The use of ontologies within the Neuroscience Information Framework, a neuroscience-centered portal for searching and accessing diverse resources

Maryann Martone, Ph. D.
University of California, San Diego
NIF Team

Amarnath Gupta, UCSD, Co Investigator
Jeff Grethe, UCSD, Co Investigator
Gordon Shepherd, Yale University
Perry Miller
Luis Marenco
David Van Essen, Washington University
Erin Reid
Paul Sternberg, Cal Tech
Arun Rangarajan
Hans Michael Muller
Giorgio Ascoli, George Mason University
Sridevi Polavarum
Anita Bandrowski, NIF Curator
Fahim Imam, NIF Ontology Engineer
Karen Skinner, NIH

Lee Hornbrook
Kara Lu
Vadim Astakhov
Xufei Qian
Chris Condit
Stephen Larson
Sarah Maynard
Bill Bug

Karen Skinner, NIH, Program Officer
What does this mean?

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

Data

People

Information systems
The Neuroscience Information Framework: Discovery and utilization of web-based resources for neuroscience

UCSD, Yale, Cal Tech, George Mason, Washington Univ

- Provides access to neuroscience resources on the web
- Provides simultaneous search of multiple types of information, organized by category: Databases, literature, web pages
- Supported by an expansive ontology for neuroscience
- Utilizes advanced technologies to search the “hidden web”, i.e., information that can’t be found by Google
- Text mining tools for literature
- Database mediators

http://neuinfo.org

Supported by NIH Blueprint
Where do I find...

- Data
- Software tools
- Materials
- Services
- Training
- Jobs
- Funding opportunities

- Websites
- Databases
- Catalogs
- Literature
- Supplementary material
- Information portals

...Lots and lots of them
What is NIF?

A dynamic inventory of Web-based neuroscience resources: data, materials, and tools accessible via any computer connected to the Internet.

An initiative of the NIH Blueprint for Neuroscience Research, the Neuroscience Information Framework advances neuroscience research by enabling discovery and access to public research data and tools worldwide through an open source, networked environment.

NIF in action

Search for Neuroscience Resources

NIF Tools | Vocabularies | Webinars | Community News | Developers

NIF Community

What's this?

Follow Us on Twitter
Facebook
Subscribe to our RSS
Email
Print

Is your Resource Registered with NIF?
Get your Registered with NIF Button.

News & Events

Neuroscience 2008

Current Release Ver. 2.0

Release Notes
NIF 2.0

- New look and feel
- Better query expansion
- Category browsing of NIF registry
- Ontology-based ranking of web results
- First export of data in RDF
- First web services released
- First MyNIF features
- Release of DISCO tool suite
Integrated views and gene search
Guiding principles of NIF

• Builds heavily on existing technologies (BIRN, open source tools)
• Information resources come in many sizes and flavors
• Framework has to work with resources as they are now
  – Federated system; resources will be independently maintained
  – But…moving forward there are things that resource providers can do that will make things a lot easier
• No single strategy will work for the current diversity of neuroscience resources
• Trying to design the framework so it will be as broadly applicable as possible to those who are trying to develop technologies
• Interface neuroscience to the broader life science community
• Take advantage of emerging conventions in search and in building web communities
Registering a Resource to NIF

Level 1
NIF Registry: high level descriptions from NIF vocabularies supplied by human curators

Level 2
Access to deeper content; mechanisms for query and discovery; DISCO protocol

Level 3
Direct query of web accessible database
Automated registration
Mapping of database content to NIF vocabulary by human
The NIF Registry
Level 3

• Deep query of federated databases with programmatic interface
• Register schema with NIF
  – Expose views of database: try to create views that are simple and easy to understand for NIF users
  – Map vocabulary to NIFSTD
• Currently works with relational and XML databases
  – RDF capability planned for NIF 2.5 (April 2010)
• Works with NIF registry: databases also annotated according to data type and
Level 2: Updates and deeper integration

- DISCO involves a collection of files that reside on each participating resource. These files store information describing:
  - attributes of the resource, e.g., description, contact person, content of the resource, etc. -> updates NIF registry
  - how to implement DISCO capabilities for the resource

- These files are maintained locally by the resource developers and are “harvested” by the central DISCO server.

- In this way, central NIF capabilities can be updated automatically as resources evolve over time.

