The NIFSTD and BIRNLex Vocabularies: Building Comprehensive Ontologies for Neuroscience

William J. Bug
Center for Research in Biological Systems, University of California, San Diego, San Diego, CA, USA

Giorgio A. Ascoli
Krasnow Institute for Advanced Study, George Mason University, Fairfax, VA, USA

Jeffrey S. Grethe
Center for Research in Biological Systems, University of California, San Diego, San Diego, CA, USA

Amarnath Gupta
Center for Research in Biological Systems, University of California, San Diego, San Diego, CA, USA

San Diego Supercomputer Center, University of California, San Diego, San Diego, CA, USA

Christine Fennema-Notestine
Department of Psychiatry, University of California, San Diego, CA, USA

Department of Radiology, University of California, San Diego, CA, USA

Angela R. Laird
Research Imaging Center, University of Texas Health Science Center, San Antonio, TX, USA

Center for Neural Informatics, Structure, and Plasticity and Molecular Neuroscience Department, San Antonio, TX, USA

Stephen D. Larson
Department of Neuroscience, University of California, San Diego, San Diego, CA 92093-0446, USA

Daniel Rubin
Center for Biomedical Informatics Research and Department of Radiology, Stanford University, Stanford, CA, USA

Gordon M. Shepherd
Center for Medical Informatics, Yale University, New Haven, CT, USA

Jessica A. Turner
Department of Psychiatry and Human Behavior, University of California, Irvine, Irvine, CA, USA

Maryann E. Martone
Department of Neuroscience, University of California, San Diego, San Diego, CA 92093-0446, USA

Center for Research in Biological Systems, University of California, San Diego, San Diego, CA, USA

Abstract

A critical component of the Neuroscience Information Framework (NIF) project is a consistent, flexible terminology for describing and retrieving neuroscience-relevant resources. Although the original NIF specification called for a loosely structured controlled vocabulary for describing neuroscience resources, as the NIF system evolved, the requirement for a formally structured ontology for neuroscience with sufficient granularity to describe and access a diverse collection of information became obvious. This requirement led to the NIF standardized (NIFSTD) ontology, a comprehensive collection of common neuroscience domain terminologies woven into an ontologically consistent, unified representation of the biomedical domains typically used to describe neuroscience data (e.g., anatomy, cell types, techniques), as well as digital resources (tools, databases) being created throughout the neuroscience community. NIFSTD builds upon a structure established by the BIRNLex, a lexicon of concepts covering clinical neuroimaging research developed by the Biomedical Informatics Research Network (BIRN) project. Each distinct domain module is represented using the Web Ontology Language (OWL). As much as has been practical, NIFSTD reuses existing community ontologies that cover the required biomedical domains, building the more specific concepts required to annotate NIF resources. By following this principle, an extensive vocabulary was assembled in a relatively short period of time for NIF information annotation, organization, and retrieval, in a form that promotes easy extension and modification. We report here on the structure of the NIFSTD, and its predecessor BIRNLex, the principles followed in its construction and provide examples of its use within NIF.

Keywords

Neuroscience Information Framework; NIF standardized; Biomedical Informatics Research Network; Web Ontology Language

---

Since the first major database of biomedical literature, Index Medicus, was assembled over a century ago by the first United States Surgeon General, John Shaw Billings (Lydenberg 1924), it has been recognized that investigators do not always use the same terms to describe objects and processes under study, even within a limited research domain. The lack of a standard vocabulary is one of the major barriers to making a broad scope of biomedical information collectively searchable. With the advent of an ever-evolving, diverse system like the World Wide Web, the need for a shared semantic framework for domains like neuroscience has become more critical if individual researchers and automated search agents are to access and utilize the most up-to-date information. Based on current, successful efforts to provide this manner of semantic disambiguation in the biomedical sciences such as the broadly scoped Unified Medical Language System (Schuyler et al. 1993), the Gene Ontology (Ashburner et al. 2000), and the Biomedical Informatics Research Network (BIRN) federated query mediation framework (Astakhov et al. 2006), it is clear one needs both to provide a means to unify the representation of concepts and to accommodate the lexical variety in use by practicing neuroscientists.

Assembling the necessary vocabularies for both annotating and searching neuroscience resources was a primary goal of the NIF project, a project designed to provide the means for describing and searching for neuroscience-relevant resources on the Web (Gardner et al. 2008a). The purpose of the NIF vocabularies was to ensure neuroscientists would be able to search the broad collection of resources indexed within the NIF integrated system (Gupta et al. 2008; Marenco et al. 2008b; Müller et al. 2008) from the vantage of the underlying concepts being described, as opposed to the idiosyncratic terms used to describe them. Linked to this goal was the need to make certain the NIF infrastructure provided a means to specify the conceptual equivalence and relatedness of entities represented across different resources—e.g., different curated databases, research articles, websites, so that users would not be left with the burden of arbitrating such relations across large and disparate search results.

The NIF employed a two-pronged approach to the creation of vocabulary resources: (1) NIF hosted a set of expert workshops to glean the preferred terms used by neuroscientists themselves to describe their data and created a hierarchy of these terms based on the way scientists would typically use them; (2) NIF created a more formal ontology for neuroscience with machine-processable semantics that could be used by the NIF federation system for refining and expanding queries and for organizing search results. The output of the terminology workshops was used to construct a loosely structured hierarchy of terms expressed in BrainML, an XML schema developed for neuroscience data (Xiao et al. 2002) and is described in more detail in the accompanying manuscript by Gardner et al. (2008b). This vocabulary, which we term NIFBasic, was designed primarily for human use in annotating and searching the NIF Registry, a database of neuroscience resources available on the web. However, for the more challenging problem of providing deeper queries of the content of individual neuroscience resources, including the neuroscience literature and databases registered to the NIF, a more formally structured and granular terminological resource was required. This requirement led to the construction of a more expansive ontology for neuroscience, termed NIFSTD (for NIF Standardized Vocabulary).

The NIFSTD builds heavily on the structure and design of the Biomedical Informatics Research Network (BIRN) Lexicon (BIRNLex), a large ontology-based vocabulary developed and maintained by the BIRN project initially for the annotation and query of brain imaging data across scales. BIRNLex was a pioneering effort for assembling practical vocabulary resources for the purposes of data federation across multiple areas of neuroscience concerned with neuroimaging. Through application of a set of best practices on the construction of ontologies emerging from the Open Biomedical Ontology (OBO) community (Smith et al. 2007), BIRNLex was able to assemble a large set of modular ontologies, each standardized to the same upper level ontology, the Basic Formal Ontology (BFO; Grenon et al. 2004), and ontology of generic relations, the OBO Relation Ontology (OBO-RO; Smith et al. 2005). At the same time, BIRNLex established a set of tools and practices for domain-specific expansion of the foundational ontologies and a means to import additional vocabularies as needed for annotation and searching of data. In this report, we describe the construction, scope and rationale of the NIFSTD and show how the basic foundation established by the BIRNLex was able to scale for the much more expansive set of vocabularies required for the broader domain of neuroscience as a whole.

Methods

The NIFSTD ontology (http://purl.org/nif/ontology/nif.owl) is expressed in the now ubiquitous Web Ontology Language (OWL— http://www.w3.org/TR/2004/REC-owl-features-20040210/) which is built on top of the Semantic Web Resource Description Framework (RDF— http://www.w3.org/TR/rdf-primer/). A wide variety of tools exist for browsing, visualizing, editing, searching, and reasoning upon formal semantic representations created using the OWL format. In particular, the Jena OWL/RDF Java library (http://jena.sourceforge.net/) was used for bulk conversion of terminologies into OWL, and the Protege-OWL ontology editor (http://protege.stanford.edu/) was extensively used for manual curation of these files. NIFSTD holds to the OWL Description Logic (OWL-DL) dialect to ensure it can support automated reasoning through use of a common OWL reasoner such as Pellet (Sirin et al. 2007).

Structure of NIFSTD

NIFSTD was built as a set of modules, each covering a distinct orthogonal domain of relevance to neuroscience. A list of these modules is provided in Table 1 Column 1. A complete list of ontologies, vocabularies and data resources referenced in this paper along with their URL's

and abbreviations is provided in Table 2. Through the use of the foundational and generic ontologies listed below, each of these modules was represented in a standardized manner. This approach not only follows the powerful modularization ontology design pattern (http://odps.sourceforge.net/), but can also be more easily extended to provide highly nuanced representations to meet the need of emerging neuroscientific research domains:

- BFO: http://www.ifomis.org/bfo/—Grenon et al. (2004)

- OBO-RO: http://obofoundry.org/ro/—Smith et al. (2005)

- The Ontology of Phenotypic Qualities (PATO— http://www.obofoundry.org/cgi-bin/detail.cgi?id=quality—Gkoutos et al. (2005)

Concepts across domains are related to one another through a set of specific object properties specified in the OBO-RO such as *located in, contains, inheres in, participates in*, etc.. These relational properties mostly exist as inverse pairs—e.g., part of and has part (see below for more detail on relations).

Each entity in NIFSTD is identified by a unique identifier and is accompanied by a variety of supporting annotations such as a definition, synonymous terms, and links to equivalent terms in other terminologies (Table 3). These properties were developed largely through the import of similar properties from the Dublin Core Metadata and the Simple Knowledge Organization System (SKOS). Some of the primary properties are:

- *Preferred label*—the default human readable term used

- *Synonym*—an alternative term in common use (including a select set of distinct synonym types such as ncbiTax GenbankCommonName, ncbiTaxScientificName, etc.)

- *Definition*—a clear, concise, human-readable definition for the entity

- *Defining citation*—contains standard citation reference for an entity definition (including definingCitationID and definingCitationURI to incorporate accession numbers from bibliographic databases or web references)

- *Curator*—person who contributed the class or annotations to the class

- *External source ID*—identifies a synonymous term in an external ontology or vocabulary (there are also many distinct external ID annotation properties for common vocabularies such as UMLS CUI (UMLS Concept Unique Identifier), MeSHID, NeuroNamesID, etc., along with a NIFID to link to the coarse-level NIFBasic categories used in the NIF Registry)

- *Curation status*—indicates the extent of curation applied to date (e.g., curated, uncurated, raw import, definition incomplete, hierarchy location temporary, pending final vetting)

- *Dates*—createDate and modifiedDate are a part of standard versioning practice

- *Obsolete properties*—isReplacedBy and hasFormer ParentClass—obsoleted classes receive these properties which also serve as a part of the versioning practice to help track the evolution of concepts

These annotation properties add lexical enrichment (e.g., synonyms) for text mining literature and database content, promote automated curation of NIFSTD and BIRNLex and match standard annotation syntax in order to leverage community curation tools. They also allow the automated tracking of the status of a term and cross referencing to the original source of the term, as well as other terminology resources. The example (Table 3) shows the neuroanatomical

region class corresponding to "amygdala" which includes external cross-references to the NeuroNames vocabulary, the UMLS CUI, and other supporting curatorial details.

