The craze over bio-ontologies:

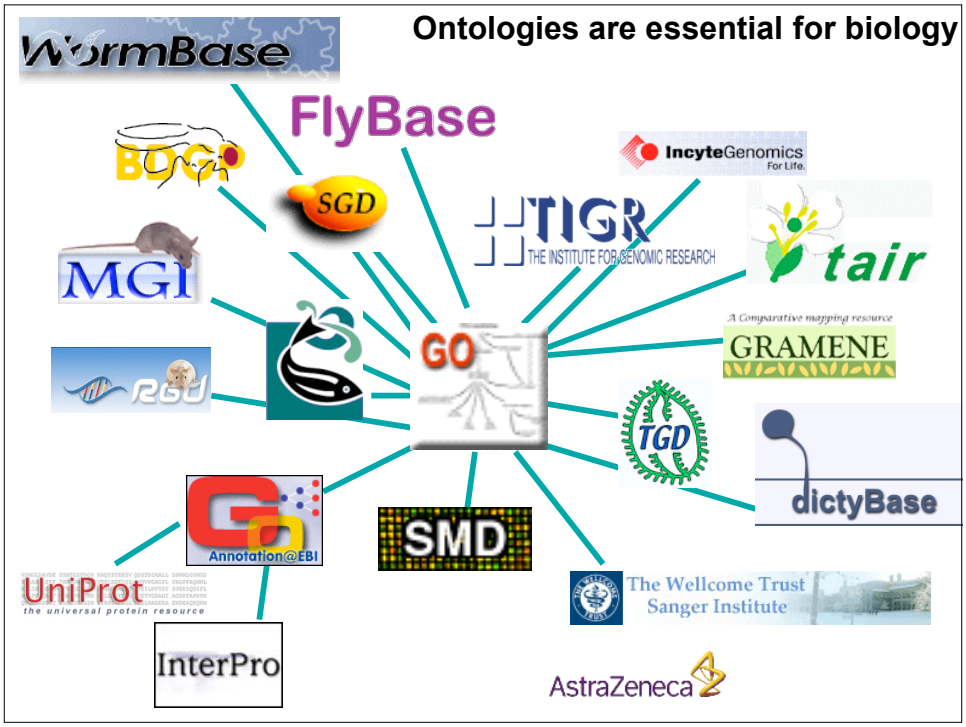
What happens when professional take ontology building into their own hands

Mark A. Musen

Stanford University

NATIONAL CENTER FOR
BIOMEDICAL ONTOLOGY

Ontologies are essential for biology



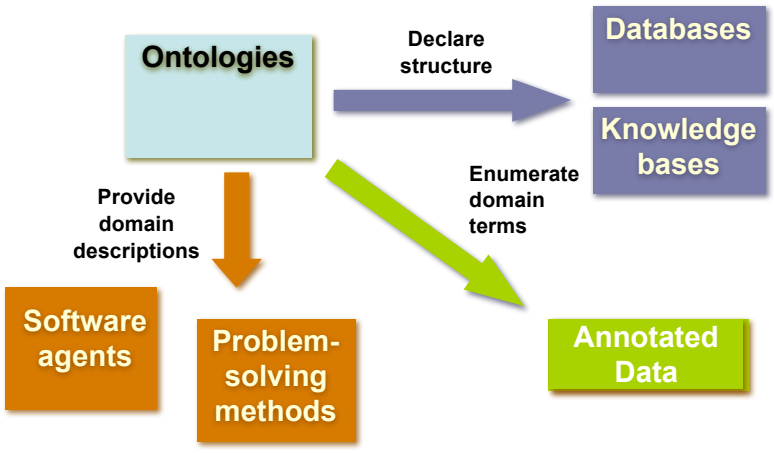
Biologist have adopted ontologies

- To provide canonical representation of scientific knowledge
- To annotate experimental data to enable interpretation, comparison, and discovery across databases
- To facilitate knowledge-based applications for
  - Decision support
  - Natural language-processing
  - Data integration

A Portion of the OBO Library

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

Ontologies are just the beginning



Foundational Model of Anatomy

- Long-term project at University of Washington to create a comprehensive ontology of human anatomy
- 72K concepts, 1.9M relationships
- One of the largest and best developed ontologies in biomedicine

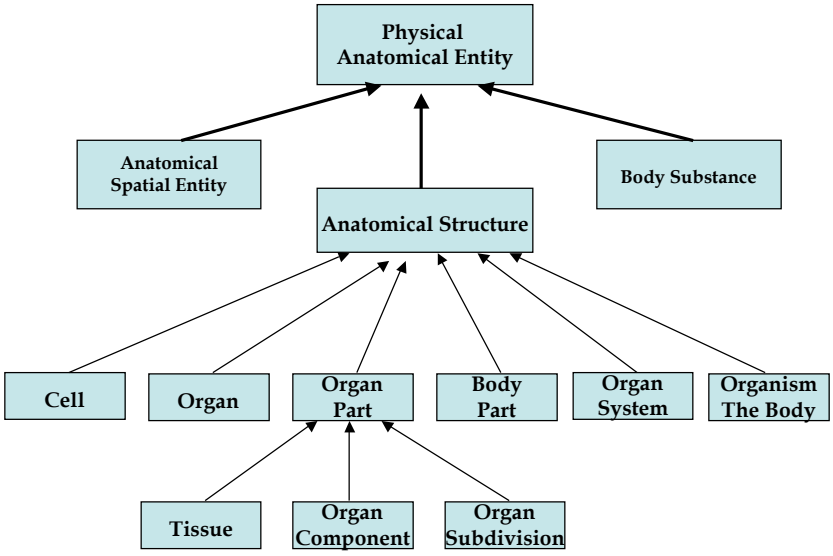
Structural Informatics Group
Structural Informatics Group
Structural Informatics Group

