


Pictorial-guided ontology navigation

The diagram illustrates the process of pictorial-guided ontology navigation. It starts with a pictorial image of a plant cell (right), which is used to identify and select specific components (red boxes) in a hierarchical ontology (left). A green arrow points from the image to the ontology, and another green arrow points from the ontology back to the image, indicating a bidirectional navigation process.

Work by Nigam Shah

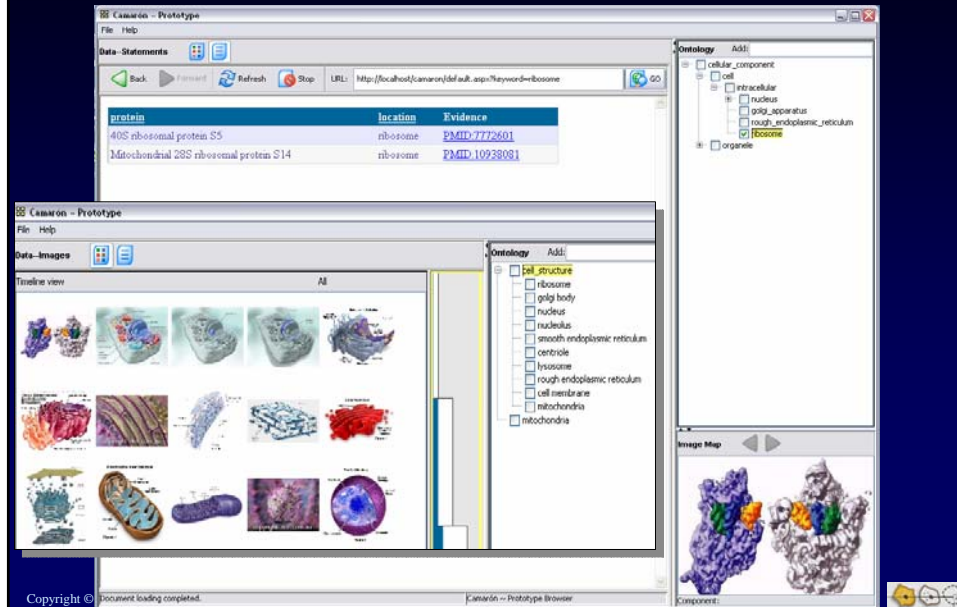
Copyright © Daniel L. Rubin 2006



Work by Nigam Shah

Copyright © Daniel L. Rubin 2006

Navigating by different image scale

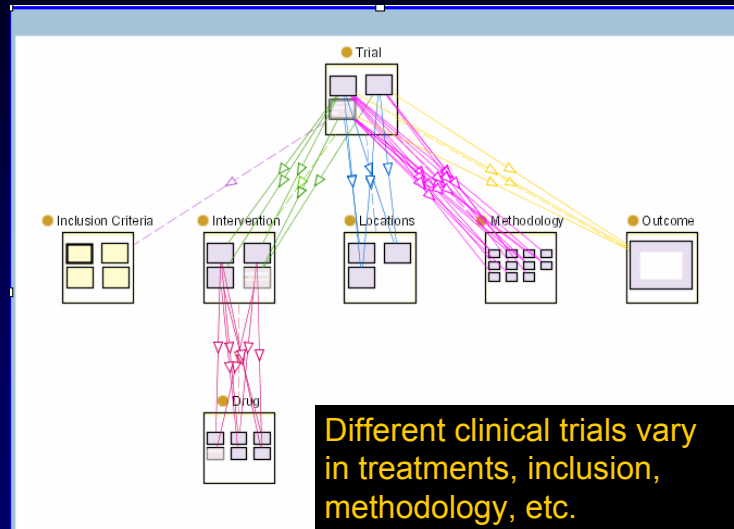


Biomedical ontology challenges

- Find ontologies or terms of interest
- Visualize and navigate ontologies and annotated data
- Support distributed, collaborative ontology development
- Enable community-based evaluation of ontology quality
- Use ontologies to annotate data, and use annotations to make discoveries