Design Principles

The NIFSTD was constructed according to a set of practices established for the BIRNLex by the BIRN Ontology Task Force (http://www.nbirn.net/research/ontology/ontology_taskforce.shtm). As far as possible, these practices follow those established by the OBO Foundry project (Smith et al. 2007), and supported by the National Center for Biomedical Ontology (NCBO; http://www.bioontology.org/). These principles are designed to avoid duplication of effort and ensure that work performed under one domain has maximum utility to the broader community by conforming to standards that promote reuse.

Re-use of Available Distilled Knowledge Sources

Wherever possible, existing terminologies and ontologies were reused to cover domains that were required by the BIRN and NIF projects (Table 1). These community vocabularies were culled from a variety of sources, ranging from fully structured ontologies to loosely structured controlled vocabularies. A more detailed discussion of each of these domains and issues regarding their import is provided in the "Results" section.

Distinct, Orthogonal Concept Domains

Each of the OWL modules in NIFSTD consists of a conceptually orthogonal or distinct domain (Table 1). Orthogonality is one of the primary OBO Foundry principles critical to ensuring maximal re-usability of the ontology. The modularity helps minimize dependencies and ensure re-use by enabling users to accept only those domains they need for annotating. If an ontology contains one or more domains overlapping with an existing module, files must be mapped extensively to specify semantic equivalencies thus creating an added dependency and curatorial burden.

Single Inheritance

Each class within a domain has only a single parent class.

Unique Concept Identifiers

Each entity[1] within the ontology is assigned a unique identifier that serves as the name of the class. Human-readable labels are assigned through the *preferred label, synonym, abbreviation* and other *lexical variant* annotation properties.

Universal Resource Identifiers

In the semantic web, complete Universal Resource Identifiers (URIs) are used to maintain the identity of a given entity. In the case of a class in NIFSTD, the complete URI is the URI for the OWL module where it resides along with the specific ID (or local name in XML) for the class within that file—e.g., http://purl.org/nbirn/birnlex/ontology/BIRNLex-Anatomy.owl#birnlex_1699 is the URI for middle cerebellar peduncle.

---

[1]"Entity" refers to a unique class within NIFSTD. In this report, entity is usually used synonymously with "concept".

Definitions

OBO Foundry practice requires all concepts receive clear and specific human readable definitions structured in Aristotelian form: "A is a B which has C", e.g., "the globus pallidus is a brain region which is found within the basilar region of the vertebrate telencephalon." Without definitions, there is no way to guide the annotation choices made by curators which leads to terms being used in unanticipated ways that confound concept-based data federation. As is quite common even with well-utilized terminologies, not all terms in NIFSTD have definitions at this time. The curation_status annotation property tracks entities that are still lacking final definitions; this property is updated as definitions are added (uncurated) and finalized (curated).

Lexical Variants

NIFSTD includes the variety of accepted synonymous terms used to identify a distinct concept. These terms serve as an aid to annotators and help when using the ontology to index a large text corpus that often employ a variety of synonyms to identify a specific concept. Lexical variants also include alternative spellings and antiquated terms no longer in common use. In addition to synonymous terms, external identifiers are included from one or more external sources where equivalent concepts exist, e.g., UMLS CUIs, NCBI Taxonomy IDs, or NeuroNames IDs. This inter-terminology mapping helps to enable automatic data federation and querying against existing data sets already annotated with such IDs.

Representation of Concept Relations that Are Unambiguous, Distinct, and Constrained

NIFSTD utilizes the OBO-RO for specifying relationships between entities. Use of the OBO-RO serves both to separate the representation of different types of relations (e.g., "is a" vs. "part of") and to limit to proliferation of relation types. The former requirement is critical to enabling maximal algorithmic parseability of relations. For instance, it has been documented that the computational power of the Gene Ontology is limited by the fact that it mixes the depiction of "is a" and "part of" relations in a single hierarchical graph (Smith et al. 2003). At the same time, it is equally vital that the number of relations not be overly expansive, as each relation brings with it a computational burden — the computer code required to interpret the meaning of that relation.

Bridge Files and Object Properties

In order to maintain the orthogonal nature of the ontology domain modules, the cross-domain relations are specified in separate ontology bridge files rather than incorporated into the individual modules. In this way, the main domain files—e.g., anatomy, cell type, disease, etc.—remain independent of one another. Using these bridge files, the dependencies need only be introduced by those applications that require them, such as the NIF system, which requires a description of the anatomical location of nerve cell types. These relations currently reside in the NIF Cell module, but they are being moved to a separate files, called "bridge files" (see "Results" section for explanation), so that other applications which seek to use the underlying nerve cell domain ontology, but do not necessarily intend to import those relations, can do so. Bridge files can also choose either to import the referenced domain ontologies in their entirety or to take a more minimal approach and simply declare the classes they need to reference.

Use of Standard Expressive Formal Semantic Formats to Support Knowledge Discovery

The current use of OWL for representing the NIFSTD semantic framework provides both the ability to employ current OWL and RDF tools to assemble and edit the ontology, as well as a means to support a rich semantic mining capability to NIF in the future.

Multi-layer Versioning Policy

NIFSTD provides distinct levels of versioning for its content to make it possible for humans and computer code to choose the level of version information required for tracking changes in the ontology. These levels include:

- *Calculated ontology digest:* A process is run nightly to determine whether any of the NIFSTD OWL modules have changed in any way. If they have, a unique digest string associated with the overall content (Message Digest 5 [MD5]) is generated and logged. This mechanism provides a very efficient means of ensuring other algorithmic processes across the NIF infrastructure that must be re-run when one of the ontology elements has changed only need be executed when a change has actually occurred. This is very coarse level versioning, as no detail is logged regarding the nature of the change—only that a change has taken place.

- *Curation dates (created and modified date):* Each OWL module and each class within the module includes creation and modification dates. These curation properties provide a means for algorithms and human curators both to establish the chronology of ontology concept evolution and to determine when a change has taken place down to the level of individual classes. The nature of the change is included as a comment. Given the evolution of class-level changes in these OWL modules will be rapid and ongoing, this level of detailed curatorial annotation provides a critical means of reconstructing the overall evolution of the file. Any software task that utilizes a set of classes can automatically track when changes take place that may affect their processing.

- *File level versioning:* Each individual OWL file also receives a version number. When major changes are made to that file—e.g., large numbers of classes are added or retired, extensive relations are added to existing classes—then the version number is incremented. There are major and minor version numbers that are selectively adjusted based on the magnitude of the change.

- Retiring antiquated concept definitions, tracking former ontology graph position and replacement concepts: According to OBO Foundry policy, when a concept or class in the ontology has changed significantly or is otherwise no longer valid, then the class and its ID are "retired". In NIFSTD and BIRNLex, retiring a class means that the old class is removed from its current position in the concept taxonomy and made a child of the single class "retired_class". By retiring classes as opposed to deleting them altogether, the URI lives on and can still be used to update existing annotations created by one of the users of the NIFSTD. Retired classes also have the annotation properties former-ParentClass and isReplacedBy which again provide a means for both algorithms and humans to follow the chronology of NIFSTD and to update antiquated annotations. While the class ID is retired any time that the definition of the concept changes, the preferred label assigned to the obsolete class may be assigned to a new class.

Importing a New Ontology

The process of importing a new vocabulary into the NIFSTD varies depending upon its state (Table 1).

- If a vocabulary already uses OWL, the OBO-RO and the BFO and is orthogonal to existing modules, the import simply involves adding an owl:import statement to the main ontology file (nif.owl).

- If an existing orthogonal ontology is in OWL but does not use the same foundational ontologies as NIFSTD, then an ontology bridge file is constructed declaring the deep

level semantic equivalencies such as foundational objects and processes. Relations are drawn from the OBO-RO as needed.

- If the external terminology is organized but has not been represented in OWL, or does not use the same foundation as NIFSTD, then the terminology is adapted to OWL/RDF in the context of the NIFSTD foundational layer ontologies.

The last case requires the most effort on the part of the NIF ontology curators. This adaptation can be performed manually if there is a significant need for manual vetting as was done when incorporating the NeuroNames hierarchical vocabulary into BIRNLex. Some progress towards an automated solution has been achieved, as shown in the following example for the NIF Molecule module. As a starting point for covering molecules of import to neuroscientists, we employed a semi-automated mechanism to convert the International Union of Pharmacology (IUPHAR) voltage-gated ion channel (Clapham et al. 2003; Hofmann et al. 2003; Gutman et al. 2003; Catterall et al. 2003a, b) and G-protein coupled receptor (Foord et al. 2005) nomenclatures (see description below) into OWL. This algorithmic process begins with a terminology in spreadsheet form, where columns and rows are mapped to classes, annotation properties and relations (Fig. 1).

The conversion process utilizes the Jena OWL/RDF open source Java library to the necessary classes and associated properties in OWL (see "Methods").