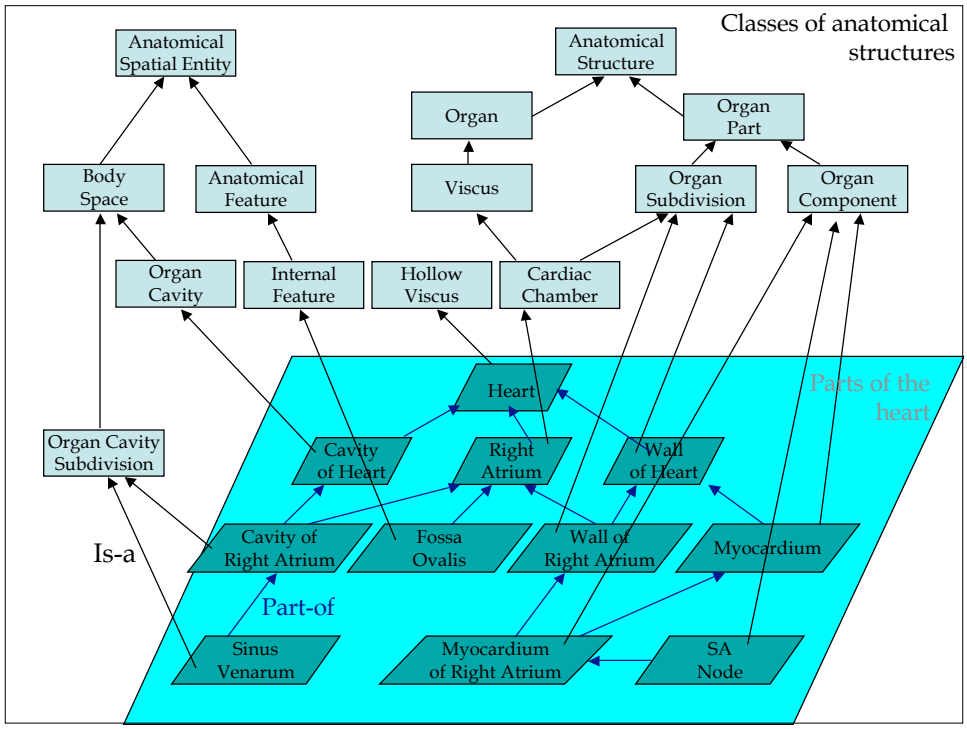
SIG

Overview Projects Personnel Publications Docs Local Info Products

Biological Structure Biomedical Informatics Computer Science and Engineering University of Washington

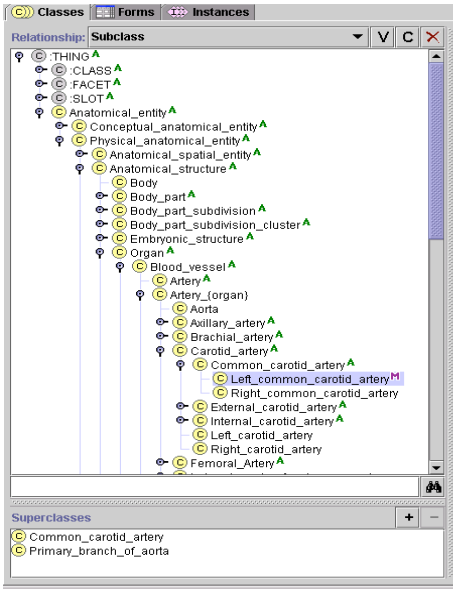
Top level of the Foundational Model of Anatomy





We really want ontologies in electronic form

- Ontology contents can be processed and interpreted by computers
- Interactive tools can assist developers in ontology authoring

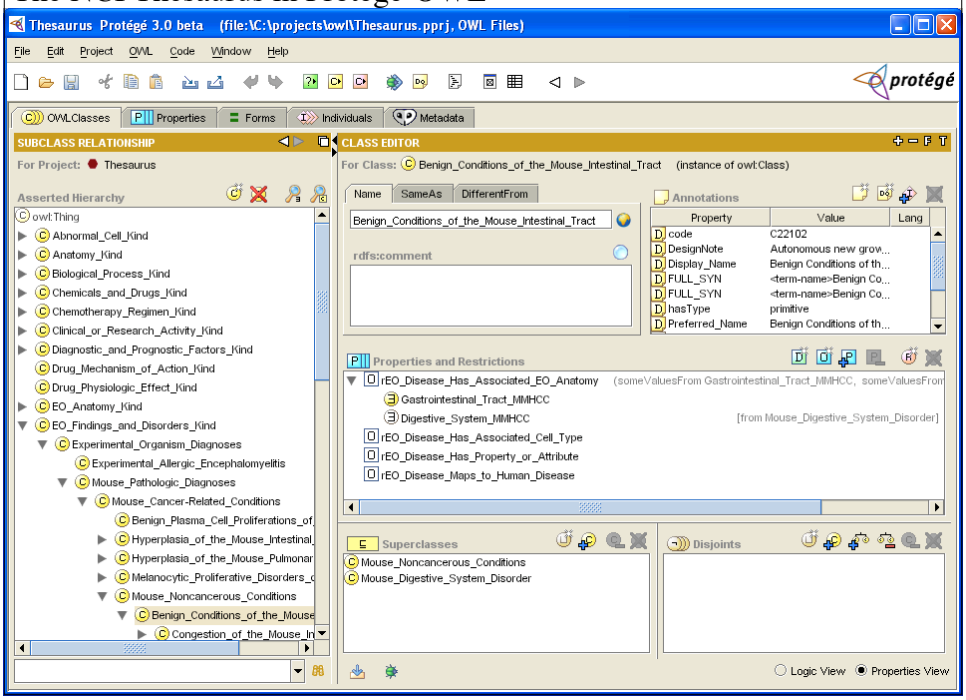


The FMA demonstrates that distinctions are not universal

- Blood is not a tissue, but rather a body substance (like saliva or sweat)
- The pericardium is not part of the heart, but rather an organ in and of itself
- Each joint, each tendon, each piece of fascia is a separate organ

These views are not shared by many anatomists!