Copyright © Daniel L. Rubin 2006

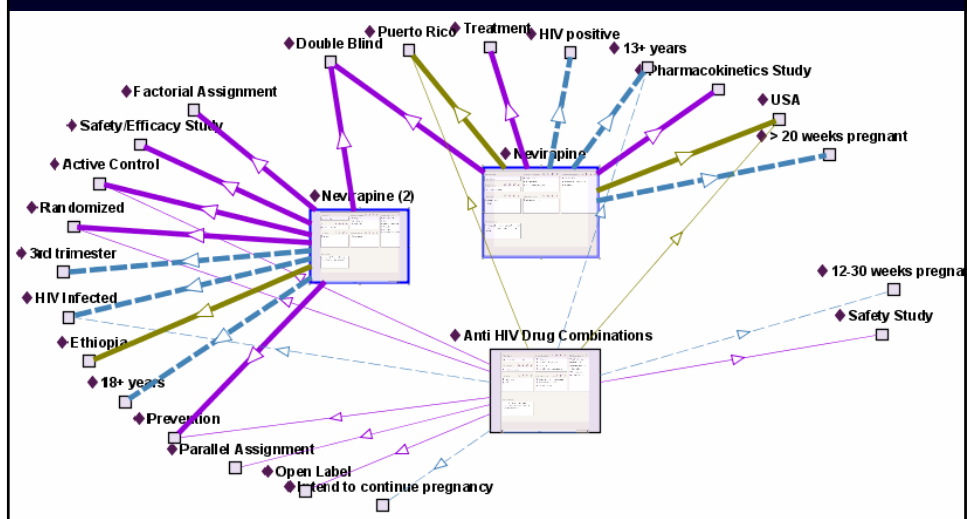
Visualizing ontology-annotated clinical trial data



Work by Maleh Hernandez

Copyright © Daniel L. Rubin 2006

Visualizing differences in Nevirapine trials



Copyright © Daniel L. Rubin 2006

Biomedical ontology challenges

- Find ontologies or terms of interest
- Visualize and navigate ontologies and annotated data
- Support distributed, collaborative ontology development
- Enable community-based evaluation of ontology quality
- Use ontologies to annotate data, and use annotations to make discoveries

Copyright © Daniel L. Rubin 2006



Challenges for community ontology development

- Need to communicate ways to improve & evolve ontologies
 - Missing attributes (e.g., definitions)
 - Class too broad, should be split or deleted
 - Class should be moved, renamed
- Current approach: email lists, F2F meetings
 - Ontology feedback is disconnected from the ontology
 - Cannot determine what parts of ontologies are stable, contentious, or evolving

Copyright © Daniel L. Rubin 2006



Example email communications from fugo-discuss

- I'd like to propose a few relationships between some higher level classes:
study executes study_design
study_design has factor owl:Thing
- What is the definition of biomaterial?
- Should biomaterial be a subclass of FuGO_54 study_object?

Copyright © Daniel L. Rubin 2006



Ontology “marginal notes”

- Structured annotations on ontologies and their contents
- Capture community feedback on ontologies
- Localized to parts of ontology to which they apply
- Make explicit the types of ontology evolutionary changes

Copyright © Daniel L. Rubin 2006



Ontology marginal notes

The screenshot displays the 'Marginalia_Instance_38' window, which is an instance of the Marginalia application. The window contains several input fields and a text area for creating a marginal note. The 'User Name' field is filled with 'C. C.', and the 'User Type' is set to 'Reviewer'. The 'User Action' is 'Create', and the 'Ontology Component' is 'Ontology Relation'. The 'Granularity Of Comment' is set to 'Relation'. The 'User Comments' text area contains the text: 'Please add relation "has_organ" to this class in order to indicate the organ of origin'. The 'Marginal Note To' dropdown menu is set to 'Curator'. In the background, other windows of the Marginalia application are visible, showing different instances and components. At the bottom left, there is a credit line: 'Work by Ravi Kaustubh' and 'Copyright © Daniel L. Rubin 2006'.

