Ontology-based annotation of multiscale imaging data: Utilizing and building the Neuroscience Information Framework

Maryann E. Martone

University of California, San Diego
What does this mean?

• 3D Volumes
• 2D Images
• Surface meshes
• Tree structure
• Ball and stick models
• Little squiggly lines

Has part
Specific Aims

- Development of multiscale Phenotype and Trait Ontology for neurological disease
- Development of an intuitive ontology-based environment for multiscale image annotation and analysis
- Extension of the current BIRN Integration Environment, now part of the Neuroscience Information Framework
  - Matching animal model to disease
  - Alignment of ontology tools/data from NCBO and data integration tools and large imaging data repository from NIF
Neuroscience Information Framework

• Funded by the NIH Blueprint
• How can we provide a consistent and easy to implement framework for those who are providing resources, e.g., data, and those looking for resources
  • Both humans and machines
• Strong foundation for data integration
• Means to query “hidden” web content in databases
• Interface for searching across multiple types of resources
• Ontologies for neuroscience
  • A consistent way to describe data
  • “Concept-based searching”
  • Sorting and understanding results

http://neuinfo.org
Integrated Search

- NIF Registry
  >450 web resources annotated by humans with NIF vocabularies
- NIF Neuroscience Web
  Custom web index built using open source web tools (Nutch) from the NIF registry
- Neuroscience literature
  ~70,000 articles, fully indexed using Textpresso tool
- NIF Data Federation
  Web accessible databases registered to NIF mediator for deep content searching
- Other portals
  Existing web resources that are themselves portals to resources
  Science.gov

http://neuinfo.org
NIF Architecture

Gupta et al., Neuroinformatics, 2008 Sep;6(3):205-17
Current Structure of NIFSTD

- NIF 0.5: http://purl.org/nif/ontology/nif.owl
- NCBO was instrumental in getting us started; incorporated BIRNLex
- Built from existing ontologies/vocabularies where possible
- Single inheritance trees with minimal cross domain and intradomain properties
  - Building blocks for additional ontologies
- Meant to be maximally useful by humans and information systems
How NIFSTD is used in NIF

• Controlled vocabulary for describing type of resource and content
  – Database, Image, Parkinson’s disease

• Concept-mapping of database content
  – Concept mapping tools from BIRN
    • Map table names, field names and values
    • Brodmann.2 (SUMSDB) = birnlex_1733

• Textpresso literature mining

• Search: Mixture of mapped content and string-based search using ontology
  – Originally used strict mappings
    • “You can search for anything you want as long as it’s a Purkinje cell”
  – More effective search and organization of results
Resource Descriptors

- Data
- Software
- Material
- Service
- Funding
- Training

*Open issue: Harmonization of NIF Resource descriptors and BRO. What does that mean?*

*Biositemaps + NIF Disco protocol*
NeuroLex Wiki

- Easy to add new classes, synonyms, definitions
- Critical when annotating data
- Easy to modify existing entries
- Easy to navigate NIFSTD “is a” hierarchy and generate custom tables, e.g., all brain regions and their definitions
- Can set up templates to simplify input
- Other groups: BioMedGT

http://neurolex.org

Stephen Larson
Jinx: Ontology Based Image Annotation

Stephan Lamont

Draws from NIF classes; can add own
Tying annotations to spatial regions

Mark Gibson, Nicole Washington, Suzie Lewis, Willy Wong, Asif Memon, Sarah Maynard, Stephen Larson
Image Annotation of 3D Microscopic Imaging data

- Annotation during segmentation
- Fluid process
  - Constantly reassigning classes and reorganizing trees
  - Must be able to add new classes at will
- Must be able to use your own tool
- How can ontologies help automated and semi-automated segmentation?
  - How can segmentation inform ontologies
Cellular Knowledge Base

Find all instances of spines that contain membrane bound organelles

- Ontology + instances
- RDF triple store; SPRKQL
- Knowledge-based queries
- Content-based retrieval
- Contains knowledge from literature, CCDB
- Query interface fairly simple

CCDB: data properties and experimental details
CKB: Biological view

Willy Wong, Amarnath Gupta, Bill Bug
Phenotypes: NIF + PATO
Define a Phenotype Template

A template for describing instances of phenotypes contained in images or the literature was created on top of the NIF ontologies; brings together experimental and biological entities.

Sarah Maynard, Stephen Larson, Bill Bug, Amarnath Gupta, Chris Mungall, Suzie Lewis
Phenotype 037

Alpha-Synuclein Protein

inheres in

is borne by

Sprague Dawley 037

is bearer of

aggregated 037

Find all diseases that have a phenotype in which Alpha-Synuclein is aggregated:

- is bearer of some (inheres in some
  (‘Alpha-Synuclein Protein’ and is bearer of some Aggregated))

- Used Protégé 4 beta and Pellet 1.5

'Levy Body Disease'
'Multiple System Atrophy'
'Parkinsons disease'
Substantia Nigra degenerates = dopamine neurons in substantia nigra decrease in number

Multiscale phenotypes
Experimental vs Biological Phenotypes

- The significance of a staining pattern is not always known
- Electron microscopic evidence indicates that these two phenotypes are related
- Effective addition of human knowledge

Class Level Description

\[ \text{inheres_in some (Alpha-Synuclein Protein and (is_bearer_of some Aggregated))} \]
OBD

OBD finds information content and semantic similarity based on a reasoned database in order to find similar phenotypes

Chris Mungall, Suzanna Lewis
Issues

• How do we truly manage co-development?

• Modular ontologies like the NIFSTD/BIRNLex were not well served by the BioPortal
  • Doesn’t reflect all the mappings that were done
  • What about when we use the same concept?
  • Would like to be able to search within a single ontology from the Explore view

• Flexible tools for application, utilization and building of ontologies in multiple contexts
  • How can we build them to maximize their utility in the broadest number of applications
  • No single approach works for everything
  • Cannot be caught in the ontology wars
    • Would also like some explicit declarations of who the combatants are
  • Best practices: URI’s, versioning

• Creating views on ontologies: hiding semantic complexity from the end user
  • Build from simple modules
  • Abstract from complex ontologies

• For collaborative ontology building by domain experts, the Wiki approach appears to be much more powerful
  • Better integration of Semantic Wiki’s into ontology workflow

• Ontology-based image annotation
  • Make more use of spatial information contained in images to describe data
  • Automated and semi-automated annotation and ontology building; analogous to NLP

• Disease phenotypes: Major challenge but great opportunity
  • No single way to handle phenotypes but some overarching consistency will go a long way