The NCI Thesaurus in Protégé-OWL



Gene Ontology Consortium

- Founded in 1998 as a collaboration among scientists responsible for developing different databases of genomic data for model organisms (fruit fly, yeast, mouse)
- Now, essentially all developers of all model-organism databases participate
- Goal: To produce a dynamic, controlled vocabulary that can be applied to all organism databases even as knowledge of gene and protein roles in cells is accumulating and changing

Gene Ontology (GO)

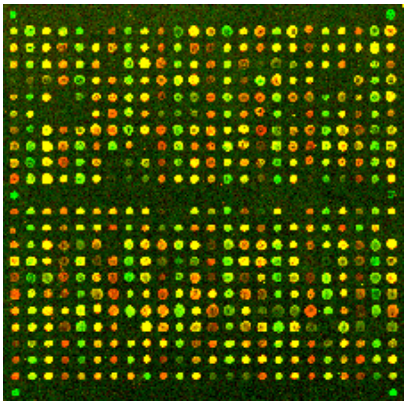
- Comprises three independent "ontologies"
  - **molecular function** of gene products
  - **cellular component** of gene products
  - **biological process** representing the gene product's higher order role.
- Terms express attributes of gene products in the collaborating databases, allowing queries across databases, providing linkage of biological information across species

GO is wildly successful

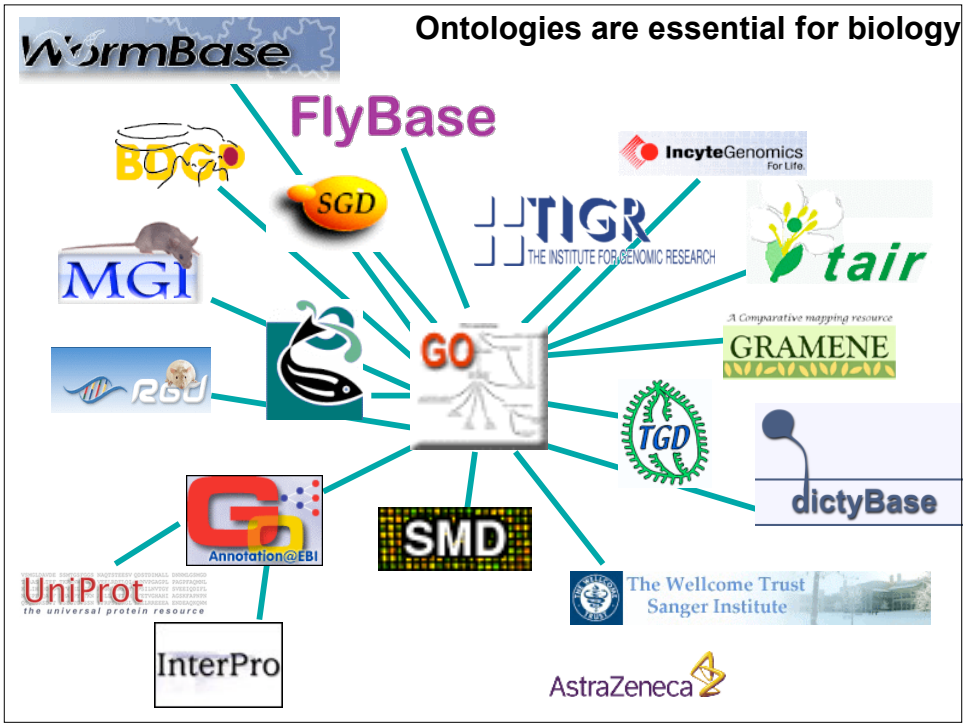
- The explosion of data associated with high-throughput experiments has demanded a mechanism for creating uniform data descriptions
- A large consortium of biologists from around the world contribute to GO on a regular basis
- It's now impossible to work in most areas of computational biology without making use of GO terms

The Microarray revolution ensured GO's success!

- Genes are expressed at different levels in different cells and tissues under different conditions
- Microarrays allow biologists to record changes in gene function across entire genomes
- Result: *Vast* amounts of gene expression data desperately needing cataloging and tagging



Ontologies are essential for biology



A Portion of the OBO Library

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

Knowledge workers seem trapped in a pre-industrial age

- Most ontologies are of relatively small scale
- Most ontologies are built and refined by small groups working arduously in isolation
- Success rests heavily on the particular talents of individual artisans, rather than on standard operating procedures
- Ontology developers often rally around particular “gurus” who have self-proclaimed skills and often strong biases
- There are few technologies on the horizon to make this process “faster, better, cheaper”

Creating bio-ontologies is a widespread cottage industry

- Professional Societies
  - MGED: Microarray Gene Expression Data Society Ontology
  - HUPO: Proteomics Standards Initiative
- Government
  - NCI Thesaurus
  - NIST: Process Specification Language
- Open Biomedical Ontologies
  - GO
  - Dozens of other biomedical ontologies

OBO
open biological ontologies

Lots of ontology builders are devoted 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 ontology engineers to go around

The University of Manchester

MANCHESTER 1824

A wonderful keynote talk from the meeting on *Standards and Ontologies for Functional Genomics* in 2004

The Capulets and Montagues

A plague on both your houses?

Professor Carole Goble
University of Manchester, UK

Warning:
This talk contains sweeping generalisations

MANCHESTER 1824

The University of Manchester

Prologue

Two households, both alike in dignity,
In fair genomics, where we lay our scene,
(One, comforted by its logic's rigour,
Claims ontology for the realm of pure,
The other, with blessed scientist's vigour,
Acts hastily on models that endure),
From ancient grudge break to new mutiny,
When "being" drives a fly-man to blaspheme.
From forth the fatal loins of these two foes
Researchers to unlock the book of life;
Whole misadventured piteous overthrows
Can with their work bury their clans' strife.
The fruitful passage of their GO-mark'd love,
And the continuance of their studies sage,
Which, united, yield ontologies undreamed-of,
Is now the hours' traffic of our stage;
The which if you with patient ears attend,
What here shall miss, our toil shall strive to mend.

