



Ontologies for neuroscience: what are they and what are they good for?

Stephen D. Larson* and Maryann E. Martone*

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

Current information technology practices in neuroscience make it difficult to understand the organization of the brain across spatial scales. Subcellular junctional connectivity, cytoarchitectural local connectivity, and long-range topographical connectivity are just a few of the relevant data domains that must be synthesized in order to make sense of the brain. However, due to the heterogeneity of the data produced within these domains, the landscape of multiscale neuroscience data is fragmented. A standard framework for neuroscience data is needed to bridge existing digital data resources and to help in the conceptual unification of the multiple disciplines of neuroscience. Using our efforts in building ontologies for neuroscience as an example, we examine the benefits and limits of ontologies as a solution for this data integration problem. We provide several examples of their application to problems of image annotation, content-based retrieval of structural data, and integration of data across scales and researchers.

Keywords: neuroinformatics, neuroanatomy, databases, subcellular anatomy, data integration

Edited by:

Jan G. Bjaalie, International
Neuroinformatics Coordination Facility,
Sweden; University of Oslo, Norway

Reviewed by:

Mihail Bota, University of Southern
California, USA
Raphael Ritz, INCF, Sweden

* Correspondence:



Dr. Maryann E. Martone is Professor-in-Residence in the Department of Neurosciences at the University of California, San Diego. She also serves as Co-Director of the National Center for Microscopy and Imaging Research (NCMIR), a technology development center supported by the US National Institutes of Health, dedicated to advancing 3D imaging of cells and tissues using light and electron microscopy. Dr. Martone was trained as a neuroanatomist, with expertise in light and electron microscopy. In recent years, she has focused on the development of informatics tools for light and electron microscopic data, including the creation of databases, ontologies and tools for annotation and query of neuroscience imaging data. mmartone@ucsd.edu

INTRODUCTION

Cellular networks in the brain are fundamentally multi-scale with relevant data derived from subcellular junctional connectivity, cytoarchitectural local connectivity, and long-range topographical connectivity. Experimental limitations make it difficult to study all these scales simultaneously. Consequently, experimental methodologies tend to reveal only a limited aspect of nervous system organization. However, to generate hypotheses across scales, we must analyze the nervous system across spatial dimensions spanning several orders of magnitude. Experimental technologies are now able to reveal organization within these scales, yet the development of tools to synthesize these data into more coherent models of brain structure and function is lagging behind.

The amount of neuroscience data now publicly available is significant, with contributions from both large scale efforts like the Allen Brain Atlas (Lein et al., 2007) and individual

neuroscientists. While many neuroscientists admirably are making their data available in publicly accessible databases or web sites (Ascoli et al., 2007), increased availability of data has not occurred within an overarching **information framework** that promotes data exchange and synthesis. Because such a framework is not used routinely by those creating data resources, each database or source tends to use its own terminology and is structured, reasonably so, around its own particular data needs. Asking even straightforward questions that span data sources, e.g., “What genes are expressed in cerebral cortex?” requires a human to confront and reconcile multiple different definitions of the cerebral cortex and descriptions of gene expression from source to source. Comparing data between sources quickly becomes a matter of comparing apples to oranges, as definitions of terms will frequently be difficult to find, incongruous, or expressed from irreconcilable viewpoints.

Information framework

An information framework is a systematic method for assembling knowledge and data within a given domain. A good information framework should allow scientists and other knowledge workers to easily extract, add, and recombine knowledge and data from various sources in a flexible way.

Ontology

An ontology is a formal representation of knowledge in a domain that takes advantage of first-order logic, standardized relationships, and modern data exchange standards such as the Web Ontology Language (OWL). Ontologies consist of a set of classes that represent concepts defining a field and the relationships among these classes.

For a multidisciplinary science like neuroscience, frameworks are also crucial for providing the necessary conceptual bridges to link across disparate disciplines. The time-varying data and research protocols of an electrophysiologist and the spatial-varying data and research methodologies of the microscopist are linked through the neural structures they study. Humans are able to make the connections between a physiological trace recorded from a cortical pyramidal neuron by researcher A and a 3D tree structure derived from the same type of cell from researcher B because they have the requisite knowledge of how these data types relate to the underlying biology. But in most cases, an automated information system does not. The data types produced by different researchers rarely express the linkages that allow those data to be put into the broader context of the brain. The lesson to be drawn from this is that just making data digital doesn't make it ready for integration.