- The developers of each resource choose which DISCO capabilities their resource will utilize

Luis Marenco, MD, Rixin Wang, PhD, Perry L. Miller, MD, PhD, Gordon Shepherd, MD, DPhil
Yale University School of Medicine
Interoperability: DISCO-Biositemaps

- Foundational DISCO very similar to Biositemaps
- NIF DISCO recently reconciled its basic resource description with Biositemaps
- NIF can now import Biositemaps

Anita Bandrowski and Luis Marenco
DISCO Level 2 Interoperation

• Level 2 interoperation is designed for resources that have only Web interfaces (no database API).

• Different resources require different approaches to achieve Level 2 interoperation. Examples are:
  • CRCNS - requires metadata tagging of Web pages
  • DrugBank - requires directed traversal of Web pages to extract data into a NIF data repository
  • GeneNetwork - requires Web-based queries to achieve “relational-like” views using “wrappers”
DrugBank Example

The DrugBank Web interface showing data about a specific drug (Phentoin).
DrugBank Example (continued)

This DISCO Interoperation file specifies how to extract data from the DrugBank Web interface automatically.
A NIF user views data retrieved from DrugBank in response to a query in a transparent, integrated fashion.
Four Things You Can Do to Make Your Database More Interoperable

So here are our top 4 barriers to data federation in the NIF:

1) Unstable identifiers: Every time the database updates, the identifiers change and all pre-indexed links to those data records break;

2) Access: For increased utilization of the data, stable access needs to be provided either through a public connection to the database, a periodic dump of the database contents or through structured web services;

3) Sessions: For general information results and data should be accessible using a static (i.e. non session based or stateless) URL;

4) Vocabulary: Use a standard terminology and avoid symbolic notations where possible.
How are ontologies used in NIF?

• Search: query expansion
  – Synonyms—try to smooth over differences without explicit mapping
  – Related classes
  – “concept based queries”: what I mean not what I say

• Annotation:
  – Resource categorization
  – Entity mapping—incremental process

• Ranking of results
  – NIF Registry; NIF Web
Modular ontologies for neuroscience

- NIF covers multiple structural scales and domains of relevance to neuroscience
- Incorporated existing ontologies where possible; extending them for neuroscience where necessary
- Normalized under the Basic Formal Ontology: an upper ontology used by the OBO Foundry
- Based on BIRNLex: Neuroscientists didn’t like too many choices
- Cross-domain relationships are being built in separate files
- Encoded in OWL-DL, but also maintained in a Wiki form, a relational database form and any other form it is needed in
Balancing act

• Different schools of thought as to how to build vocabularies and ontologies

• NIF is trying to navigate these waters, keeping in mind:
  – NIF is for both humans and machines
  – Our primary concern is data
  – We have to meet the needs of the community
  – We have a budget and deadlines

• Building ontologies is difficult even for limited domains, never mind all of neuroscience, but we’ve learned a few things
  – Reuse what’s there: trying to re-use URI’s rather than map when possible
  – Make what you do reusable: adopt best practices where feasible
    • Numerical identifiers, unique labels, single asserted simple hierarchies
  – Engage the community
  – Avoid “religious” wars: separate the science from the informatics
  – Start simple and add more complexity
    • Create modular building blocks from which other things can be built
What are we doing?

• Strategy: Create modular building blocks that can be knit into many things
  – Step 1: Build core lexicon (NeuroLex)
    • Classes and their definitions
    • Simple single inheritance and non-controversial hierarchies
    • Each module covers only a single domain
  – Step 2: NIFSTD: standardize modules under same upper ontology
  – Step 3: Create intra-domain and more useful hierarchies using properties and restrictions
    – Brain partonomy
  – Step 4: Bridge two or more domains using a standard set of relations
Neurolex

• More human centric
• Synonyms and abbreviations were essential for users
  • Can’t annotate if they can’t find it
  • Can’t use it for search if they can’t find it
  • Facilitates semi-automated mapping
• Contains subsets of ontologies that are useful to neuroscientists
  – e.g., only classes in Chebi that neuroscientists use
• Wanted the community to be able to see it and use it
  – Simple understandable hierarchies
Category: Amyotrophic Lateral Sclerosis