Results

Current Scope of NIFSTD

As of this writing, the NIFSTD ontology framework includes the following major domains: organismal taxonomy, anatomy, nerve cell types, subcellular anatomy, nervous system function, nervous system dysfunction, phenotypic qualities, investigation provenance and molecules (Table 1; Fig. 2). For all of these domains, coverage has been focused on those entities and relations within those domains that are of most significant interest to neuroscientists. Much of the current content of NIFSTD was taken directly from BIRNLex (Fig. 2) and expanded as necessary for the NIF project. Details about the conversion process and the decisions for their inclusion are given for each of these domains in the following:

Organismal Taxonomy—This domain re-used the BIRNLex-OrganismalTaxonomy OWL module, embellishing it as necessary. BIRNLex, in turn, derived the taxonomy from several public sources including the NCBI Taxonomy and the Global Biodiversity Info Facility (GBIF: http://data.gbif.org/welcome.htm). The primary driver for including specific organismal classes in NIFSTD was whether or not the organisms were used generally in neuroscience, as determined by the NIF terminology workshops (described in Gardner et al. 2008a, b). These include organisms from across a broad swath of the phylogenetic tree from primates and rodents on through a variety of invertebrates. Starting with these general organismal categories, we proceeded down the NCBI Taxonomy tree to capture the required organisms down to the level of strains in general use by neuroscientists. We also captured the parent classes back to the Kingdom taxonomic level. The OWL classes created for these organisms included an annotation property pointing back to the record in NCBI Taxonomy. As this knowledge source explicitly states that it is not the authoritative view of organismal taxonomy, we used the GBIF online taxonomy database as the arbiter of the actual taxonomic relations. For the most part, these agree with the NCBI Taxonomy. When present in GBIF, an organism class also was assigned an annotation property to track the GBIF ID as well. Rodents, especially mice, are a special case. In this sub-domain, we relied primarily on the Mouse Genome Information mouse strain information repository (http://www.informatics.jax.org/external/festing/mouse/STRAINS.shtml), as well as the

Jackson Labs mouse catalog for more specific strain descriptions (http://jaxmice.jax.org/findmice/index.html). We also used the International Mouse Strain Resource (IMSR—http://www.informatics.jax.org/imsr/index.jsp) for strains outside of Jackson Labs. Apart from the most common strains, many of the specific strains, substrains, or transgenic mice were driven by the needs of researchers in the BIRN project, where this module was originally constructed. Rat strains are covered more coarsely. Here the driver for inclusion was the current data sets that have been concept mapped for the NIF data federation.

Macroscopic Anatomy (Anatomy)—Neuroanatomy plays a central role in categorizing and organizing neuroscience information. A decision was made early on in the assembly of the BIRNLex Anatomy OWL module to start with the venerable NeuroNames primate brain region partonomy (Bowden and Dubach 2003; Bowden et al. 2007). A considerable amount of expert effort has been invested over two decades to reconcile various neuroanatomical nomenclatures into this canonical taxonomy that included extensively specified synonymies. We chose to incorporate the partonomy consistently into a core "is a" taxonomy of brain regions by creating partonomic meta-classes. For instance, we created a "Regional part of diencephalon" class that contains subclasses such as "Thalamus", "Hypothalamus", "Superficial feature part of diencephalon", etc. This method enables one to capture the partonomy without mixing "part of" relations into the subclass relations which still represent "is a" relations. The containment can then be encoded using horizontal OWL ObjectProperty relations. For instance, we specify "Regional part of diencephalon" as being "part of" "Diencephalon". By specifying this transitive relation on the regional metaclass, it is inherited by all its subclasses, so that "Thalamus" is also a "part of" "Diencephalon" (Fig. 3). This approach is not unlike that used by the Foundational Model of Anatomy (Martin et al. 2001, 2003). We did not import the FMA directly because it was too large and expansive as a whole, extending well beyond the nervous system. We have also worked to broaden the scope to non-primate vertebrates. To date, this task has largely consisted of specifying synonymies and some additional classes drawn from the Brain Architecture Management System (BAMS; Bota et al. 2005). Cortical surface parcellation schemes were embellished using the Brodmann cytoarchitectural regions as represented in the SumsDB (Van Essen 2005).

Cell NIF—provided significant additional content for nerve cell types, both neurons and glia and other cells encountered within the nervous system. Although an existing cell type ontology was available (OBO Cell Ontology—http://www.obofoundry.org/cgi-bin/detail.cgi?id=cell), the coverage of nerve cells was minimal and insufficient for the neuroscience community. To assemble a more comprehensive list of nerve cells, a compendium of types was pooled from the SenseLab curated nerve cell physiology NeuronDB repository (Crasto et al. 2007), the Neuromorpho.org cell morphological model repository (Ascoli et al. 2007), and the Cell-Centered Database (CCDB) nerve cell types derived from the associated Subcellular Anatomy Ontology (Martone et al. 2008; Larson et al. 2007). These types were pooled within a spreadsheet that also listed cell body anatomical locations, released transmitters, circuit types, and other cellular properties. Using the semi-automated input mechanism described under methods, the contents of the spreadsheet were imported into NIFSTD where each cell type became a class and properties, e.g. transmitters, were automatically specified using appropriate OWL ObjectProperties.

Subcellular Anatomy (SAO)—To provide a formal framework for describing subcellular organelles and associated anatomical structures, we utilized the Subcellular Anatomy Ontology (SAO—Larson et al. 2007) in its entirety. SAO was created to provide a formal model of anatomy for the diverse light and electron microscopic images stored in the Cell Centered Database (CCDB; Martone et al. 2008). SAO was designed to fill the gap in describing structure at the subcellular level, e.g., parts of cells. It explicitly maps into the Gene Ontology Cellular

Component hierarchy, though in covering such a scope of structural entities for nervous system material many additional entities and relations have been required. These are being expressed on the same foundation as the BIRNLex and NIFSTD—i.e., BFO + OBO-RO. As we began to add SAO, it was clear we needed to eliminate certain duplicate entities between SAO, and the Nerve Cell types and Molecules represented in NIFSTD. We are currently working to create a "molecule lite" and "nerve cell lite" set of OWL modules that both NIFSTD and SAO can each import as a base, which will avoid creating duplicate entities, while still providing each ontology the ability to independently evolve the detailed inter-relatedness of those entities. The other major domains of structure covered by SAO, e.g., parts of neurons and glia, will also be constructed as separate modules in the future to avoid entangling relationships that will limit reuse.

Nervous system Dysfunction (Disease)—For this domain, we employed subsets of the following high-level categories from MeSH: Nervous system disease (MeSH ID D009422), Muscular disease (MeSH ID D009135) and Eye disease (MeSH ID D005128). The majority of listed diseases—284/333 classes—are listed as "Nervous system disease". We also created the category "Multisystem disease" to include neuromuscular disorders and other syndromes that present with a significant number of symptoms across a variety of body systems. MeSH is a multi-parent hierarchy and many diseases do in fact inhabit multiple nodes within the overall MeSH terminology graph. In holding with OBO Foundry principles we sought to include only a single inheritance graph. When choosing the parent to include we biased our choice toward categorizations that implied a grouping by presenting symptoms, as opposed to other criteria such as a proposed etiology like genetic disorders or autoimmune disease. These other facets of the various diseases will ultimately be included using horizontal, OWL ObjectProperty relations, as opposed to using them as "is a" relations thus leading to multiple inheritance. MeSH also served as a source for definitions and synonyms. For certain MeSH diseases, an ENTRY TERM may imply a subclass. When this could be corroborated through other sources, those terms were created as their own distinct classes. The diseases were also enhanced with supporting definitions and links to the NINDS online disorder index (http://www.ninds.nih.gov/disorders/disorder_index.htm), when a given disease was listed in that index.

Nervous System Function (Sensory|Cognitive|Behavior)—This module resulted from a collaboration between members of the BIRN Ontology Task Force and curators of the Brainmap.org fMRI repository (Fox et al. 2005). Terms used by Brainmap.org to describe sensory, cognitive, and behavioral paradigms employed to collect dynamic MRI images during the execution of specific brain functions were re-organized in a manner that promoted incorporating these concepts into the BIRNLex/NIFSTD BFO + OBO-RO foundation.

Investigation and Resources—Once again, the BIRNLex ontology initiated the investigative details primarily scoped to the neuroimaging domain. This was built with the understanding this sort of experimental provenance will ultimately be covered by the Ontology of Biomedical Investigation (OBI—http://obi.sourceforge.net/). We built on this base to add physiological techniques and some coarse-level descriptions of molecular experiments, again expecting more detail in that domain will come soon from OBI. As this module will ultimately derive from OBI, it also contains representational descriptors for published resources. In the next version of NIFSTD, these descriptors will ultimately reside in a module of their own that will collectively meet the resource descriptor needs of the NIF, OBI, NITRC (http://nitrc.org), and the Resourceome projects (Dinov et al. 2008).

Phenotypic Qualities—We reused the OBO Foundry Ontology of Phenotypic Qualities (PATO—http://www.obofoundry.org/cgi-bin/detail.cgi?id=quality). Phenotype is being used

in its most broad sense meaning any observable quality. Using this well-founded description of biomaterial and biological process qualities enables ultimately making diverse yet meaningful comparisons across a broad scope of biological phenomena and a broad swath of organism types. For instance, one will be able to collectively query for processes involving cell degeneration regardless of whether the cells are from an invertebrate such as Aplysia or human tissue. PATO is built on the same foundation we are using in developing NIFSTD/ BIRNLex—i.e., BFO + OBO-RO and is informed by years of experience by PATO curators who have also been seminal contributors to and users of the Gene Ontology (Harris et al. 2004).

Molecule—A Jena-based algorithm (see "Methods") was developed to adapt the IUPHAR terminologies. As an example of how this algorithm seeks to fully utilize the information encoded in the IUPHAR nomenclature, we provide a view below of the OWL representation for one of the voltage-gated Sodium Channels (type 1). The following fields have been specified in the IUPHAR nomenclature repository for each of the channel types derived from human genes: channel name, parent class, GENBANK transcript accession number, preferred gene name, chromosomal map location, other names, auxiliary subunits, minimum pore amino acid sequence analyzed, and channel aa sequence. We first manually defined a set of parent classes that are specified as being subclasses of an appropriate macromolecular class such as "voltage-gated sodium channel" (Fig. 4). Given the channel nomenclature, it is then possible to parse out a name for an appropriate intermediate channel class—e.g., "Sodium channel type 1". The "channel name" was then used to create the specific channel class with the "other names" added as synonym annotations. The gene name and accession number were used to create a class for the related gene as a type of Sequence Ontology (SO) gene that includes the map position as an annotation property. The map position string was parsed to determine which chromosome number was indicated, which in turn was used to create a class for that chromosome as a type of "SO:nuclear chromosome". The two types of amino acid sequence are added as types of "SO:mature protein region". Finally, if an additional subunit was listed, it was added as a part of the channel. The result of this process is depicted in the figure below for the Nav1.4 Sodium channel both in graph form (Fig. 4). In the NIFSTD file, intervening channel classes were included with the intention of capturing the sense of whether the channel is composed of one or more main subunits that are either the same (homomer) or different (heteromer).