Work by Ravi Kaustubh
Copyright © Daniel L. Rubin 2006

Biomedical ontology challenges

- Find ontologies or terms of interest
- Visualize and navigate ontologies and annotated data
- Support distributed, collaborative ontology development
- Enable community-based evaluation of ontology quality
- Use ontologies to annotate data, and use annotations to make discoveries

Copyright © Daniel L. Rubin 2006

Ontology metadata and peer review

- Variable ontology quality; no venue for community rating of ontologies
- Building a peer review platform for ontologies based on “Web of Trust”
- Providing tools to enable community to evaluate and improve ontology quality

Copyright © Daniel L. Rubin 2006



Peer review of ontologies

Metadata

- Source_Metadata
- Content_Metadata
 - Quantifiable_Content_Metadata → No_of_Concepts
 - NonQuantifiable_Content_Metadata → No_of_Relations
- Usage_Metadata
- Utility_Metadata
- Support_Metadata
- Licensing_Costs
- Identification_Metadata → No_of_Axioms
- ThirdParty_Metadata

Metadata Hierarchy

- OWLSpecific_Metadata
- RDFSspecific_Metadata
- DAGSpecific_Metadata
- Core_Metadata
- Miscellaneous_Metadata

Inferred Hierarchy

User Reviews

8 of 11 people found the following review helpful:

★★★★★ **MGED User centered review**, 17 Oct 2005 18:50:23 GMT

Reviewer: Nigam Shah [See all my reviews](#)

The boundaries between MIAME concepts, the MAGE-OM and the MGED ontology is not well defined. However, before commenting on the MGED ontology it is imperative to understand -- at least at some level

Metadata Ontology

Work by Kaustubh Supekar

Web-based tools to enter ontology metadata, post reviews, and rate reviews

Copyright © Daniel L. Rubin 2006

Biomedical ontology challenges

- Find ontologies or terms of interest
- Visualize and navigate ontologies and annotated data
- Support distributed, collaborative ontology development
- Enable community-based evaluation of ontology quality
- Use ontologies to annotate data, and use annotations to make discoveries

Copyright © Daniel L. Rubin 2006



Preliminary results

- Two unrelated biomedical knowledge sources (ZFIN, OMIM)
- Each annotated using ontologies to describe phenotypes
- Search for similar phenotype annotations → discover disease genes
- Example: holoprosencephaly genes

Copyright © Daniel L. Rubin 2006





Encoding disease phenotypes

Phenotype

(clinical sign) = entity + quality

P₁ = eye + hypoteloric

P₂ = midface + hypoplastic

P₃ = kidney + hypertrophied

Syndrome = P₁ + P₂ + P₃
(disease)

= holoprosencephaly

Copyright © Daniel L. Rubin 2006

Finding human disease genes

Gene homology?