In our work, we have used the problem of multiscale brain imaging to investigate how data acquired by different researchers, using different techniques and data types, can be made available and ready for integration. As one of the foundational strategies, we employ formal ontologies of neuronal structure to provide the conceptual bridge between often disparate and fragmented data. In this focused review, we discuss ontologies for neuroscience – what they are, how they are constructed and how they can be used. Rather than covering the field extensively as in a classical review¹, we provide a practical and personal perspective on the benefits and limits of ontologies as a solution for the data integration problem, using Subcellular Anatomy Ontology (SAO; Larson et al., 2007) as an example. We try, as much as possible, to make the often arcane world of ontologies understandable to a neuroscience audience by avoiding jargon whenever possible. We also discuss some of the lessons we have learned on how to build ontologies for neuroscience, and show how these have been applied in building ontologies on a broader scale for the Neuroscience Information Framework (NIF) project (Gardner et al., 2008a)². Due to space limitations, we do not cover all other ontology efforts in neuroscience but refer readers to recent papers by Bota and Swanson (2008a), Gardner et al. (2008b) and Bowden et al. (2007) for other worthy efforts and different viewpoints on building neuroscience terminologies.

¹A broad overview of neuroscience-related ontologies can be found in Bug et al. (2008).

²All of the ontologies and tools mentioned are available through the Open CCDB Wiki (<http://openccdb.org/wiki>).

INTRODUCTION TO ONTOLOGIES

An ontology is a formal representation of knowledge in a domain that takes advantage of first-order logic, standardized relationships, and, in the information age, modern data exchange standards such as the Web Ontology Language (OWL) (www.w3.org/TR/owl-features). Ontologies consist of a set of classes that represent concepts defining a field and the relationships among these classes. These are distinguished from other ways of organizing knowledge, such as controlled vocabularies and taxonomies, by the richness and expressiveness of relationships. A controlled vocabulary can be thought of as the backbone of an ontology. It is a set of terms in a subject domain that may have been given definitions and unique identifiers, but which has no explicit relationships among these terms. A taxonomy adds to a controlled vocabulary by further organizing terms according to one or more classification criteria. A very well known taxonomy is the tree of life, where living organisms are classified into Kingdom, Phylum, and Class, etc. An ontology builds upon a taxonomy by adding the ability to define other relationships between entities beyond identifier, definition, and place in the taxonomic hierarchy. Relationships such as “part of” allow entities³ within the ontology to be related to one another across the taxonomic hierarchy. These relationships themselves are entities that can be rigorously defined based on what is required to describe the knowledge domain.

In order to manage and perform computations on the different relationships between entities, ontologies are usually encoded into a language that allows a machine to manage and utilize the information. Tools that act like the “word processor” of an ontology such as Protégé (<http://protege.stanford.edu>) use a language like OWL as a convenient standard^{4,5}. Once this knowledge has been captured in a machine processable

³At this point, we switch from “term”, which connotes the words used to refer to a thing, to “entity”, which connotes the thing itself. The entities of an ontology are its classes, properties, restrictions, and instances; the building blocks of its structure. The terms of an ontology are the knowledge content of its classes.

⁴OWL is not the only formalism for encoding ontologies and it should be kept in mind that the notion of an ontology is independent of the language in which it is encoded.

⁵The resource description framework (RDF) is commonly associated with ontologies. RDF is a particular language for encoding knowledge in an exchangeable way, rather than a way of organizing knowledge. RDF can be used to encode controlled vocabularies, taxonomies, or ontologies, as well as other kinds of information. Currently, OWL 1.0 adds formalisms on top of RDF to express ontologies. OWL 2.0 has defined an additional exchange format that encodes OWL directly in XML that does not use RDF.

* Correspondence:



Stephen D. Larson received his Bachelor of Science and Master of Engineering from the Massachusetts Institute of Technology in Electrical Engineering and Computer Science. After working as a software engineer in the financial sector in New York City, he was accepted into the PhD program in the Department of Neurosciences at the University of California, San Diego. During his studies at UCSD, he has published two first-author papers in the field of neuroinformatics. His work focuses on assembling multiscale models of the cerebellum through the combination of spatial and semantic knowledge. slarson@ncmir.ucsd.edu

form, it becomes easier to exchange and utilize this knowledge within information systems. The ontology can be e-mailed or posted on the internet, merged with descriptions of other domains, split apart, and modified algorithmically.