- ID: bimlax_12566
- Definition: A degenerative disorder affecting upper MOTOR NEURONS in the brain and lower motor neurons in the brain stem and SPINAL COPD. Disease onset is usually after the age of 50 and the process is usually fatal within 3 to 5 years. Clinical manifestations include progressive weakness, atrophy, FASCICULATION, hyperreflexia, DYSARTHRIA, dysphagia, and eventual paralysis of respiratory function. Pathologic features include the replacement of motor neurons with fibrous ASTROCYTES and atrophy of anterior SPINAL NERVE ROOTS and corticospinal tracts (MeSH).
- Synonym: Gehrig's Disease
- Synonym: Lou Gehrig's Disease
- Synonym: Maladie de Charcot
- Definition Source: MeSH_defSource
- Definition Citation: Adams et al., Principles of Neurology, 6th ed, pp1089-94
- Acronym: ALS
- Curation Status: uncurated
- Created Date: 2007-10-05

Query for more information

Click here to find more about Amyotrophic Lateral Sclerosis

Facts about Amyotrophic Lateral Sclerosis

- Acronym: ALS
- Created: 5 October 2007
- Curation Status: uncurated
- Defining Citation: Adams et al., Principles of Neurology, 6th ed, pp1089-94
- Definition: A degenerative disorder affecting upper MOTOR NEURONS in the brain and lower motor neurons in the brain stem and SPINAL COPD. Disease onset is usually after the age of 50 and the process is usually fatal within 3 to 5 years. Clinical manifestations include progressive weakness, atrophy, FASCICULATION, hyperreflexia, DYSARTHRIA, dysphagia, and eventual paralysis of respiratory function. Pathologic features include the replacement of motor neurons with fibrous ASTROCYTES and atrophy of anterior SPINAL NERVE ROOTS and corticospinal tracts (MeSH).
- Synonym: Gehrig's Disease, Lou Gehrig's Disease, and Maladie de Charcot

Category: Neurodegenerative disease

Subcategories

This category has the following 3 subcategories, out of 3 total.

A
- Amyotrophic Lateral Sclerosis

P
- Progressive Bulbar Palsy

S
- Spinal Muscular Atrophy

Category: Motor Neuron Disease
NeuroLex Wiki

- Uses Semantic Media wiki software
- Each class becomes a category page
- Good way to train neuroscientists on ontology construction
- Supports automatic classification based on properties
- Has custom forms for different entities, e.g., brain regions vs neurons
- When the parent is assigned, the correct form is provided
- Has simple human understandable properties
- Curated

http://neurolex.org

Stephen Larson
Maintaining multiple versions

- NIF maintains the NIF vocabularies in different forms for different purposes
  - Neurolex Wiki: Lexicon for community review and comment
  - NIFSTD: set of modular OWL files normalized under BFO and available for download
  - NCBO Bioportal for visibility and services
  - Ontoquest: NIF’s ontology server
    - Relational store customized for OWL ontologies
    - Materialized inferred hierarchies for more efficient queries
NIF Architecture

Gupta et al., Neuroinformatics, 2008 Sep;6(3):205-17
NIF Ontology Curation Workflow

1. Add/Edit NeuroLex Terms/Categories
2. Bulk Upload Request of Terms
3. Identify Valid Contributions
4. Update NIFSTD (Testing)
5. Testing in OntoQuest
6. Testing in BioPortal
7. Keep Persistent Link to Older Versions
8. Update Project Wiki Release Notes
9. Update NIFSTD (Production)
10. Update NeuroLex
11. Update OntoQuest
12. Update BioPortal
13. Update TextPresso Bucket
Reuse principle in practice

• Biggest impediment to query across distributed data repositories is terminology

• Reuse of community ontologies good idea

• NIF tries to do this but…
  – Ontologies aren’t ready
  – Didn’t know an ontology existed
    • Divergence $\rightarrow$ convergence?
  – Ontologies aren’t constructed in a way where they could be utilized across different applications
• Working with NCBC (Biomedical Resource Ontology) and NITRC to come up with single resource ontology and information model
• Reconciling current versions; trying to move forward jointly
• Same classes, different views?