The IUPHAR G-protein coupled receptor nomenclature files included a slightly different complement of information. For one, in addition to gene accession numbers, they provided both transcript and protein accession numbers for human, mouse, and rat. They also listed a nomenclature code for each receptor along with a list of known ligands. The graph below provides an example of how this information is translated into OWL using the alpha 2A type adrenoceptor as an example (Fig. 5). A representation like this was created for all three species. Note that in both the case of voltage-gated ion channels and the G-protein coupled receptors we restricted ourselves to using relations from or soon to be added to the OBO-RO, so as to promote maximal interoperability with other ontologies (Smith et al. 2007).

Creating Hierarchies Using OWL

As these neuroscience domains are represented using OWL to capture the complex inter-relatedness, one can use community-based OWL and RDF tools to algorithmically search or infer complex conceptually specified sets of NIFSTD annotated records. For instance, some of the neuroanatomical containment partonomies are asserted in the current NIFSTD anatomy module through the "is a" relationships specified within the OWL module (Fig. 3). Others can be inferred with common OWL reasoners or specified using RDF based queries (i.e., SPARQL). In this way, data annotated using standardized and globally unique IDs that are

represented in RDF can be queried based on the partonomy (e.g., "return all hits on thalamus or any region contained within the thalamus", "return all hits on basal ganglia excluding those on pallidum", etc.) using RDF and/or OWL tools. Not all such partonomies have yet been specified in the neuroanatomical module of NIFSTD, but over time, additional ObjectProperties are being added to fully encode all such relations. This enrichment of interrelations will also be applied across domains in the ontology (Fig. 6). Already there are relations stipulating the brain regions in which specific nerve cell types are found and the circuit types they participate in. Some nerve cells also have transmitters specified.

Use of the NIFSTD

The prime requirement for the NIFSTD vocabulary was to provide a uniform conceptual framework and associated set of terms to query the variety of resources made searchable through the NIF—i.e., neuroscience resource descriptions, web site content and associated data repositories linked to those same resources, and a related neuroscientific literature corpus. The NIFSTD has been employed to provide semantic mark up of source databases registered to the NIF data federation (See Gupta et al. 2008 and Marenco et al. 2008b) for details about the annotation process). In this process, concepts contained within data sources, e.g., databases, are mapped to the unique identifiers within NIFSTD. The NIFSTD also provides the semantic underpinnings of the integrated NIF system, allowing conceptbased queries across multiple neuroscience resources without explicit mapping of data sources (Gupta et al. 2008). An example of how the NIFSTD vocabulary enhances search of the NIF is shown in Fig. 7, illustrating the advanced search interface of the NIF Beta Release (available through http://nif.nih.gov). In this case, a user enters "neurodegenerative" into the advanced search option of the NIF. The NIF first returns terms in NIFSTD matching the string. After selecting a NIFSTD term, the user can expand the term, returning the direct parent and children of the term. In the example shown, the children are neurodegenerative diseases. The user can further expand the search to include synonyms. Synonyms are joined by an "OR" condition, i.e., the NIF searches for "Huntington's disease OR Huntington's chorea OR HD". Once the query is composed, the NIF uses these terms to search all the NIF resources simultaneously. In the example shown, results from the NIF Registry, a curated database of neuroscience relevant resources, are shown. Other reports in this issue describe the NIF search capabilities in great detail (Müller et al. 2008; Gupta et al. 2008).

Viewing the NIFSTD Vocabularies

The NIFSTD and BIRNLex vocabularies are available as. owl files (http://purl.org/nif/ontology/nif.owl and http://purl.org/nbirn/birnlex/ontology/birnlex.owl, respectively), which may be viewed using Protégé or similar ontology tools. However, these tools generally require a fair amount of expertise to use. To create more human friendly viewing environments, both NIFSTD and BIRNLex have been uploaded into the NCBO BioPortal (http://www.bioontology.org/ncbo/faces/index.xhtml). In the BioPortal, a user can search for specific terms, browse the overall ontology concept tree, select specific concepts to display in the graph viewer, and view associated concept properties. An upcoming release of the Bioportal will also support community feedback on ontology concepts through a discussion forum for each concept.

Discussion

The NIFSTD was constructed to provide a significantly fine-grained and formal ontology for neuroscience that supports machine-based access and reasoning through the NIF federated information system. Through the application of emerging best practices within the ontology community, NIFSTD was constructed in a rather short time frame in a way that will promote its further evolution and reuse in other applications.

Concept-Based Searching

The NIF system allows for "concept-based" queries that utilized the NIF vocabularies. By concept-based, we mean a search that probes data sets based on shared meaning of content, as opposed to matching of terms they have in common. Thus, nucleus as part of cell and nucleus as part of brain each map to a unique identifier such that the NIF search should easily distinguish between the two. Terms such as "Parkinson's disease" and "Parkinson's syndrome" both link appropriately to the same concept, thus searches using either synonymous term leads to the same set of results.

This initial phase of the NIF project has clearly demonstrated concept-based searching can be implemented across a diverse set of neuroscience-related resources (Gupta et al. 2008). NIFSTD-based concept searches were most effective when implemented against data repositories that have been registered with the NIF query mediator and concept-mapped down to the individual tuple level (Marenco et al. 2008b). However, as the amount of data explicitly mapped to NIFSTD is small, we have also utilized the semantic relations within NIFSTD to provide more powerful search even in the absence of explicit mappings. As described in the results, NIF expands searches for a concept such as Parkinson's disease to include parent and children terms, and lexical variants like synonyms. In the coming phase of the NIF project, NIFSTD will also explore how to utilize information within the NIFSTD to disambiguate homonymous concepts such as "nucleus as part of cell'" and "nucleus as part of brain". We will also explore how to utilize within the NIF infrastructure more extended entity relations such as partonomies and relations between structure and function.

NIFSTD and OBO Best Practices

The NIFSTD was constructed using principles and practices espoused by the OBO Foundry project and put into practice by the BIRN project. In particular, the structure of NIFSTD and much of the content was copied directly from BIRNLex. BIRNLex was created to provide a lexicon of concepts utilized by neuroscientists to describe neuroimaging experiments across scales. It was designed to be used by the BIRN mediator, a database federation engine that provides cross query of multiple databases through concepts mapped to a shared ontology (Astakhov et al. 2006). The BIRN project initially planned to use the comprehensive UMLS vocabulary (Schuyler et al. 1993), which, like BIRNLex and NIFSTD, imports existing terminology resources and provides cross mappings. However, BIRN participants found the UMLS difficult to use because of the lack of human readable definitions, the duplication of concepts and because of the sparse semantic network used to connect them. The BIRNLex began as an effort to provide a human-curated subset of concepts within UMLS, and both BIRNLex and NIFSTD maintain a link to the appropriate UMLS CUI. However, in order to serve the needs of the BIRN and NIFSTD projects, we also incorporated additional vocabularies that are not included in UMLS.

The NIFSTD and BIRNLex provide good examples of how application of OBO Foundry principles can facilitate reuse of ontologies across applications. BIRNLex went through several iterations as these practices were developed and employed, finally settling on a modular structure with single inheritance class hierarchies, each covering a distinct domain. Although in the current version of NIFSTD described in this report (v 0.5), the modular structure is not completely implemented, reflecting the many iterations of development of BIRNLex before this principle was formalized, we expect that with the release of NIFSTD v1.0, the modularity principle will be realized.

By using existing terminologies and tools, we were able to create a very large vocabulary resource for the NIF project in a matter of months. In re-using the OBO-RO relations, we ensure other ontologies designed to use those relations will be commensurate with relations as defined

in NIFSTD. By having unique IDs for each concept, we provide other users of the NIFSTD ontologies an unambiguous means of referencing concepts. The versioning policy in place provides some support for evolving existing annotations, as necessary changes are made to the concepts in NIFSTD. Human readable definitions for each concept help to promote clarity in the use of these concepts for annotation. The human readable definitions are not meant to be authoritative, but to provide a clearly defined standard that may be applicable or not by an individual researcher.

By following the OBO Foundry principles, we have tried to ensure that NIF modules can be utilized by other communities with minimal effort. The rapid construction of NIFSTD was possible through the direct import of BIRNLex modules, including the extensive set of annotation properties allowing tracking of class provenance, providing proof of principle that following these practices promotes flexible and efficient use of ontologies. Because both BIRNLex and NIFSTD use the same upper level ontology, they can exchange modules easily, e.g., the NIFSTD cell type module developed through NIF. Although the two ontologies largely cover the same domain, they will be maintained as separate entities because in the future we anticipate that BIRN will extend beyond the neuroscience domain. As BIRN seeks to cover domains outside of neuroscience, the required ontology coverage will be built out in separate OWL files that will not be linked into the NIFSTD semantic framework, allowing continued interaction between on-going NIFSTD and BIRNLex development.

Expressivity and Advantages of OWL

Using OWL has brought with it considerable expressivity over simply using a general markup language such as various XML schemas. XML-based schemas are primarily used to define hierarchical data models. The core XML semantics do not directly support a means for creating is_a subsumptive hierarchies, a relation that is critical when constructing descriptions of biomedical reality. The over-arching need for representing is_a hierarchies when describing complex, real-world data sets such as those found in the life sciences (e.g., mouse is_a rodent is_a mammal, etc.) is one of the primary reasons the Semantic Web RDF formalism provides such parent-child is_a relation in its core semantics. RDF can be used both to create subsumptive hierarchies of entities and the properties which are used to relate one entity to another (e.g., proper_part_of is_a type of part_of relation). The semantics specify the SubClassOf and SubPropertyOf relations are transitive, thus enabling child entities to inherit all the properties of the chain of parent classes. OWL builds on RDF, adding an additional layer to its semantics that helps support more expressive set operations such as those found typically in logical programming languages—e.g., defining equivalent and disjoint sets, defining new sets based on union or intersection of existing sets, etc. OWL also adds enriched property relations (ObjectProperties) which provide the basis for inferring when a given entity belongs to a defined set and include the ability to declare standard set property qualities such as transitivity, inverse, reflexivity, etc.. Because of this enriched but generalized expressivity, OWL and RDF have become much more prevalent in bioinformatic studies in the last several years, and the variety of tools and programming libraries built around these languages provide a very significant foundation on which to construct such applications. In the case of constructing terminologies to describe complex neuroscience data sets, there is no doubt the additional expressive semantics RDF and OWL provide are indispensable. Were one to construct this *de novo* using XML, one would merely be duplicating what RDF and OWL have already done, quite literally so, as both formalisms include an XML serialized format.