Human holo-
prosencephaly

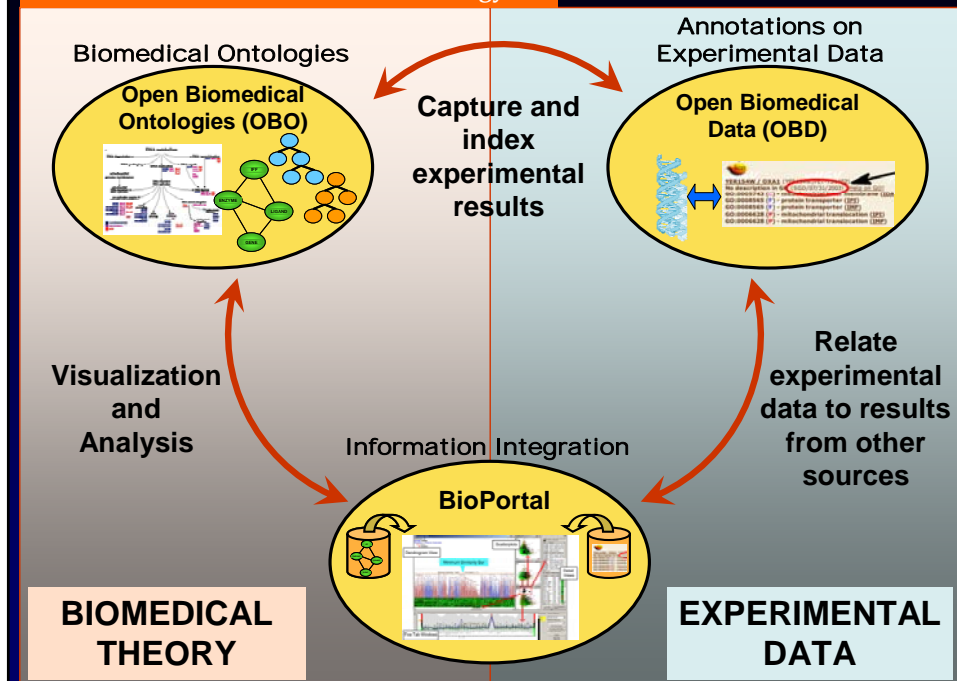
Zebrafish
oep

Human gene: <i>SHH</i>	Zebrafish gene: <i>shh</i>	Zebrafish gene: <i>oep</i>
Ref: OMIM:142945	Ref: ZDB-GENE-980526-166	Ref: ZDB-GENE-990415-198
Entity: prosencephalon development	Entity: prosencephalon development	Entity: prosencephalon development
Attribute: arrested	Attribute: reduced	Attribute: arrested
Entity: brain	Entity: brain	Entity: brain
Attribute: small	Attribute: small	Attribute: small
Entity: brain ventricle		Entity: brain ventricle
Attribute: single		Attribute: single
Entity: midface		Entity: midface
Attribute: hypoplastic		Attribute: hypoplastic
Entity: eye morphology		Entity: eye morphology
Attribute: abnormal		Attribute: abnormal
Entity: eye		Entity: eye
Attribute: single		Attribute: single
Entity: eye	Entity: eye	Entity: eye
Attribute: hypoteloric	Attribute: hypoteloric	Attribute: hypoteloric
Entity: nose		Entity: nose
Attribute: abnormal morphology		Attribute: abnormal morphology
Entity: nostril		Entity: nostril
Attribute: single		Attribute: single
Entity: kidney		Entity: kidney
Attribute: hypertrophied		Attribute: hypertrophied

2. Orthologs in human may cause disease? with similar phenotypes

Similar phenotypes

National Center for Biomedical Ontology



Acknowledgements

- National Center for Biomedical Ontology



- **Executive Team:** Mark Musen, Suzanna Lewis, Daniel Rubin, Sima Misra
- **cBio staff:** Natasha Noy, Tim Redmond, Lynn Murphy, Archana Verbakam, Chris Mungall, John Day-Richter, Mark Gibson, ShengQiang Shu, Nicole Washington, Harold Solbrig, Deepak Sharma, James Buntrock, Tom Johnson, Chris Callendar
- **Collaborators:** Michael Ashburner, Monte Westerfield, Ida Sim, Chris Chute, Barry Smith, Peggy Storey, Richard Olshen, Werner Ceusters, Deborah McGuinness
- **Students & post-docs:** Kaustubh Supekar, Nigam Shah, Fabian Neuhaus, Tricia d'Entremont, Maria-Elena Hernandez, Sean Falconer, Ravi Tiruvury




- Funded through NIH Roadmap for Medical Research grant U54 HG004028




- **Program officer:** Peter Good (NIGMS)
- **Lead Science Officer:** Carol Bean (NCRR)

Copyright © Daniel L. Rubin 2006





THE NATIONAL CENTER FOR
BIOMEDICAL ONTOLOGY



Thank you.

Contact information

Center: feedback@bioontology.org

Copyright © Daniel L. Rubin 2006