SUBCELLULAR ANATOMY ONTOLOGY

Our entrée into the world of ontologies occurred through our work in building informatics resources for cellular and subcellular data derived from the nervous system. In our previous article, we described a formal ontology of subcellular anatomy (Larson et al., 2007), specifically constructed to describe data from light and electron microscopic imaging and provide the conceptual bridge between whole brain anatomy and macromolecular scales. The SAO builds upon existing ontologies for subcellular components (e.g., the Gene Ontology) and extends them for the nervous system. Its 835 classes encompass cells, parts of cells and supracellular structures like synapses⁶. The SAO is encoded in OWL and uses 68 relationships between its terms such as “*is a*” and “*has part*” derived from the OBO relations ontology (Smith et al., 2005), e.g., Neuron *is a* Nerve cell; Pyramidal cell *has part* Axon. Through these relationships, the SAO allows us to relate macromolecules to subcellular structures, parts of cells to a whole cell or to higher-order brain structures. Thus, any part of a neuron may be localized to a brain region, recognizing the fact that neurons are large cells whose parts span many brain regions.

LESSONS LEARNED

Formalizing knowledge about poorly understood biological systems presents many obstacles to the development of ontologies. Those who are tasked to do so can find it a daunting and ultimately unproductive task. However, through multiple iterations and by application of ontology best practices formulated and promulgated by the Open Biological Ontologies (OBO) community, we distilled a few guidelines that help make the problem manageable (Bug et al., 2008). First and foremost, we limited our scope to design ontologies for the purpose of applying them to data. The ontologies were designed to provide the links between data acquired by a researcher and the biological concepts used to communicate about their meaning and significance. Our goal at first was not, therefore, to encapsulate within the ontology everything that we know about biological systems, but rather to create a structure that enabled clear communication about data.

⁶The SAO contains 43 neuron classes that have been identified across multiple species.

STRUCTURING NEUROSCIENCE KNOWLEDGE:

CLASSES VS. INSTANCES

Because we are designing systems to be applied to data, we found it useful to draw a clear distinction between *classes* and *instances*. In an ontology, **classes** represent the canonical description of an entity while instances represent individual examples of that entity. The Purkinje cell class will have a definition that is consistent with what is generally known about the Purkinje neuron. In contrast, an **instance** of the class Purkinje cell refers to a specific Purkinje cell that has been encountered in an experiment or described in a published report. Thus, we distinguish between Purkinje cells in general (class) and a specific Purkinje cell under investigation (instance). However, it should be noted that this definition of instances is not always consistent across ontologies. For example, Bota and Swanson (2008), consider all members of a class to be instances.

Having these separate views confers several advantages when confronting the complexity of biological systems. Early in our efforts to build ontologies, we tied ourselves in knots trying to capture all of the class-level rules that define something like a Purkinje neuron. We attempted to conservatively define those rules that must be true such as the characteristic number of dendrites or diameter of the cell body. We quickly ran into the well-known problem that these properties exist as wide ranges of legitimate values that differ from species to species and across time and perhaps space. Encoding this information into the class-level descriptions is problematic. On the technical level, ontology languages such as OWL are not particularly good at dealing with numerical values such as ranges or probabilities. On a philosophical level, we know that we have very few examples of even well-studied cells such as Purkinje neurons from which to make generalizations. Because we do not yet have the capability of studying biological organisms across scales without significantly perturbing them, we also know that much of what we observe is colored by the experimental procedures used to prepare and image biological specimens.

As we discuss in Larson et al. (2007), we turned to the ontological *instance* as the vehicle by which individual examples of Purkinje neurons encountered during an experiment or in the literature could be described in a consistent way, and through which biological objects were associated with experimental and data properties. A specific Purkinje neuron is stained according to a particular protocol and imaged with a particular type of microscope. It has numerous primary dendrites and other attributes that can be

Class

In an ontology, classes represent the canonical description of an entity. The Purkinje cell class will have a definition that is consistent with what is generally known about the Purkinje neuron. Thus, we distinguish between Purkinje cells in general (class) and a specific Purkinje cell under investigation (instance).