As mentioned above, the core OWL semantics include relations for asserting hierarchies, as shown in Fig. 3, and also to infer hierarchies based the set of properties assigned to a class. This latter capacity is critical when dealing with the inherent complexity of biological objects within ontologies. Entities such as nerve cells or brain regions are complex, and can exist within

many hierarchies depending upon the point of view. Trying to construct all of these hierarchies as separate ontologies generally is non-productive for groups trying to create re-usable ontologies, as it leads to an open ended problem. For this reason, the BIRN OTF found that sticking to the OBO Foundry principle of single inheritance trees was invaluable in moving forward with the BIRNLex. Application of this principle leads to fairly flat and prosaic hierarchies that on their own were of minimal expressivity. However, through the judicious assignations of properties to a class, a large number of asserted hierarchies can be automatically generated as required. A good example of this utility is the NIF Cell Type module. The list of nerve cells contained within NIFSTD is fairly flat; nerve cells are listed as either neuron or glia. The core nerve cell IS_A hierarchy does not specify where a cell is found, what neurotransmitter it uses or its physiological properties. These features are specified through a set of properties assigned to the nerve cell class. Using the asserted classification capabilities of OWL, we can generate inferred hierarchies for each of these properties (Larson et al. 2007).

Coverage of NIFSTD

There are two primary areas where NIFSTD coverage is currently minimal or lacking: molecules and cross-module relations. Molecules have been partially covered with a focus on molecules mediating cellular excitability, e. g., ion channels and receptors. As described below, a different approach will be needed for a more inclusive scope and a more complete representation at all required levels of granularity. This solution will also make more extensive use of the basic concepts in both the OBO Sequence Ontology (Eilbeck et al. 2005) and the OBO Protein Ontology (Natale et al. 2007). Cross domain relations have been covered only minimally so far. For instance, there are some anatomical locations designated for many of the nerve cell types, though not all have been given a designation to date. Clearly, only the most specific types of neuron can be assigned a specific location and even then, certain types of cells have a wide range of occurrence throughout the brain. This information will be augmented over time. Enhancements to the nerve cell representation will also include focus on incorporating the Petilla Convention nerve cell qualities (Ascoli et al. 2008). Another representational detail that requires specifying relations across domains is cross-species anatomy. This has yet to be directly addressed in the BIRNLex-Anatomy module that contains the CNS and PNS regional entities used by NIFSTD. Several projects now underway seek to address this issue (Baldock and Burger 2005; Zhang and Bodenreider 2005; Mabee et al. 2007) and NIF will take advantage of the results when they are made available.

The Challenge of Molecules

Our wealth of molecular knowledge has been amplified by whole genome sequencing and continues to accelerate in post-genome era expression studies leading to a profusion of such information. The mouse chromosome alone has over 20,000 expressed genes each of which nearly all have a several alternative transcripts, not to mention the combinatorial explosion that genes subject to somatic recombination gives rise to. A system designed to search across neuroscience information must provide a means to access resource data based on the molecular concepts they reference. If all of the molecular entities are to be tracked and unambiguously identified when they occur in individual records within an available resource, the tools used to build and apply the semantic framework must scale to many millions of concepts.

The primary focus in NIFSTD to date has been on molecules mediating cellular excitability (ion channels and receptors). We chose to test these tools and techniques by starting first with a restricted set of molecules for which a highly curated, normalized terminology already existed —the IUPHAR Nomenclature Committee's compendium terminology for voltage-gated ion channels and G-protein coupled receptors. Collectively these represented approximately ~750 genes and expressed peptide isoform sets. Voltage-gated ion channels are given for human

only whereas G-protein coupled receptors were characterized for three separate species—human, mouse, and rat. Once one accounts for the genes, associated mRNA transcripts, and even a simplified view of the transcribed peptides macromolecular receptors across all three organisms, the number of concepts involved grows to approximately 3,000-4,000 concepts which are richly inter-related with each other and with a variety of molecular ligands.

Creating algorithmic tools to construct this according to the principles above was difficult but not impossible, and these are the primary molecules currently present and searchable via the NIFSTD ontology. Unfortunately, the available tools and current version of OWL are not capable of scaling this representation up to cover all of the 100,000s of expressed genes across dozens of organisms. An alternative means to approaching the long-term goal of providing a rich and parseable representation of this intricate molecular detail is to start with a the more tractable objective of simply being able to unambiguously identify equivalent concepts for genes, transcripts, proteins, and drugs when these are encountered across the various resources. Though there have been preliminary efforts to do this in RDF by a variety of groups (NeuroCommons, Entrez Neuron (Lam et al. 2006), National Library of Medicine's creation of URIs (Sahoo et al. 2008)), the most comprehensive and stable source currently hosting this information is the collection of NCBI-associated data repositories such as Entrez Gene, Entrez Protein, RefSeq, Homologene, OMIM, etc.. Though they are not designed to support richly expressed views of these molecules, they are very much intended to provide a means to uniquely identify specific molecules, the names and symbols used to reference them, and a rudimentary sense of how various molecules relate to one another—e.g., genes on a chromosome, or peptides derived from particular genes and their homologs across species. NCBI also provides algorithmic access to this information in the form of query-based web services. The NIF engineers are in the process of adapting the current NIF infrastructure so that it will be able to make use of NCBI molecular identification capabilities when resolving NIF user queries.

Conclusion

The NIF project to date has demonstrated the creation of a broadly-scoped ontology (NIFSTD) to support concept-based searches against a wide range of neuroscience resources. When users/ scientists related to a resource invest the time to map appropriate semantic descriptors into the NIFSTD, very satisfactory concept-based searching is possible against that resource. General coverage of the bulk of the required domains—organisms, anatomy, disease, cell type, technique, resource type—was achieved to the extent required to support concept-driven queries of concept-mapped repositories, and to support term based searching of indexed web pages and available literature (Gupta et al. 2008; Müller et al. 2008). Although the initial coverage of molecules of excitability proved promising, an alternative solution will be required to scale molecule coverage to the 1,000,000 concept level as required. This, along with more detailed cross-domain relations, will be some of the primary work done on NIFSTD in the coming phase of NIF development.

Information Sharing Statement

The NIFSTD and BIRNLex ontologies are available at http://purl.org/nif/ontology/nif.owl and http://purl.org/nbirn/birnlex/ontology/birnlex.owl respectively. The NIF is offered under BSD and MIT compatible OS licenses (http://opensource.org/licenses).

Acknowledgments

References

Ascoli GA, Donohue DE, Halavi M. NeuroMorpho.Org: A central resource for neuronal morphologies. The Journal of Neuroscience 2007;27:9247–9251. [PubMed: 17728438]doi:10.1523/JNEUROSCI. 2055-07.2007

Ascoli GA, Alonso-Nanclares L, Anderson SA, Barrionuevo G, Benavides-Piccione R, Burkhalter A, et al. Petilla terminology: Nomenclature of features of GABAergic interneurons of the cerebral cortex. Nature Reviews Neuroscience 2008;9:557–568.doi:10.1038/nrn2402

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. Nature Genetics 2000;25:25–29. [PubMed: 10802651]doi:10.1038/75556

Astakhov V, Gupta A, Grethe JS, Ross E, Little D, Yilmaz A, Ellisman M, et al. Semantically based data integration environment for biomedical research. Proc. of 19th IEEE Symp. on Comp. Based Med. Sys (CBMS'06). 2006

Baldock R, Burger A. Anatomical ontologies: Names and places in biology. Genome Biology 2005;6Epub 2005 Mar 15

Bota M, Dong HW, Swanson LW. Brain architecture management system. Neuroinformatics 2005;3:15–48. [PubMed: 15897615]doi:10.1385/NI:3:1:015

Bowden DM, Dubach MF. NeuroNames 2002. Neuroinformatics 2003;1:43–59. [PubMed: 15055392] doi:10.1385/NI:1:1:043

Bowden DM, Dubach M, Park J. Creating neuroscience ontologies. Methods in Molecular Biology (Clifton, N.J.) 2007;401:67–87.

Catterall WA, Goldin AL, Waxman SG, International Union of Pharmacology. International Union of Pharmacology. XXXIX. Compendium of voltage-gated ion channels: Sodium channels. Pharmacological Reviews 2003a;55:575–578. [PubMed: 14657413]doi:10.1124/pr.55.4.7

Catterall WA, Striessnig J, Snutch TP, Perez-Reyes E, International Union of Pharmacology. International Union of Pharmacology. XL. Compendium of voltage-gated ion channels: Calcium channels. Pharmacological Reviews 2003b;55:579–581. [PubMed: 14657414]doi:10.1124/pr.55.4.8

Clapham DE, Montell C, Schultz G, Julius D, International Union of Pharmacology. International Union of Pharmacology. XLIII. Compendium of voltage-gated ion channels: Transient receptor potential channels. Pharmacological Reviews 2003;55:591–596. [PubMed: 14657417]doi:10.1124/pr.55.4.6

Crasto CJ, Marenco LN, Liu N, Morse TM, Cheung KH, Lai PC, et al. SenseLab: New developments in disseminating neuroscience information. Briefings in Bioinformatics 2007;8:150–162. [PubMed: 17510162]doi:10.1093/bib/bbm018

Dinov ID, Rubin D, Lorensen W, Dugan J, Ma J, Murphy S, et al. iTools: A framework for classification, categorization and integration of computational biology resources. PLoS ONE 2008;3:e2265. [PubMed: 18509477]doi:10.1371/journal.pone.0002265

Eilbeck K, Lewis SE, Mungall CJ, Yandell M, Stein L, Durbin R, et al. The sequence ontology: A tool for the unification of genome annotations. Genome Biology 2005;6:R44. [PubMed: 15892872]Epub 2005 Apr 29doi:10.1186/gb-2005-6-5-r44

Foord SM, Bonner TI, Neubig RR, Rosser EM, Pin JP, Davenport AP, et al. International Union of Pharmacology. XLVI. G protein-coupled receptor list. Pharmacological Reviews 2005;57:279–288. [PubMed: 15914470]doi:10.1124/pr.57.2.5

Fox PT, Laird AR, Fox SP, Fox PM, Uecker AM, Crank M, et al. BrainMap taxonomy of experimental design: Description and evaluation. Human Brain Mapping 2005;25:185–198. [PubMed: 15846810] doi:10.1002/hbm.20141

Gardner D, Akil H, Ascoli GA, Bowden DM, Bug W, Donohue DE, et al. The Neuroscience Information Framework: A data and knowledge environment for neuroscience. Neuroinformatics. 2008athis issue

Gardner D, Goldberg DH, Grafstein B, Robert A, Gardner EP. Terminology for neuroscience data discovery: Multi-tree syntax and investigator-derived semantics. Neuroinformatics. 2008bthis issue

Gkoutos GV, Green EC, Mallon AM, Hancock JM, Davidson D. Using ontologies to describe mouse phenotypes. Genome Biology 2005;6:R8. [PubMed: 15642100]Epub 2004 Dec 20 doi:10.1186/gb-2004-6-1-r8

Grenon P, Smith B, Goldberg L. Biodynamic ontology: Applying BFO in the biomedical domain. Studies in Health Technology and Informatics 2004;102:20–38. [PubMed: 15853262]

Gupta A, Bug W, Marenco L, Qian X, Condit C, Rangarajan A, et al. Federated access to heterogeneous information resources in the Neuroscience Information Framework (NIF). Neuroinformatics. 2008this issue

Gutman GA, Chandy KG, Adelman JP, Aiyar J, Bayliss DA, Clapham DE, et al. International Union of Pharmacology. XLI. Compendium of voltage-gated ion channels: Potassium channels. Pharmacological Reviews 2003;55:583–586. [PubMed: 14657415]doi:10.1124/pr.55.4.9

Harris MA, Clark J, Ireland A, Lomax J, Ashburner M, Foulger R, et al. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Research 2004;32:D258–D261. [PubMed: 14681407]doi:10.1093/nar/gkh066

Hofmann F, Biel M, Kaupp UB, International Union of Pharmacology. International Union of Pharmacology. XLII. Compendium of voltage-gated ion channels: Cyclic nucleotide-modulated channels. Pharmacological Reviews 2003;55:587–589. [PubMed: 14657416]doi:10.1124/pr.55.4.10

Lam HY, Marenco L, Shepherd GM, Miller PL, Cheung KH. Using web ontology language to integrate heterogeneous databases in the neurosciences. AMIA… Annual Symposium Proceedings/AMIA Symposium. AMIA Symposium 2006:464–468. [PubMed: 17238384]2006

Larson SD, Fong LL, Gupta A, Condit C, Bug WJ, Martone ME. A formal ontology of subcellular neuroanatomy. Frontiers in Neuroinformatics 2007;1:3. [PubMed: 18974798]doi:10.3389/neuro.11/003.2007

Lydenberg, HL. John Shaw Billings: Creator of the National Medical Library and its Catalogue, First Director of the New York Public Library. American Library Association. The Merrymount Press; Boston: 1924.

Mabee PM, Arratia G, Coburn M, Haendel M, Hilton EJ, Lundberg JG, et al. Connecting evolutionary morphology to genomics using ontologies: A case study from Cypriniformes including zebra fish. Journal of Experimental Zoology. Part B Molecular and Developmental Evolution 2007;308:655–668. [PubMed: 17599725]doi:10.1002/jez.b.21181

Marenco L, Li Y, Martone ME, Sternberg PW, Shepherd GM, Miller PL. Issues in the design of a pilot concept-based query interface for the Neuroinformatics Information Framework. Neuroinformatics. 2008bthis issue

Martin RF, Mejino JL Jr, Bowden DM, Brinkley JF 3rd, Rosse C. Foundational model of neuroanatomy: Implications for the Human Brain Project. Proc AMIA Symp 2001:438–442. [PubMed: 11825226]2001

Martin RF, Rickard K, Mejino JL Jr, Agoncillo AV, Brinkley JF, Rosse C, et al. The evolving neuroanatomical component of the Foundational Model of Anatomy. AMIA… Annual Symposium Proceedings/AMIA Symposium. AMIA Symposium 2003:927. [PubMed: 14728433]2003

Martone ME, Tran J, Wong WW, Sargis J, Fong L, Larson S, et al. The cell centered database project: An update on building community resources for managing and sharing 3D imaging data. Journal of Structural Biology 2008;161:220–231. [PubMed: 18054501]doi:10.1016/j.jsb.2007.10.003

Müller HM, Rangarajan A, Teal TK, Sternberg PW. Textpresso for neuroscience: Searching the full text of thousands of neuroscience research papers. Neuroinformatics. 2008this issue

Natale DA, Arighi CN, Barker WC, Blake J, Chang TC, Hu Z, et al. Framework for a protein ontology. BMC Bioinformatics 2007;8(Suppl 9):S1. [PubMed: 18047702]doi:10.1186/1471-2105-8-S9-S1

Sahoo SS, Bodenreider O, Rutter JL, Skinner KJ, Sheth AP. An ontology-driven semantic mashup of gene and biological pathway information: Application to the domain of nicotine dependence. Journal of Biomedical Informatics 2008;41:752–765. [PubMed: 18395495]Feb 29. Epub ahead of print

Schuyler PL, Hole WT, Tuttle MS, Sherertz DD. The UMLS Metathesaurus: Representing different views of biomedical concepts. Bulletin of the Medical Library Association 1993;81:217–222. [PubMed: 8472007]

Sidhu AS, Dillon TS, Chang E. Deployment of protein ontology framework. International Journal of Applied Bioengineering 2007;1:34–41.

Sirin E, Parsia B, Grau BC, Kalyanpur A, Katz Y. Pellet: A practical OWL-DL reasoner. Web Semantics: Science, services and agents on the World Wide Web 2007;5:51–53.doi:10.1016/j.websem. 2007.03.004

Smith B, Williams J, Schulze-Kremer S. The ontology of the gene ontology. AMIA… Annual Symposium Proceedings/AMIA Symposium. AMIA Symposium 2003:609–613. [PubMed: 14728245]2003

Smith B, Ceusters W, Klagges B, Köhler J, Kumar A, Lomax J, et al. Relations in biomedical ontologies. Genome Biology 2005;6:R46. [PubMed: 15892874]Epub 2005 Apr 28doi:10.1186/gb-2005-6-5-r46

Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, et al. The OBO Foundry: Coordinated evolution of ontologies to support biomedical data integration. Nature Biotechnology 2007;25:1251–1255.doi:10.1038/nbt1346

Van Essen DC. A population-average, landmark- and surface-based (PALS) atlas of human cerebral cortex. NeuroImage 2005;28:635–662. [PubMed: 16172003]doi:10.1016/j.neuroimage.2005.06.058

Xiao XP, Abato M, Knuth KH, Gardner D. An abstract semantic metalanguage for developing and interfacing data description languages. Biophysical Journal 2002;82(1 part 2):823. [PubMed: 11806924]

Zhang S, Bodenreider O. Alignment of multiple ontologies of anatomy: Deriving indirect mappings from direct mappings to a reference. AMIA… Annual Symposium Proceedings/AMIA Symposium. AMIA Symposium 2005:864–868. [PubMed: 16779163]2005

Fig. 1.
Terminology spreadsheet for automated import of the IUPHAR ion channel database. This direct import was used to create OWL classes, annotations and related properties, under NIFSTD Molecule. Further description of this process is described in the "Results" section
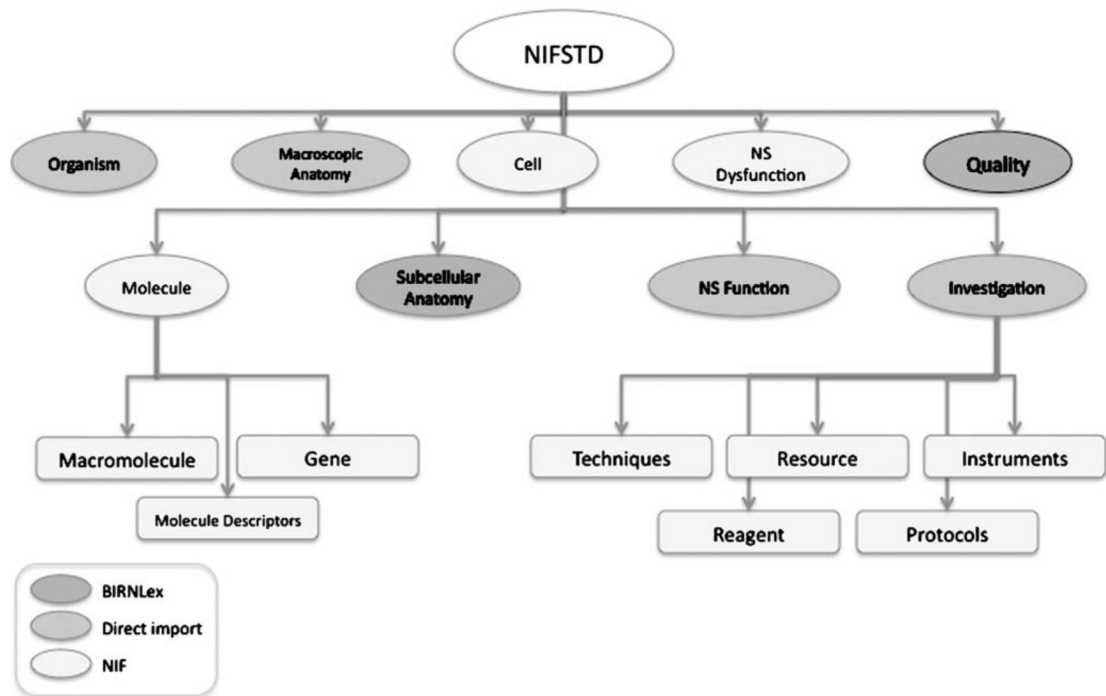
Fig. 2.
The semantic domains covered in the NIFSTD v0.5 OWL ontology. Each of the domains specified within the ovals are covered by a separate OWL module (see Table 1). The umbrella file http://purl.org/nif/ontology/nif.owl imports each of these modules when opened in Protégé. These domains are covered either through import of the corresponding BIRNLex module, a module constructed by NIF or a direct import of an existing ontology (see key for color code). Each of the modules, in turn, may cover multiple subdomains, some of which are shown in the *rectangular boxes. NS* = nervous system
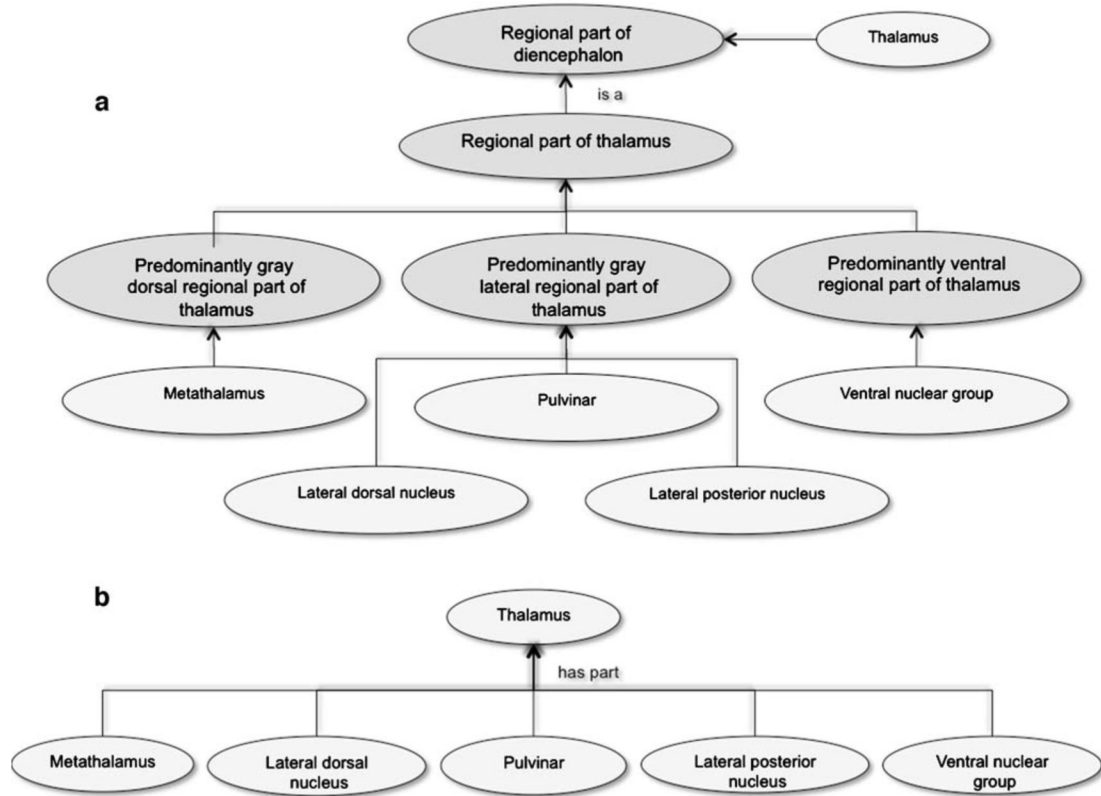
Fig. 3.
View of some mammalian thalamic brain regions in NIFSTD. a Core "is a" hierarchy for
"Regional part of diencephalon"; **b** Partonomy of diencephalon computed using OWL
ObjectProperties and restrictions that relate the regional part of thalamus to the thalamus. Only
a portion of the classes covering thalamic entities is shown here
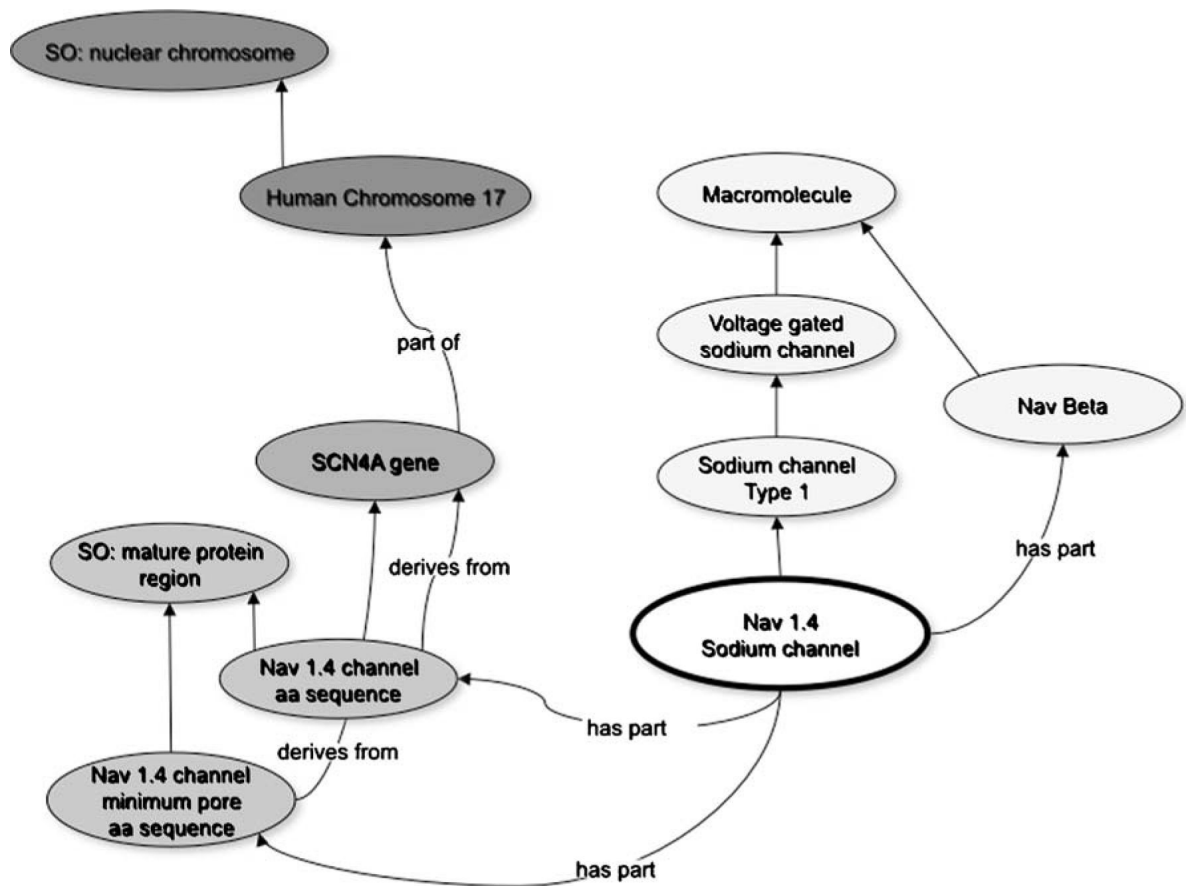
Fig. 4.
Graph view of Nav Sodium Channel representation in NIFSTD. NIF Molecule defines the relationships among the macromolecule Nav 1.4 sodium channel (*bold oval*), its auxiliary subunit (*Nav Beta*), the channel and pore regions (aa sequence), the gene encoding these sequences (SCNA4 gene) and its corresponding chromosome location. The different *shadings* indicate that the concepts come from different hierarchies within the NIF molecule module. *Unlabeled arrows* represent "is a" relationship, e.g., Nav1.4 Sodium channel is a Type 1 Sodium channel
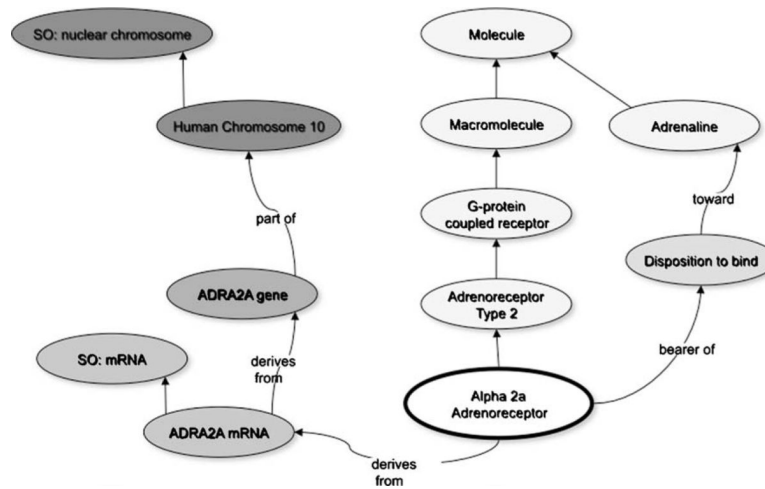
Fig. 5.
Graph view of Alpha 2A Adrenoceptor G-protein coupled receptor (*bold oval*) representation in NIFSTD Molecule. Through the classes and relationships within NIFSTD Molecule, the macromolecule is related to its ligand, mRNA and ultimately the chromosome containing the DNA sequence through additional relationships. The different shadings indicate that the concepts come from different hierarchies within the NIF molecule module. *Unlabeled arrows* represent "is a" relationship, e.g., e.g., Alpha 2a Adrenoreceptor is a Type 2 Adrenoreceptor
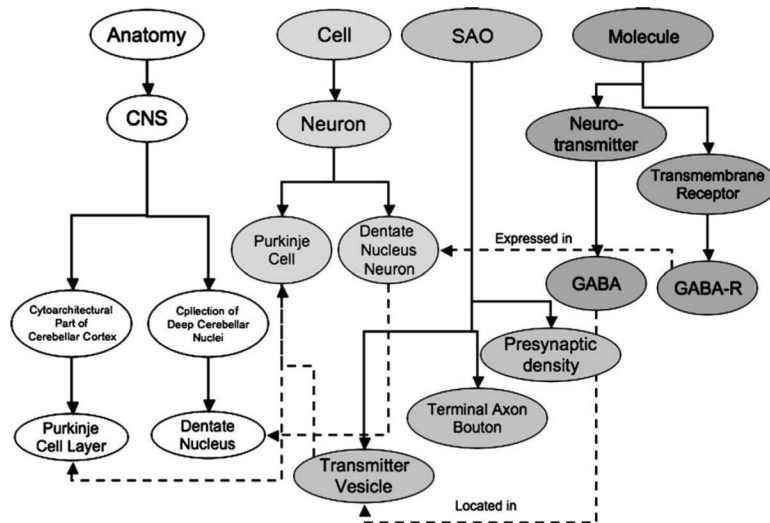
Fig. 6.
Example of cross-domain relations that can be built among NIFSTD modules (NOTE: Current NIFSTD has yet to add expressed_in relations)

Fig. 7.
Sample search of the NIF using the NIFSTD vocabularies to expand and refine the search.
Query results from the NIF Registry, a human curated database of neuroscience relevant sites
on the web, are shown

Table 1

Domains and subdomains covered by NIF, along with the vocabularies imported from external sources for each, and the corresponding NIF OWL module

Domain	External sources	Import or adapt to OWL	NIF module	Unique classes (as of 4/28/08)	Comment
Organism taxonomy	NCBI Taxonomy, GBIF, ITIS, IMSR, Jackson Labs mouse catalog	adapt	http://purl.org/nbirn/birnlex/ontology/BIRNLex-OrganismalTaxonomy.owl	760	Specifically the taxonomy of model organisms in common use by neuroscientists
Molecules	IUPHAR ion channels and receptors, Sequence Ontology (SO); pending: NCBI Entrez Gene, NCBI Entrez Protein, NCBI RefSeq, NCBI Homologene; NIDA drug lists, PDSP Ki, ChEBI, and Protein Ontology	Adapt IUPHAR; import SO	http://purl.org/nif/ontology/NIF-Molecule.owl	3,883	Tested OWL representation techniques on this limited number of molecules (~750). See below for more detail on how molecules in general are to be addressed in NIFSTD
Sub-cellular anatomy	Sub-cellular Anatomy Ontology (SAO)	Import	http://ccdb.ucsd.edu/SAO/1.2.8/SAO.owl	827	SAO covers general cellular structure—referencing the Gene Ontology Cellular Component taxonomy—and more nerve cell-specific structures needed to characterize ultrastructural studies of the nervous system
Cell	CCDB, NeuronDB, NeuroMorpho.org terminologies;	Adapt	http://purl.org/nif/ontology/NIF-Cell.owl	128	

Domain	External sources	Import or adapt to OWL	NIF module	Unique classes (as of 4/28/08)	Comment
	pending: OBO Cell Ontology				
Multi-scale representation of Nervous System Mac Macroscopic anatomy	NeuroNames extended by including terms from BIRN, SumsDB, BrainMap.org, etc	Adapt	http://purl.org/nbirn/birnlex/ontology/BIRNLex-Anatomy.owl	1,398	
Nervous system function	Sensory, Behavior, Cognition terms from NIF, BIRN, BrainMap.org, MeSH, and UMLS	Adapt	http://purl.org/nbirn/birnlex/ontology/BIRNLex-SenseCogBehavior.owl	152	
Nervous system dysfunction	Nervous system disease from MeSH, NINDS terminology; pending: OMIM	Adapt	http://purl.org/nbirn/birnlex/ontology/BIRNLex-Disease.owl	333	
Phenotypic qualities	PATO	Import	http://purl.org/nbirn/birnlex/ontology/BIRNLex-OBO-UBO.owl	1949	Imported as part of the OBO foundary core
Investigation: reagents	Overlaps with molecules above, especially RefSeq for mRNA, ChEBI, Sequence ontology; pending: Protein Ontology	Adapt and import	http://purl.org/nbirn/birnlex/ontology/BIRNLex-Investigation.owl	n.a.	
Investigation: instruments		Import	http://purl.org/nbirn/birnlex/ontology/BIRNLex-Investigation.owl	566	BIRNLex-Investigation imports a BIRNLex-OBI-Proxy file being assembled in parallel with the Ontology of Biomedical Investigation (OBI) This proxy will be replaced by OBI itself, once there is a

Domain	External sources	Import or adapt to OWL	NIF module	Unique classes (as of 4/28/08)	Comment
					full production release of OBI
Investigation: protocols and plans	Biomaterial transformations, assays, data collection, data transformation	Import	http://purl.org/nbirn/birnlex/ontology/BIRNLex-Investigation.owl	(Included in 566 above)	same as above —i.e., ultimately derived from OBI
Investigation: resource type	NIF, OBI, IATR/ NITRC, NCBC Resourceome ontology (BRO)	Mostly adapt, except for OBI	http://purl.org/nbirn/birnlex/ontology/BIRNLex-Investigation.owl	(Included in 566 above)	Will ultimately be a single ontology shared by NITRC, Resourceome, OBI, and NIF

Also indicated is whether the source was in OWL or needed to be adapted, the number of unique classes (concepts) under each domain/subdomain and any comments about the import

Table 2

List of terminology resources used to construct BIRNLex/NIFSTD, along with their URL's, abbreviation and a reference for more information if available

Ontology/vocabulary	URL	Abbreviation	Reference
Basic Formal Ontology	http://www.ifomis.org/bfo/	BFO	Grenon et al. (2004)
Biological Relations Ontology	http://obofoundry.org/ro/	OBO-RO	Smith et al. (2005)
Biomedical Research Network Lexicon	http://birnlex.nbirn.net/ontology/birnlex.owl	BIRNLex	
Biomedical Resource Ontology	Under development	BRO	Dinov et al. (2008)
Brain Architecture Management System	http://brancusi.usc.edu/bkms/	BAMS	Bota et al. (2005)
BrainMap.org	http://brainmap.org/		Fox et al. (2005)
Cell Centered Database	http://ccdb.ucsd.edu	CCDB	Martone et al. (2008)
Chemical Entities of Biological Interest	http://www.ebi.ac.uk/chebi	CHEBI	
Dublin Core Metadata	http://dublincore.org/		
Foundational Model of Anatomy	http://sig.biostr.washington.edu/projects/fm/	FMA	Martin et al. (2003)
Gene Ontology	http://www.geneontology.org/	GO	Ashburner et al. (2000)
Global Biodiversity Info Facility	http://data.gbif.org/welcome.htm	GBIF	
Integrated taxonomic information system	http://www.itis.gov/	ITIS	
International Mouse Strain Resource	http://www.informatics.jax.org/imsr/index.jsp	IMSR	
International Union of Basic and Clinical Pharmacology	http://www.iuphar.org/	IUPHAR	Catterall et al. (2003a, b)
Medical Subject Headings	http://www.nlm.nih.gov/mesh/	MeSH	
Mouse Genome Information mouse strain	http://www.informatics.jax.org/external/festing/mouse/STRAINS.shtml	MGI mouse strain	
information repository		information repository	
National Center for biotechnology information entrez gene	http://www.ncbi.nlm.nih.gov/gene/	NCBI Entrez Gene	
National Center for biotechnology information entrez protein	http://www.ncbi.nlm.nih.gov/protein/	NCBI Entrez Protein	
National Center for biotechnology information Homologene	http://www.ncbi.nlm.nih.gov/sites/entrez? db=homologene	NCBI Homologene	
National Center for Biotechnology Information taxonomy	http://www.ncbi.nlm.nih.gov/sites/entrez? db=Taxonomy	NCBI taxonomy	
National Institute of Mental Health Psychoactive Drug Screening Program	http://pdsp.med.unc.edu/	NIMH PDSP	
National Institute of neurological disorders and stroke online disorder index	http://www.ninds.nih.gov/disorders/disorder index.htm	NINDS disorder index	
National Institute on Drug Abuse drug lists	http://www.drugabuse.gov/drugpages.html	NIDA drug list	
NCBI Reference Sequences	http://www.ncbi.nlm.nih.gov/RefSeq/	NCBIRefSeq	
Neuroimaging Informatics Tools and Resources Clearinghouse	http://www.nitrc.org	NITRC	
NeuroMorpho	http://neuromorpho.org		Ascoli et al. (2007)
NeuroNames	http://braininfo.rprc.washington.edu/		Bowden and Dubach (2003)

Ontology/vocabulary	URL	Abbreviation	Reference
Neuroscience Information Framework Basic vocabulary	http://brainml.org/viewVocabulary.do? versionID=785	NIFBasic	Gardner et al. (2008a, b)
Neuroscience Information Framework standardized Ontology	http://purl.org/nif/ontology/nif.owl	NIFSTD	
OBO Cell Ontology	http://www.obofoundry.org/cgi-bin/detail.cgi? id=cell	CO	
Online Mendelian Inheritance in Man	http://www.ncbi.nlm.nih.gov/sites/entrez? db=OMIM	OMIM	
Ontology of Biomedical Investigation	http://obi.sourceforge.net	OBI	
Open Biological Ontologies foundry core ontology	http://purl.org/nbirn/birnlex/ontology/obo-foundry/core/obo-foundry-core-full-import.owl#	OBO core	
Phenotype and trait ontology	http://www.obofoundry.org/cgi-bin/detail.cgi? id=quality	PATO	Gkoutos et al. (2005)
Protein Ontology	http://proteinontology.info/	PO	Sidhu et al. 2007
Resource description framework	http://www.w3.org/TR/rdf-primer/	RDF	
Sequence Ontology	http://www.sequenceontology.org/	SO	Eilbeck et al. (2005)
Simple Knowledge Organization System	http://www.w3.org/2004/02/skos/	SKOS	
Subcellular Anatomy Ontology	http://ccdb.ucsd.edu/sao	SAO	Larson et al. (2007)
Surface management system database	http://sumsdb.wustl.edu:8081/sums/	SUMSdb	Van Essen (2005)
Unified Medical Language System	http://www.nlm.nih.gov/research/umls/	UMLS	Schuyler et al. (1993)
Web ontology language	http://www.w3.org/TR/2004/REC-owl-features-20040210/	OWL	

Table 3

Neuroanatomical region class for "BIRNLex_1241" (preferred label = amygdala) and associated annotation properties

Annotation property	Value
prefLabel	Amygdala
birnlexDefinition	Subcortical brain region lying anterior to the hippocampal formation in the temporal lobe and anterior to the temporal horn of the lateral ventricle in some species; it is usually subdivided into several nuclear groups; functionally, it is not considered a unitary structure
birnlexCurator	http://purl.org/nbirn/birnlex/ontology/annotation/BIRNLex_annotation_properties.owl#Maryann Martone
birnlexCurator	http://purl.org/nbirn/birnlex/ontology/annotation/BIRNLex_annotation_properties.owl#Bill_Bug
curationStatus	http://purl.org/nbirn/birnlex/ontology/annotation/BIRNLex_annotation_properties.owl#pending final vetting
NeuronamesID	219
createDate	2006-10-05
modifiedDate	2007-10-06
NIFID	5.1.11.1.3.3
synonym	Amygdaloid body
synonym	Amygdaloid nucleus
UMLSCUI	C0002708

See text for a description of the individual annotation properties