Peter Lyster, Csongor Nyulas, David Kennedy, Maryann Martone, Anita Bandrowski
Applying NIF principles

• NIFSTD module: NIF investigation
  • Based on BFO–OBI
  • Contains objects that are related to resource types, e.g., software applications; instruments
  • Extremely human unfriendly
    – Realizable entity?
    – Over normalized

• NIFSTD module: NIF Resource
  – Independent–dependent? Punted
  – Resource categories, e.g., software resources
  – Objects from NIFSTD reclassified under resource categories using very simple logical restrictions

• NIF resource browser
  – Assigns alternate labels that are easier to understand
Evolution of the NIF Resource Ontology

<table>
<thead>
<tr>
<th>Object</th>
<th>Function</th>
<th>Target Audience</th>
<th>Data Type</th>
<th>Data Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>Materials</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Biomaterials</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Reagents</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Software</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>People</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grants</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jobs</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Information</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Service</td>
<td>General</td>
<td>Structured</td>
<td>Text</td>
</tr>
<tr>
<td></td>
<td>- Storage</td>
<td>Kids</td>
<td>- Database</td>
<td>RDF Text</td>
</tr>
<tr>
<td></td>
<td>- Production</td>
<td></td>
<td>- Atlas</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Funding</td>
<td>Student</td>
<td>Unstructured</td>
<td>Picture</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Medical</td>
<td>- Journal</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Webpage</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Job Service</td>
<td>Researcher</td>
<td></td>
<td>Video</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Community-building</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Structured               |                   |                | Structured      | Text            |
|                         |                   |                | - Database      | RDF Text        |
|                         |                   |                | - Atlas         |                 |
| Unstructured             |                   |                | Unstructured    | Picture         |
|                         |                   |                | - Journal       |                 |
|                         |                   |                | - Webpage       |                 |
|                         |                   |                |                 | Video           |
NIF Cell Ontology

• NIF isn’t building ontologies; we import and extend as necessary
  – Establishing pipelines to ontology builders to feed classes
    • e.g., Chebi, OBI, PRO, FMA, NEMO, CogPo

• One exception: neurons and glia
  – NIF is creating an ontology for neurons
  – Defining a standard set of properties by which they can be defined
NIF Cell

**A**
- Amygdala basolateral nucleus pyramidal neuron
- Amygdala cerebral ganglion metacerebral cell

**B**
- Basalis nucleus cholinergic neuron

**C**
- Cerebellum Golgi cell
- Cerebellum Lugero cell
- Cerebellum Purkinje cell
- Cerebellum basket cell
- Cerebellum granule cell
- Cerebellum stellate cell
- Cerebellum unipolar brush cell
- Ciliary ganglion cell
- Cochlear nucleus (dorsal) cartwheel cell
- Cochlear nucleus (dorsal) granule cell
- Cochlear nucleus (ventral) bushy cell
- Cochlear nucleus (ventral) multipolar cell
- Cochlear nucleus bushy cell
- Cochlear nucleus giant cell
- Cochlear nucleus glutamatergic cell
- Cochlear nucleus multipolar D cell
- Cochlear nucleus multipolar T cell
- Cochlear nucleus octopus cell
- Cochlear nucleus vertical cell
- Collliculus inferior intrinsic cell
- Collliculus inferior principal cell

**D**
- Dentate Gyrus basket cell
- Dentate gyrus HCICAP cell
- Dentate gyrus HIPP cell

**H cont.**
- Hippocampus CA1 neurogliaform cell
- Hippocampus CA1 oriens lacunosa molecular cell
- Hippocampus CA1 stratum oriens neuron
- Hippocampus CA2 basket cell broad
- Hippocampus CA2 basket cell narrow
- Hippocampus CA2 bistratified cell broad
- Hippocampus CA2 bistratified cell narrow
- Hippocampus CA2 pyramidal neuron
- Hippocampus CA3 IS-I cell
- Hippocampus CA3 IS-II cell
- Hippocampus CA3 axo-axonic cell
- Hippocampus CA3 basket cell
- Hippocampus CA3 lacunosum molecular neuron
- Hippocampus CA3 oriens interneuron
- Hippocampus CA3 oriens lacunosum molecular neuron
- Hippocampus CA3 radiatum neuron
- Hippocampus CA3 spiny CR cell