Instance

In an ontology, instances represent the individual example of a class. An instance of the class Purkinje cell refers to a specific Purkinje cell that has been encountered in an experiment or described in a published report. Thus, we distinguish between Purkinje cells in general (class) and a specific Purkinje cell under investigation (instance).

measured and described. By tying these attributes to formal ontologies like the SAO and NIFSTD, observations taken by different researchers can be aggregated together so that questions like “How many primary dendrites does a Purkinje cell have?” can be answered statistically based on all of the available instances. In this way the *instances* of the ontology can be used to integrate information about things in neuroscience and their associated properties.

NAMES, LABELS AND DEFINITIONS

A second guiding principle promoted by OBO is the use of clear definitions for each entity. When we assert that something is an instance of a class, we are asserting that the thing is identical to the class, not just close in meaning. We therefore require an explicit declaration of the meaning of the entity so that it can be applied appropriately. Thus, annotating with an ontology is somewhat different than choosing keywords to describe data or trying to determine the most appropriate category under which a paper belongs. In these cases, we often chose the closest fit even if it is not an exact match. Because meaning is paramount for consistent application, when building our ontologies, we found it useful to start with a lexicon, i.e., a dictionary of the classes accompanied by a human-readable definition. Following best practices, this human-readable definition is expressed in a way that it is consistent with the machine-readable definition, i.e., the graph structure, in the ontology. For example, the human readable definition “A Purkinje neuron is a type of neuron that is found in the Cerebellar cortex” would be reflected in the class hierarchy (Purkinje cell *is a* Neuron; Purkinje cell *has location* Cerebellar cortex).

By focusing on the definition of a thing, the name by which it is identified becomes less important. As Shakespeare wrote, “A rose by any other name would smell as sweet”. In fact, again adhering to standard practices in the ontology community, in the SAO, the name of the class is a meaningless numerical ID. To make it understandable to a human, each class is also assigned a “preferred label” as an annotation property, which can serve in lieu of the class identifier. For example, class *sao471801888* has the preferred label “Purkinje Cell”. Similarly, multiple alternate labels such as synonyms can be assigned to the same class. If the preferred label turns out to be undesirable for some reason, it can be switched without altering the structure of the ontology. However, if the definition changes, the class is retired and a new class created even though it may still have the same preferred label. Biology also certainly

suffers from its share of homographs, words that are spelled the same but with different meanings, e.g., nucleus as brain region and nucleus as cell part. As these homographs have clearly different definitions, they are distinguished in ontologies by their unique identifiers and positions within the graph.

A common misconception about ontologies is that once they are created they are intended to be rigid definitions that must be agreed to by all who use them. Rather, we see ontologies as a flexible formal medium for arriving at an explicit shared understanding of concepts that define a field and for exposing areas where such shared understanding does not yet exist. By declaring the definition of a thing, the ontology serves as a standard by which other understandings can be compared. When annotating data, the essential point is not whether the researcher agrees wholeheartedly with the entity definition, but that the definition is clear and can be applied correctly. In these circumstances, ontologies can be powerful tools to facilitate clarity in communication and data exchange across sub-disciplines.

As our understanding of ontologies and ontology languages have increased, we have begun to take advantage of more of the class level operations available in OWL to enhance the computing power of the SAO. While using the ontology as a standard way of constructing instances, the advantages of OWL as an ontology language are more fully realized at the level of classes. As a first order logic language, OWL allows the user to define a class not only through its place in the hierarchy but also through a logical definition constructed by the addition of necessary and sufficient conditions, called “restrictions”, and additional rules. Restrictions allow a description logic reasoner such as Pellet (Evren et al., 2005) to make classification inferences about these classes. For example, a set of restrictions about a Purkinje cell in the cerebellum can specify that all instances of a Purkinje cell must have a cell body and that the cell body must be located in the Purkinje cell layer, while the dendritic tree must be located in the molecular layer. Of course, we know that we are likely to encounter instances of displaced Purkinje cells that may violate this constraint, for example, during development or in a pathological condition. Such inconsistencies would be identified by an algorithm checking for consistency between the definitions of the classes and instances as being in an error state. Several steps could then be taken. These include: the definition of the class of Purkinje neuron could be

revised and made more general; the notion of a displaced Purkinje cell could be formalized; or the instance could be thrown out as not qualifying as a Purkinje cell for the purposes of the analysis. In this way, inconsistencies that might otherwise be buried inside a data set are made transparent.