Based on an idea by Shakespeare

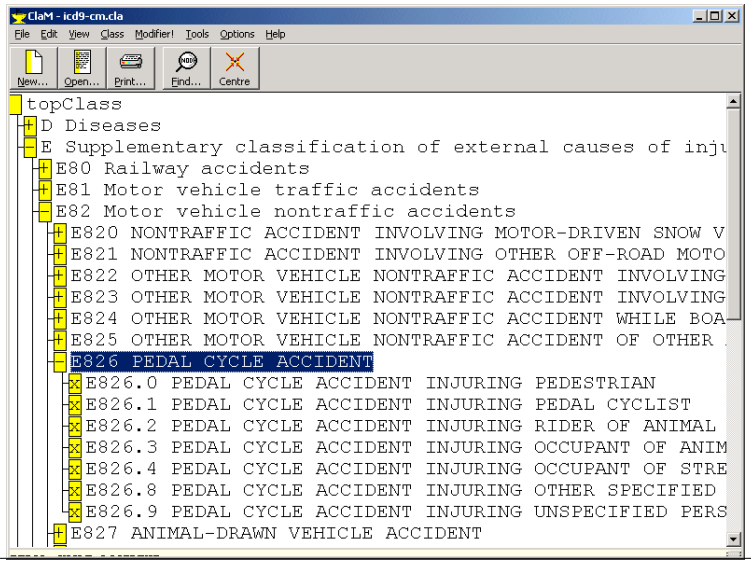
Legacy ontologies can cripple us:
The International Classification of Diseases

- An enumeration of diseases that forms the basis for all medical claims and reimbursements in the United States
- A "legacy" terminology that has its roots in 19th century epidemiology
- Created initially by biostatisticians with a pressing need to compare death statistics in different European countries
- A system that won't go away—and yet we would never create anything like it again

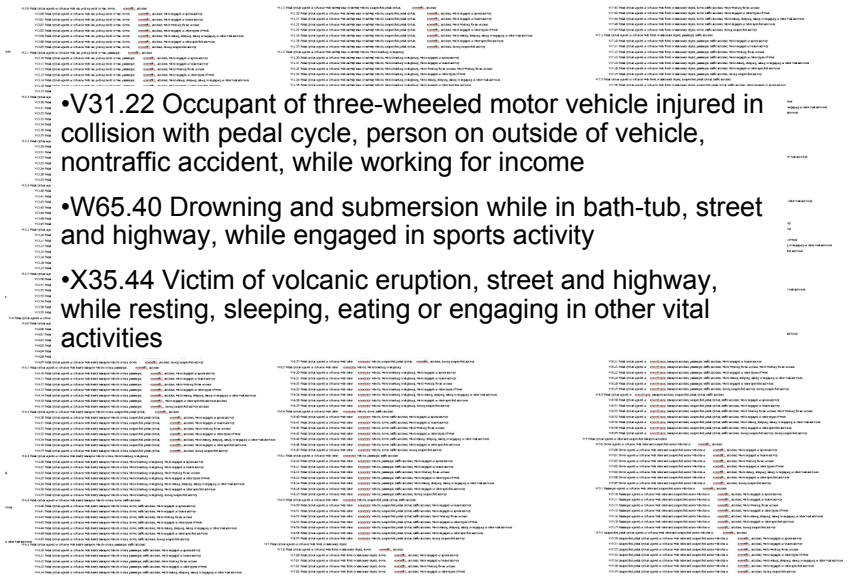
The International Classification of Diseases

724	Unspecified disorders of the back
724.0	Spinal stenosis, other than cervical
724.00	Spinal stenosis, unspecified region
724.01	Spinal stenosis, thoracic region
724.02	Spinal stenosis, lumbar region
724.09	Spinal stenosis, other
724.1	Pain in thoracic spine
724.2	Lumbago
724.3	Sciatica
724.4	Thoracic or lumbosacral neuritis
724.5	Backache, unspecified
724.6	Disorders of sacrum
724.7	Disorders of coccyx
724.70	Unspecified disorder of coccyx
724.71	Hypermobility of coccyx
724.71	Coccygodynia
724.8	Other symptoms referable to back
724.9	Other unspecified back disorders

ICD9 (1977): A Handful of Codes for Traffic Accidents



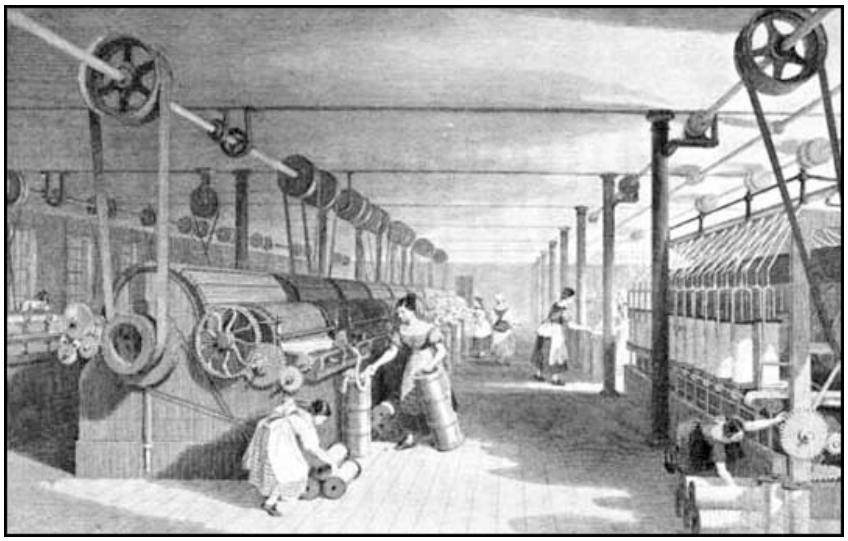
ICD10 (1999): 587 codes for such accidents



- V31.22 Occupant of three-wheeled motor vehicle injured in collision with pedal cycle, person on outside of vehicle, nontraffic accident, while working for income
- W65.40 Drowning and submersion while in bath-tub, street and highway, while engaged in sports activity
- X35.44 Victim of volcanic eruption, street and highway, while resting, sleeping, eating or engaging in other vital activities

If real ontologists could build the ICD from scratch …

- Diseases would be organized with well-defined relationships
- Diseases would be associated with computer-understandable definitions
- There would be well-defined rules for ensuring that descriptions are sensible
- There would be well-defined mechanisms for creating use-specific views of the ICD
- There would be a well-defined path to integration with bioinformatics resources that describe the molecular underpinnings of disease