**L**
- Locus coeruleus NA neuron

**M**
- Magnocellular neurosecretory cell

**N**
- Neocortex Cajal-Retzius cell
- Neocortex Martinotti cell
- Neocortex basket cell
- Neocortex bipolar cell
- Neocortex bipolar neuron
- Neocortex bouquet cell double
- Neocortex chandelier medium
- Neocortex chandelier cell
- Neocortex chandelier-starglow cell

**O cont.**
- Olfactory epithelium main supporting cell
- Olfactory receptor neuron
- Olfactory tubercle islets of Calleja DA cell
- Olfactory tubercle islets of Calleja GABA cell

**P**
- Paracapsular intercalated cell
- Pars reticulata principle neuron
- Parvalbumin interneuron
- Piriform cortex polymorphic cell
- Proprioception intersegmental cell
- Pyramidal basket cell
- Pyramidal cell

**R**
- Retina Cone M
- Retina amacrine cell
- Retina bipolar cell
- Retina ganglion cell
- Retina horizontal cell
- Retina photoreceptor cell
- Retina stellate cell aspiny

**S**
- Scarpa's ganglion cell (vestibular nerve)
- Short axon cell
- Solitary tract nucleus intrinsic cell
- Spinal cord intermediate horn motor neuron sympathetic
- Spinal cord ventral horn interneuron FRA
- Spinal cord ventral horn interneuron IA
- Spinal cord ventral horn interneuron IB
- Spinal cord ventral horn interneuron II
- Spinal cord ventral horn interneuron Renshaw
- Spinal cord ventral horn motor neuron
NIF cell bridge files use to create inferred hierarchies

• NIF cell to molecule
  – Neurotransmitters
  – Other molecules

• NIF cell to brain region
  – Each NIF cell name is a precomposed brain region plus cell type
    • Uniquely identifies all cells
  – Location assigned at the level of part of neuron not neuron class
    • Hippocampal neuron: has cell soma location hippocampus or any part

• NIF cell to qualities, e.g. morphology
  – Pyramidal neurons
NeuroLex, the Neuroscience Lexicon
A dynamic lexicon of 6,000 neuroscience terms supported by The Neuroscience Information Framework.

Hierarchies: Behavioral Activity • Behavioral Paradigms • Brain Regions • Cells • Diseases • Molecules • Nervous System Function • Subcellular Components • Resource Types • Qualities

Tables: Behavioral Activity • Brain Regions • Cell Types • Diseases • Molecules • Nervous System Function • Neurons • Neurons by Neurotransmitter • Organisms • Resources and Information Entities • Qualities

All Categories ABCDEFGHIJKLMNOPQRSTUVWXYZ

About • What's new • FAQs • NIFSTD ontologies • How to Contribute

Show me a Random Term!

Create a new cell category
Create a new brain region category
Create a new generic category

Subscribe to Neurolex mailing list

The NeuroLex project, supported by the Neuroscience Information Framework project, is a dynamic lexicon of neuroscience terms. Unlike an encyclopedia, a lexicon provides the meaning of a term, and not all there is to know about it.

The NeuroLex is being constructed to help improve the way that neuroscientists communicate about their data, so that information systems like the NIF can find data more easily and provide more powerful means of integrating data that occur across distributed resources. One of the big roadblocks to data integration in neuroscience is the inconsistent use of terminology in databases and other resources like the literature.

When we use the same terms to mean different things, we cannot easily ask questions that span across multiple resources. For example, if three databases have information about what genes are expressed in cortex, but they all use different definitions of cerebral cortex, then we cannot compare them easily.

As part of the NIF, we provide a simple search interface to many different sources of neuroscience knowledge. To build this, we need a consistent terminology to ensure that data provided by different sources is comparable.
NIF Cell in Action

Neuroscience Information Framework
NIF Evolution

Then  Now  Later

V1.0: NIF

NIF 1.5*
Summary

- NIF has tried to adopt a flexible, practical approach to assembling, extending and using community ontologies
  - We believe in modularity
  - We believe in starting simple and adding complexity
  - We believe in balancing practicality and rigor
  - We believe in single asserted hierarchies and multiple inferred hierarchies
- NIF is working through the International Neuroinformatics Coordinating Facility (INCF) to engage the community to help build out the Neurolex and “weaving the threads”
- The more different groups work together on establishing the basic frameworks for biomedical data, the less time and effort we need to spend on reconciling the 80% overlap between efforts and the more time we can spend delving into the deeper semantics of data integration