OWL also allows the definition of classes that must be disjoint from one another. When such a restriction is added, we are declaring that two classes cannot overlap with one another. For anatomical brain regions, this type of restriction can be used to define a set of cerebellar regions that, when added together, would account for the entire cerebellum. Again, we know that through gross anatomy and cytoarchitectural characterizations of brain regions, some brain regions may be declared disjoint, e.g., cerebellar cortex and the deep cerebellar nuclei, while others are not, e.g., cerebellar cortex and the cerebellar hemispheres. In this case, we can generate a set of non-overlapping territories by forming cross products between different anatomical parcellations, e.g., cerebellar cortex of cerebellar hemispheres.

One of the recommendations of the OBO Foundry, which has elicited much misunderstanding and heated argument in ontology meetings, is that ontologies be constructed as single inheritance trees. In a single inheritance tree, each entity has only a single parent (super class) and is ideally organized along a single dimension. We found that adhering to the principle of single inheritance actually helps in the construction of ontologies. This is because we do not have to manually create all possible hierarchies to which an entity belongs. Of course, we know that biological entities are complex things and can belong to many different superclasses. For example, a Purkinje cell is a member of the class “neuron”; it is also a member of the class “GABAergic neuron” and the class “spiny neuron”. We have interpreted the rule of single inheritance to mean that each class should only have a single *asserted* parent in the core ontology. However, through the use of properties and logical definitions, additional hierarchies may be *inferred*, again using reasoners applied to the ontology. In the case of neurons, the property *has neurotransmitter* is assigned to members of the class Neuron. We then create a class “GABAergic neuron” defined by a restriction that states “it is any neuron that *has neurotransmitter* GABA”. A reasoner then will classify all neurons for which this condition holds under that class (see Figure 5; Larson et al., 2007).

APPLICATION OF ONTOLOGIES TO DATA

For the past few years, we have been exploring not only the best means to develop ontologies, but how they can be effectively used within information systems to enhance data exchange and search. Here, we illustrate several uses specifically for application to imaging data obtained from light and electron microscopy housed in the Cell Centered Database (CCDB; <http://ccdb.ucsd.edu>), an on-line web-accessible database created to disseminate high resolution microscopic data to the scientific community (Martone et al., 2008).

TAGGING NEUROANATOMICAL DATA WITH ONTOLOGIES

Annotation with ontologies is typically performed after data is acquired, often by dedicated annotators or curators. While an effective model, not all databases have the resources to hire full time annotators or those with the expertise to interpret highly specialized data such as that derived from 3D electron microscopy. Within the CCDB project, we have been developing ways in which ontologies can be incorporated into biologists’ tools so that annotation occurs as a researcher is analyzing data. In the realm of electron microscopy, this approach makes sense because electron microscopists spend many hours carefully segmenting structures of interest from their data, typically through manual tracing. The SAO has been deployed through Jinx, a segmentation program designed principally for looking at micrographs that are the result of electron tomography experiments (Martone et al., 2008). While segmenting, users create objects as instances of SAO classes. Jinx retrieves the latest version of SAO from the web, so that all users of Jinx are accessing the same ontology. Users can also use a subset of SAO relations to provide information about the relationship among segmented structures, e.g., Mitochondrion.000 *has part* Cristae.000. These instances are stored within an adjunct to the CCDB called the Cellular Knowledge Base (CKB), an RDF triple store with pointers to the datasets stored within the CCDB.

QUERY OF IMAGING DATA THROUGH ONTOLOGY

At the simplest level, annotation with the SAO provides a controlled vocabulary for describing subcellular structures, thus avoiding customized and often unrecognizable names assigned to objects. However, the advantages of describing CCDB data as instances of the SAO go well beyond the benefits of a controlled vocabulary. Through the CKB, we can query CCDB data through the

relationships between segmented objects, thereby taking advantage of knowledge encoded both at the class and instance level.

To provide a simple example, the CCDB contains several datasets of neurons filled with an intracellular dye to reveal cellular morphology. If one were to issue a query to the CCDB for all examples of “GABAergic neuron”, CCDB would return zero results even though it has many examples of neurons that use GABA as a neurotransmitter (e.g., Purkinje neurons). The data model of the CCDB primarily represents information about how a dataset was produced. If an experiment does not explicitly look for GABAergic markers, the relationship between GABA and a cell type is not made explicit. The SAO, however, records knowledge about nerve cell classes, including the neurotransmitter. As described above, using the inference capabilities of OWL, the CKB can generate a list of neurons that use GABA as a neurotransmitter and then query the CCDB for instances of those classes.