NIH Roadmap ACCELERATING MEDICAL DISCOVERY TO IMPROVE HEALTH

- Overview
- NIH Roadmap Initiatives
- Funding Opportunities
- Funded Research
- Roadmap Related Activities
- Public Meetings and Workshops
- Frequently Asked Questions
- News and Information
- NIH Roadmap Institute and Center Liaisons
- Subscribe to the NIH Roadmap E-mail list

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

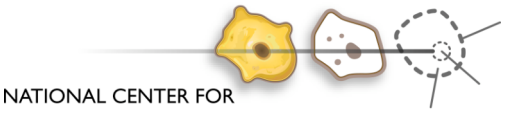
- Re-engineering the Clinical Research Enterprise Initiatives
  - Clinical Research Networks and NECTAR
  - Clinical Outcomes Assessment
  - Clinical Research Training
  - Clinical Research Policy Analysis and Coordination
  - Translational Research

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

The National Center for Biomedical Ontology

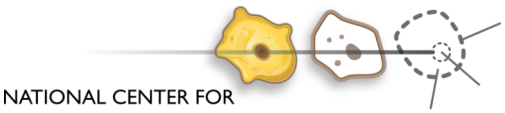
- One of three National Centers for Biomedical Computing launched by NIH in 2005
- Collaboration of Stanford, Berkeley, Mayo, Buffalo, Victoria, UCSF, Oregon, and Cambridge
- Primary goal is to make ontologies accessible and usable
- Research will develop technologies for ontology dissemination, indexing,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



NATIONAL CENTER FOR
BIOMEDICAL ONTOLOGY

Goals for the NCBO

- Our **vision** is that all biomedical knowledge and data are disseminated on the Internet using principled ontologies in such a way that the knowledge and data are semantically interoperable and useful for furthering biomedical science and clinical care.
- Our **mission** is to create software and support services for the application of principled ontologies in biomedical science and clinical care, ranging from tools for application developers to software for end-users.



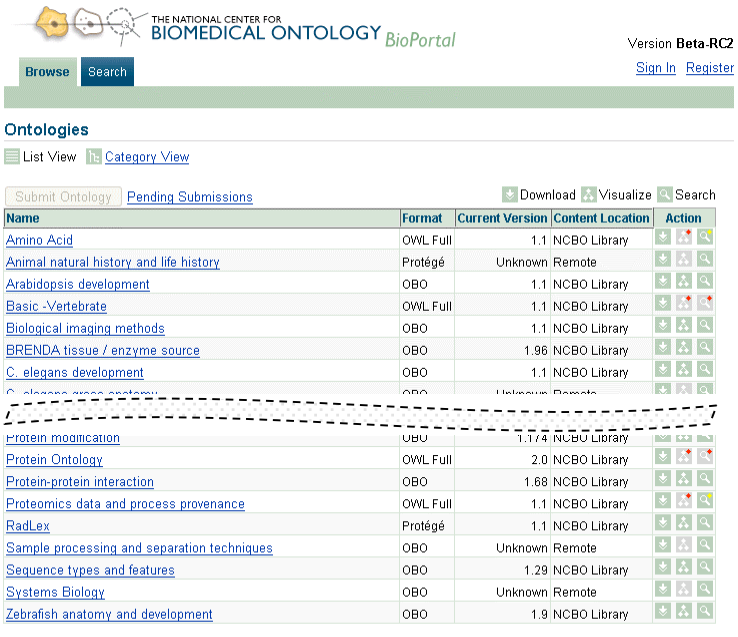
NATIONAL CENTER FOR
BIOMEDICAL ONTOLOGY

Our Center will offer

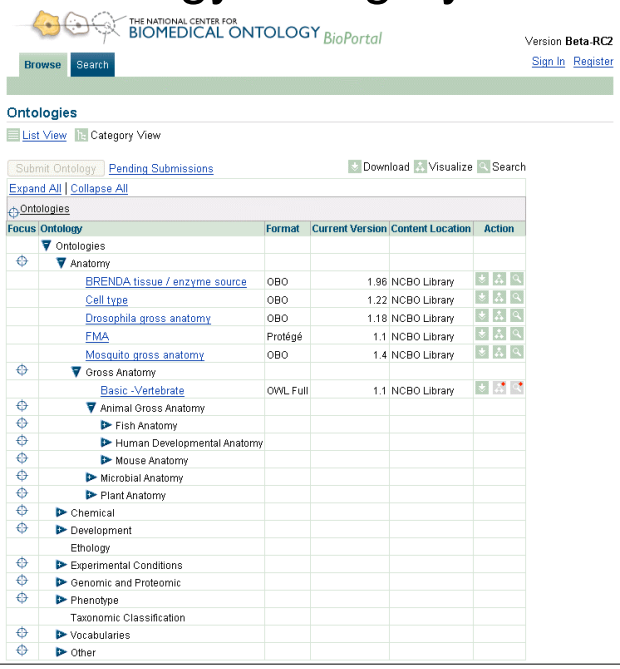
- Technology for uploading, browsing, and using biomedical ontologies
- Methods to make the online "publication" of ontologies more like that of journal articles
- Tools to enable the biomedical community to put ontologies to work on a daily basis