The goal of many imaging databases is to provide “content-based retrieval”, that is, retrieval of images based on their content and not on high level descriptions of that content. Real-time feature analysis of images remains a difficult challenge, made more so by the multidimensional and heterogeneous content of a database like the CCDB. When a user specifies relationships among segmented objects using Jinx, the content of a very complex scene is turned into a machine-parseable graph. Thus, these relationships can be used to provide content-based retrieval through the CKB. A query to “Find all instances of spines that contain membrane-bound organelles” returns an electron tomography dataset in which the dendritic spine of a Purkinje cell contains smooth endoplasmic reticulum. Note that this example takes advantage of both class-level (*Dendritic spine is a Spine*; *Smooth endoplasmic reticulum is a Membrane-bound organelle*) and instance-level operations (*Dendritic Spine.000 has part Smooth endoplasmic reticulum.000*).

SCALING UP:

THE NEUROSCIENCE INFORMATION FRAMEWORK

The value of ontologies for facilitating data exchange has been recognized in the Neuroscience Information Framework (NIF) project (<http://neuinfo.org>). Funded through the NIH Blueprint consortium, this project has as its goal the creation of a framework for describing and accessing neuroscience resources that are on the web (Gardner et al., 2008a). At the core of the NIF is an expansive ontology covering the broader

domain of neuroscience. The NIFSTD (for NIF standardized) ontology was built using the same core principles outlined previously (Bug et al., 2008) through importing existing ontologies and other terminology resources and standardizing them under the same foundational ontologies used by the SAO (Smith et al., 2005).

The NIFSTD (<http://purl.org/nif/ontology/nif.owl>) was created in OWL using a modular design, with separate modules covering the domains of gross brain anatomy, nerve cells, subcellular anatomy, molecules of excitability, nervous system function, nervous system dysfunction (disease) and technique. Whenever possible, the NIFSTD imported an existing ontology rather than re-inventing one. For example, the SAO was imported to cover subcellular structures in the nervous system. Each class is named with a numerical ID; if the class was imported from an existing ontology, the ID remained unchanged. The NIFSTD provides a preferred label for each class and a set of synonyms, acronyms and abbreviations, along with human readable definitions. Each of the modules consists of a single inheritance tree with a relatively flat class hierarchy. Cross-domain relationships, e.g., nerve cell to brain region, were deliberately not included in the core ontologies but are included in separate files called “bridge files”. This modular structure was chosen in order to make it easy for other applications to import parts of the NIFSTD ontology and build separate extensions and applications around these base ontologies for their specific applications. This approach showcases another feature of OWL – entities can have their definitions extended in a file external to the authoritative ontology that first defines them. This feature is important because it allows an authoritative ontology to contain only the most conservative statements about a domain while allowing other derivative ontologies to fill in more controversial or customized views within their local domains.

BUILDING ON NIFSTD

From the base ontologies established in the NIFSTD, we can now create more complex ontologies that contain a much richer set of intra- and cross-domain relationships. As an example, we illustrate how we are redesigning the SAO so that it is built on top of the NIFSTD core ontologies. In the original SAO, we had a class “Cell” that elaborated different types of nerve cells and “Molecule” which enumerated molecules found in the nervous system. However, when the NIFSTD imported the SAO, many duplicate classes were created because

it imported cell types from other ontologies as well. We then had to spend considerable effort removing and reconciling these duplicate classes. The new SAO imports multiple NIF modules: NIF Cell, Molecule, Anatomy and Subcellular structure, directly from NIF, rather than recreating them. Instead, the SAO confines itself to providing the relationships among these classes, e.g., Subcellular Structure *is located in* Brain Region; Subcellular structure *has part* Molecule. Subcellular structure *is part of* Cell. NIF Cell, Molecule, Brain Anatomy and Subcellular Anatomy. By utilizing the core classes of the NIFSTD, ontologies for neuroscience can be built covering almost any domain. Because they all reference the same core classes, we can aggregate information together through knowledge networks. In addition, in this way, the SAO can take advantage of community contributions to the NIFSTD ontologies from other sources.

CROSS-SCALE INFERENCES

One of the stated goals of providing a formal ontology in a logic language such as OWL is the ability to perform automated reasoning. While algorithms may not yet be able to make scientific discoveries for us, reasoning and classification tools have been developed for OWL which can significantly aid a researcher trying to wade through a sea of heterogeneous data. In previous papers (Larson and Martone, 2007), we described the use of logical rules and the SAO to infer biological structure across spatial scales from limited scenes obtained from electron microscopy. We were able to show that from an annotation of electron microscopic data showing a synapse onto a dendritic spine, connectivity across multiple scales (brain region to brain region; cell to cell) could be automatically inferred⁷. This inference was made possible because the ontology encoded the relationship between different parts of a neuron and their relationship to higher order brain regions.

LOOKING FORWARD

Ontologies can be difficult to construct and maintain and so it is important that any effort spent results in significant returns. By taking a graduated approach in their construction and deployment, we applied ontologies simply and effectively as a framework to solve some basic problems in data annotation and data retrieval. They can be combined with other information systems like relational databases and analysis

⁷See Figure 2, Larson and Martone (2007) for an example.

tools to provide a semantic means by which data are reported and queried. At the same time, we are exploring more advanced features involving reasoning and classification to see how far we can go in utilizing machine-based systems to look for patterns in data and perform the same types of conceptual leaps routinely performed by humans (e.g., reasoning about electron micrographs across spatial scales). Although valuable, we also recognize the limitations of current ontology tools for dealing effectively with very large numbers of concepts, e.g., genes and proteins, and partially overlapping classes, e.g., different brain parcellation schemes. With the construction of the NIFSTD, we believe that we have provided a solid foundation for talented knowledge-engineers to explore such issues in the context of neuroscience data. We also hope that neuroscientists will see the value of knowledge frameworks as a critical part of neuroscience in the digital age and actively participate in the refinement and utilization of these frameworks to advance the practice of neuroscience⁸.

CONTRIBUTIONS

In this review we have:

- Characterized a problem of information management crucial to understanding the organization of the brain across spatial scales and across data modalities
- Introduced ontologies, what they are, and how they can be used to describe the structure of the brain
- Enumerated several examples of the diverse applications of ontologies to help solve the information management problem in neuroscience.

ACKNOWLEDGEMENTS

Supported by NIH grants NIDA DA016602 (CCDB), NINDS RO1NS058296 and NCRR RR04050. The Neuroscience Information Framework is supported by NIH Neuroscience Blueprint Contract HHSN271200577531C via NIDA. The Protégé resource is supported by grant LM007885 from the United States National Library of Medicine. The authors wish to thank Christopher Aprea and Sarah M. Maynard for helpful comments.

⁸The Neuroscience Information Framework (<http://neuinfo.org>) has made community ontology building a priority. The subcellular anatomy ontology has been adopted as one of its core ontologies for bringing neuroscience data together. The NIF project is continuing to refine the SAO as part of the NIFSTD.