NATIONAL CENTER FOR
BIOMEDICAL ONTOLOGY

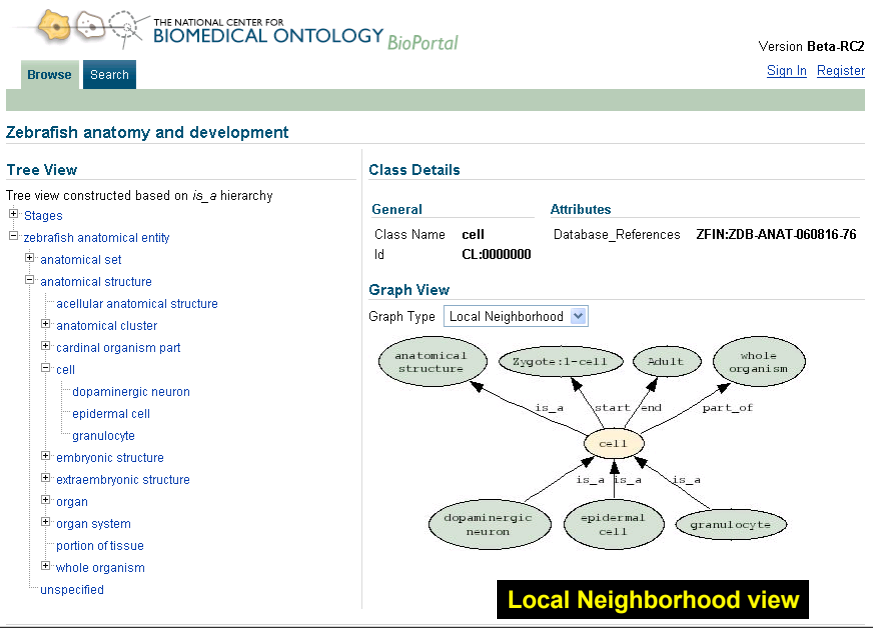
http://bioportal.bioontology.org

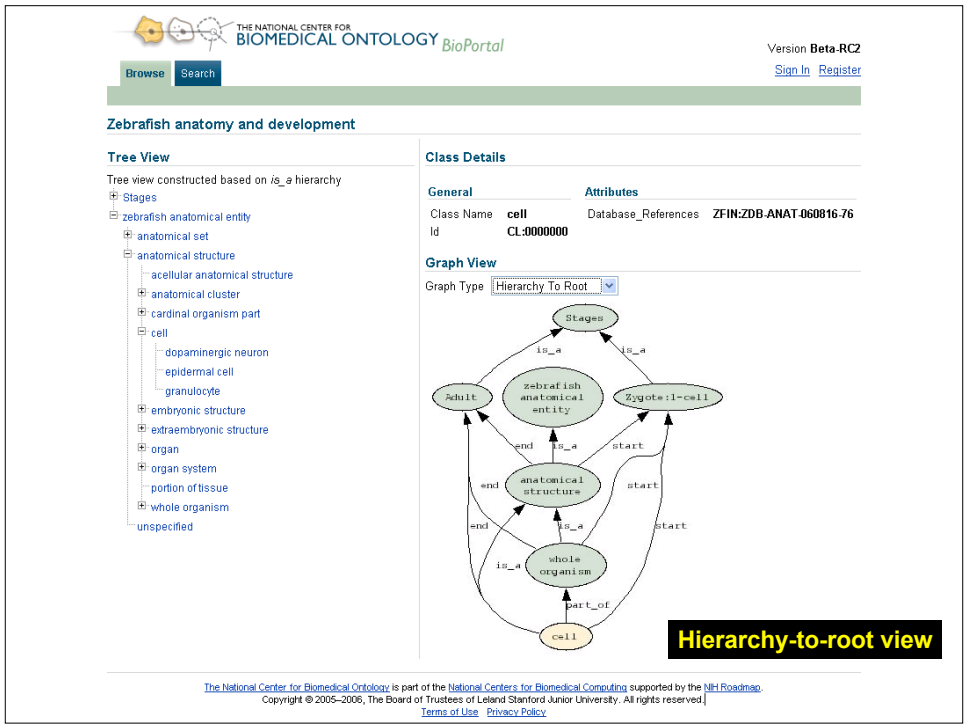


Ontology Category View

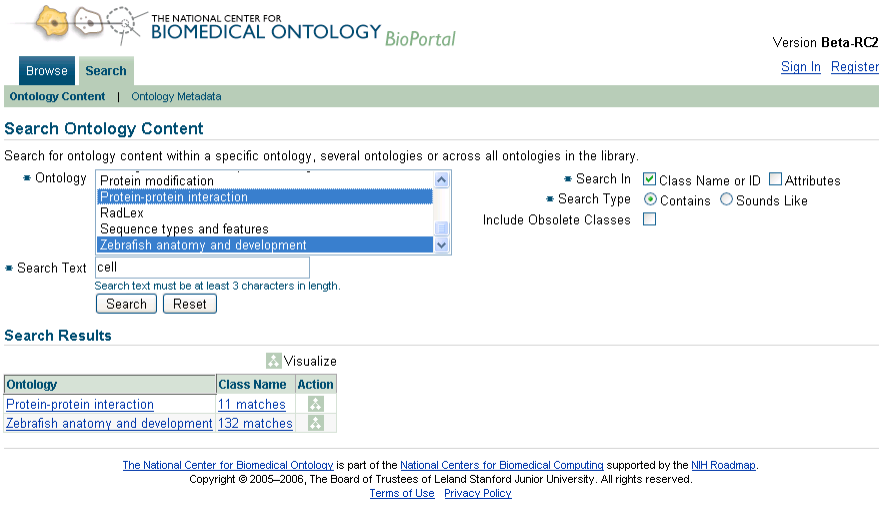


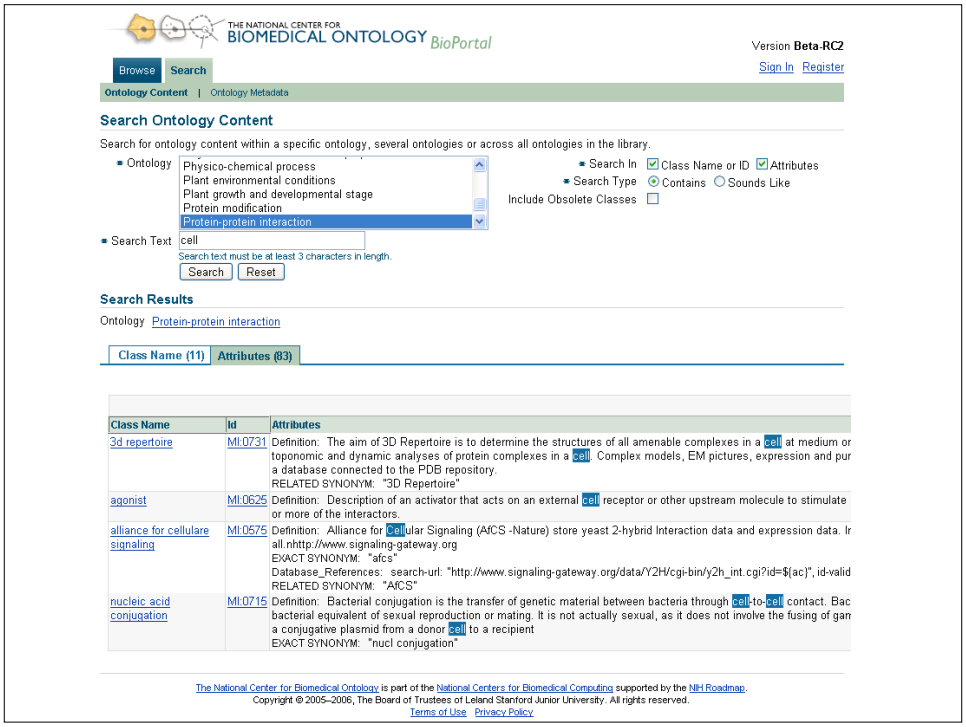
Browsing/Visualizing Ontologies





Searching for terms within ontologies in BioPortal





BioPortal's impact in the community

- National Cancer Institute
  - Deploying BioPortal locally to evaluate its use as preferred method for visualizing/navigating enterprise terminologies and ontologies
- Biomedical Informatics Research Network (BIRN)
  - Adopting BioPortal for disseminating and visualizing BIRNLex terminology to research community
- Radiological Society of North America
  - Using BioPortal to supplant home-built ontology viewer for graphical visualization of RadLex

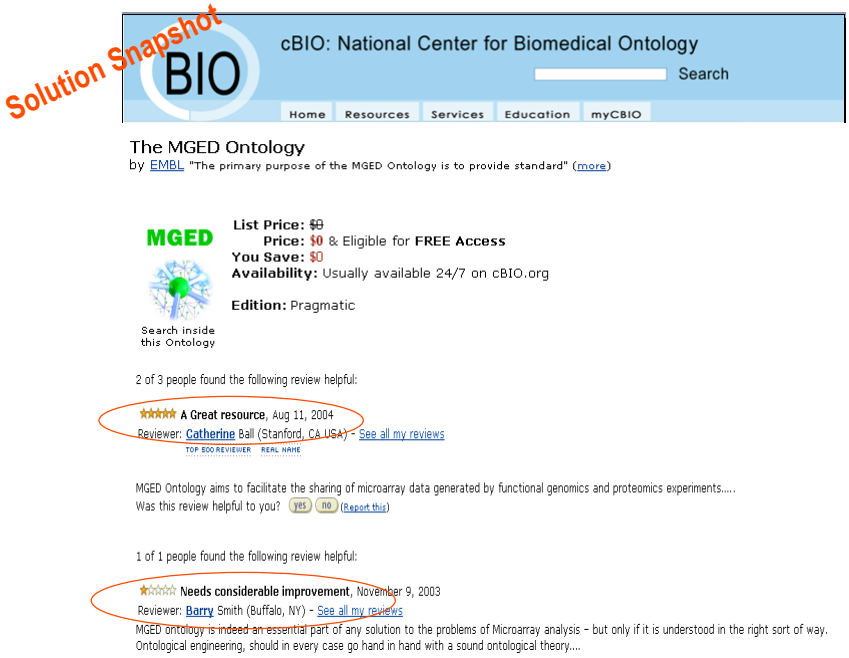
A problem in both technology and sociology

- 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?


amazon.com.
VIEW CART | WISH LIST | YOUR ACCOUNT | HELP
WELCOME | KAUSTUBH'S STORE | BOOKS | APPAREL & ACCESSORIES | ELECTRONICS | TOYS & GAMES | MUSIC | COMPUTER & VIDEO GAMES | SEE MORE STORES
SEARCH | BROWSE SUBJECTS | BESTSELLERS | THE NEW YORK TIMES® BEST SELLERS | MAGAZINES | CORPORATE ACCOUNTS | E-BOOKS & DOCS | BARGAIN BOOKS | USED BOOKS
Save up to 60% Choose from 1000+ Toys & Games! TOYSRUS.COM teamed with amazon.com Shop now
Join Amazon Prime and ship Two-Day for free and Overnight for $3.99.
The Da Vinci Code
by Dan Brown "Robert Langdon awoke slowly..." (more)
SEARCH INSIDE!
Search inside this book
See 1 customer image
Share your own customer images
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
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 skilfully 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.
Was this review helpful to you? yes no (Report this)
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 potenti reader is lead to believe it is, TDVC is turgid, jerky, and filled with clichés. The characters are characterless and stupid, merely cardboard for the author to push around like pawn 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 tc 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 peop are idiot savants. Whatever it is, it's exasperating.


amazon.com.
VIEW CART | WISH LIST | YOUR ACCOUNT | HELP
WELCOME | KAUSTUBH'S STORE | BOOKS | APPAREL & ACCESSORIES | ELECTRONICS | TOYS & GAMES | MUSIC | COMPUTER & VIDEO GAMES | SEE MORE STORES
SEARCH | BROWSE SUBJECTS | BESTSELLERS | THE NEW YORK TIMES® BEST SELLERS | MAGAZINES | CORPORATE ACCOUNTS | E-BOOKS & DOCS | BARGAIN BOOKS | USED BOOKS
Save up to 60% Choose from 1000+ Toys & Games! TOYSRUS.COM teamed with amazon.com Shop now
Join Amazon Prime and ship Two-Day for free and Overnight for $3.99.
The Da Vinci Code
by Dan Brown "Robert Langdon awoke slowly..." (more)
SEARCH INSIDE!
Search inside this book
See 1 customer image
Share your own customer images
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
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 skilfully 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.
Was this review helpful to you? yes no (Report this)
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 potenti reader is lead to believe it is, TDVC is turgid, jerky, and filled with clichés. The characters are characterless and stupid, merely cardboard for the author to push around like pawn 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 tc 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 peop are idiot savants. Whatever it is, it's exasperating.

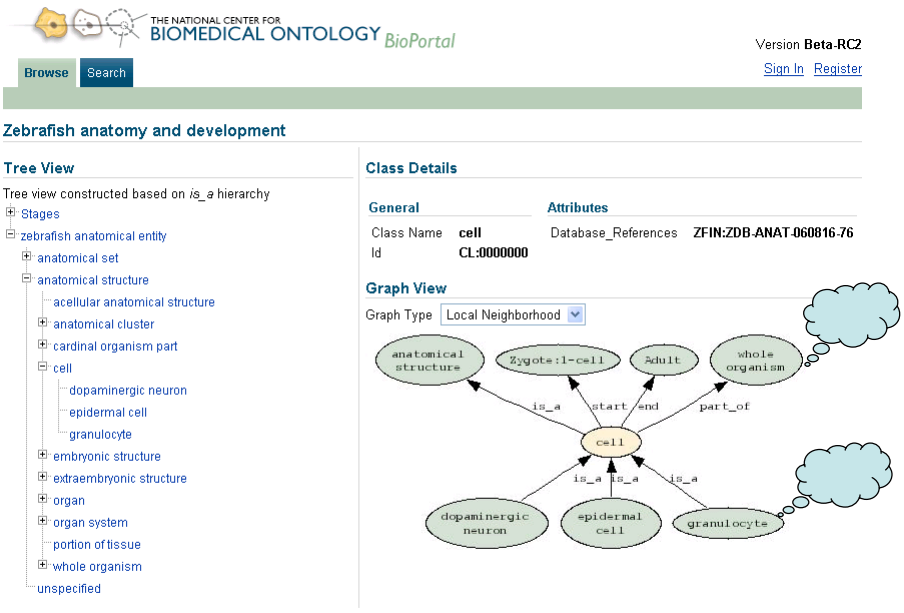
Solution Snapshot



Under development: 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”

Marginal notes in BioPortal



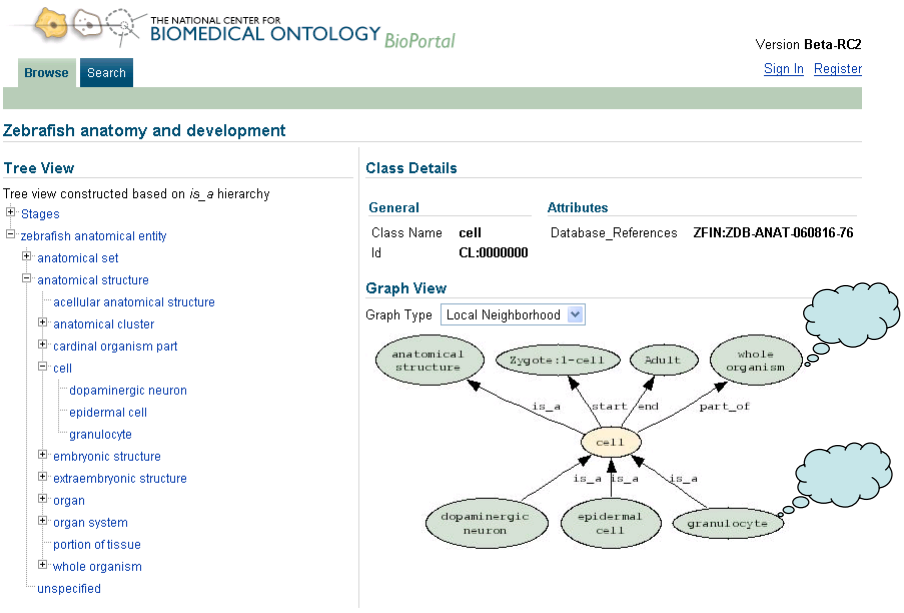
A Curated Approach for Quality Assurance

- 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

Unsettled questions in the ontology ratings game

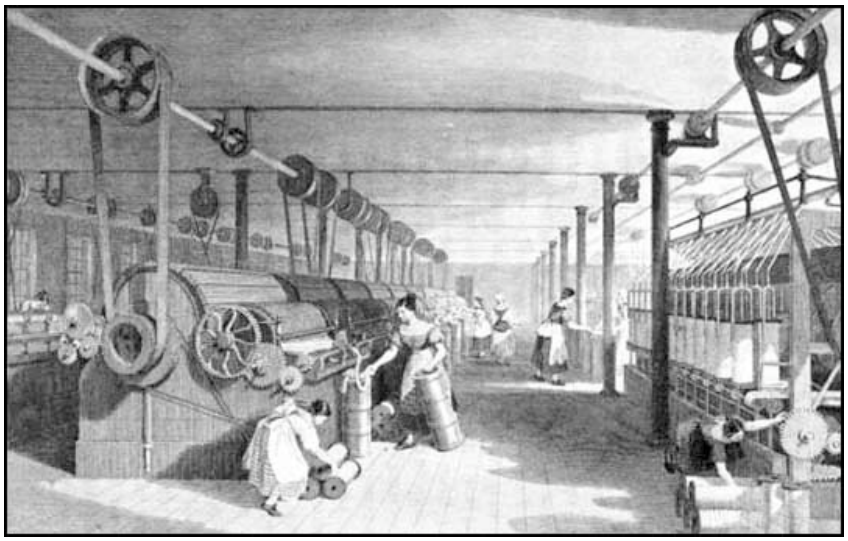
- Can the OBO Foundry scale?
- Who gets to reject an ontology on the basis of form or content?
- Who will curate the curators?
- Alternatively, will end-users really use an open rating system?
- Are there enough raters out there to provide sufficient feedback?

Marginal notes in BioPortal



BioPortal will 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 industrial revolution gave us

- Standardized, interchangeable parts
- Technologies for creating new technologies
- Tremendous increase in output
- Unparalleled incentives for innovation

The foundation is in place

- Scientific culture now recognizes the importance of ontologies
- We are beginning to articulate best practices for ontology construction
- We have a burgeoning cottage industry at work

We are moving beyond one-off ontologies and one-off tools to:

- Web services to provide access to ontologies throughout cyberspace
- Meta-data standards for ontology annotation
- Comprehensive methods for ontology alignment, indexing, and retrieval
- End-user platforms for putting ontologies to use for
  - Data integration
  - Decision support
  - Natural-language processing
  - Information retrieval
  - And applications that we have not yet thought of!


THE NATIONAL CENTER FOR
BIOMEDICAL ONTOLOGY
About
Resources
News and Events
Education
Dissemination
Collaboration
Advancing biology and medicine with tools and methodologies
for the structured organization of knowledge.
Resources
Education
Dissemination
Driving
Biological
Projects
Collaboration