REFERENCES

- Ascoli, G., Donohue, D., and Halavi, M. (2007). NeuroMorpho.Org—a central resource for neuronal morphologies. *J. Neurosci.* 27, 9247–9251.
- Bota, M., and Swanson, L. W. (2008). BAMS neuroanatomical ontology: design and implementation. *Front. Neuroinformatics* 2, 2, Epub 22 May 2008.
- Bowden, D. M., Dubach, M., and Park, J. (2007). Creating neuroscience ontologies. *Methods Mol. Biol.* 401, 67–87.
- Bug, W. J., Ascoli, G. A., Grethe, J. S., Gupta, A., Fennema-Notestine, C., Laird, A. R., Larson, S. D., Rubin, D., Shepherd, G. M., Turner, J. A., and Martone, M. E. (2008). The NIFSTD and BIRN Lex vocabularies: building comprehensive ontologies for neuroscience. *Neuroinformatics* 6, 175–194.
- Evren, S., Parsia, B., Grau, B. C., Kalyanpur, A., and Katz, Y. (2005). Pellet: A Practical OWL-DL Reasoner. UMIACS Technical Report, 2005–2068.
- Gardner, D., Akil, H., Ascoli, G. A., Bowden, D. M., Bug, W., Donohue, D. E., Goldberg, D. H., Grafstein, B., Grethe, J. S., Gupta, A., Halavi, M., Kennedy, D. N., Marenco, L., Martone, M. E., Miller, P. L., Müller, H. M., Robert, A., Shepherd, G. M., Sternberg, P. W., Van Essen, D. C., and Williams, R. W. (2008a). The neuroscience information framework: a data and knowledge environment for neuroscience. *Neuroinformatics* 6, 149–160.
- Gardner, D., Goldberg, D. H., Grafstein, B., Robert, A., and Gardner, E. P. (2008b). Terminology for neuroscience data discovery: multi-tree syntax and investigator-derived semantics. *Neuroinformatics* 6, 161–174.
- Larson, S. D., Fong, L. L., Gupta, A., Condit, C., Bug, W. J., and Martone, M. E. (2007). A formal ontology of subcellular neuroanatomy. *Front. Neuroinformatics* 1, 3.
- Larson, S. D., and Martone, M. E. (2007). Rule-Based Reasoning With a Multi-Scale Neuroanatomical Ontology. CEUR Workshop Proceedings 258, ISSN 1613-0073.
- Lein, E. S., Hawrylycz, M. J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A. F., Boguski, M. S., Brockway, K. S., Byrnes, E. J., Chen, L., Chen, L., Chen, T. M., Chin, M. C., Chong, J., Crook, B. E., Czaplinska, A., Dang, C. N., Datta, S., Dee, N. R., Desaki, A. L., Desta, T., Diep, E., Dolbeare, T. A., Donelan, M. J., Dong, H. W., Dougherty, J. G., Duncan, B. J., Ebbert, A. J., Eichele, G., Estin, L. K., Faber, C., Facer, B. A., Fields, R., Fischer, S. R., Fliss, T. P., Frensley, C., Gates, S. N., Glattfelder, K. J., Halverson, K. R., Hart, M. R., Hohmann, J. G., Howell, M. P., Jeung, D. P., Johnson, R. A., Karr, P. T., Kawal, R., Kidney, J. M., Knapik, R. H., Kuan, C. L., Lake, J. H., Laramee, A. R., Larsen, K. D., Lau, C., Lemon, T. A., Liang, A. J., Liu, Y., Luong, L. T., Michaels, J., Morgan, J. J., Morgan, R. J., Mortrud, M. T., Mosqueda, N. F., Ng, L. L., Ng, R., Orta, G. J., Overly, C. C., Pak, T. H., Parry, S. E., Pathak, S. D., Pearson, O. C., Puchalski, R. B., Riley, Z. L., Rockett, H. R., Rowland, S. A., Royall, J. J., Ruiz, M. J., Sarno, N. R., Schaffnit, K., Shapovalova, N. V., Svisay, T., Slaughterbeck, C. R., Smith, S. C., Smith, K. A., Smith, B. I., Sodt, A. J., Stewart, N. N., Stumpf, K. R., Sunkin, S. M., Sutram, M., Tam, A., Teemer, C. D., Thaller, C., Thompson, C. L., Varnam, L. R., Visel, A., Whitlock, R. M., Wohnoutka, P. E., Wolkey, C. K., Wong, V. Y., Wood, M., Yaylaoglu, M. B., Young, R. C., Youngstrom, B. L., Yuan, X. F., Zhang, B., Zwingman, T. A., and Jones, A. R. (2007). Genome-wide atlas of gene expression in the adult mouse brain. *Nature* 445, 168–176.
- Martone, M. E., Tran, J., Wong, W. W., Sargis, J., Fong, L., Larson, S., Lamont, S. P., Gupta, A., and Ellisman, M. H. (2008). The cell-centered database project: an update on building community resources for managing and sharing 3D imaging data. *J. Struct. Biol.* 161, 220–231.
- Smith, B., Ceusters, W., Klagges, B., Köhler, J., Kumar, A., Lomax, J., Mungall, C., Neuhaus, F., Rector, A. L., and Rosse, C. (2005). Relations in biomedical ontologies. *Genome Biol.* 6, R46.

Conflict of Interest Statement: This research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 23 September 2008; paper pending published: 04 November 2008; accepted: 22 March 2009; published: 01 May 2009. Citation: *Front. Neurosci.* (2009) 3,1: 67–60. doi: 10.3389/neuro.01.007.2009

Copyright © 2009 Larson and Martone. This is an open-access article subject to an exclusive license agreement between the authors and the Frontiers Research Foundation, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited.