UNIVERSITY OF CALIFORNIA, SAN DIEGO

Semantic and Spatial Multi-Scale Information Models of the Nervous System

A dissertation submitted in partial satisfaction of the
requirements for the degree
Doctor of Philosophy

in

Neurosciences with a specialization in Computational Neuroscience

by

Stephen David Larson

Committee in charge:

Professor Maryann E. Martone, Chair
Professor Ed Callaway
Professor Gert Cauwenberghs
Professor Mark H. Ellisman
Professor Larry Smarr

2012

The dissertation of Stephen David Larson is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2012

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SUPPLEMENTAL MATERIALS

ACKNOWLEDGEMENTS

Its my pleasure to thank the large community of individuals who helped make this dissertation a reality.

I cannot overstate my gratitude to my Ph.D supervisor, Dr. Maryann E. Martone, whose tireless dedication to building the information science of neuroscience inspired me and gave me a path to follow when I would have been lost. Throughout my time as a graduate student, she has believed in me and encouraged me to chart the way forward in a field in which she is an undisputed leader.

I am also indebted to Dr. Mark H. Ellisman, co-chair of my committee, who, in conjunction with Dr. Martone, enabled me to work within a fantastic group of talented technologists and engineers, to realize a shared vision of how to build the information-fueld biological sciences of the future.

My dissertation committee members, Dr. Ed Callaway, Dr. Gert Cauwenberghs, and Dr. Larry Smarr have been extremely helpful in their feedback and advice over the course of my dissertation. Their perspectives and encouragement were invaluable to helping me craft this dissertation.

Several professors in the UC San Diego community stood up for computational neuroscience and created the graduate program that recruited me, and to them I am extremely grateful to have been given the chance to cross-train in the biological sciences coming from an engineering background. To Dr. Terry Sejnowski, Dr. William Kristan, Dr. Anirvan Ghosh, and Dr. EJ Chichilnisky, thank you so much for creating a great graduate student community.

I am extremely lucky to have been surrounded by talented technologists, engineers, and scientists within the organizations of the National Center for Microscopy and Imaging Research (NCMIR), the Neuroscience Information Framework (NIF), and the Whole Brain Catalog (WBC). To all who made these organizations possible, from funding to coordinating to administrating to building – thank you so much.

Lastly I would like to acknowledge my parents, Edward J. Larson and Maria C. Larson, who gave me the love and support necessary to perservere, who raised me to be curious, conscientious, and to value excellence.

Chapter 2, in full, is a reprint of the material as it appears in Frontiers in Neuroinformatics 2007, 1:3. Larson, Stephen D.; Fong, Lisa L.; Gupta, Amarnath; Condit, Christopher; Martone, Maryann E. The dissertation author was the primary investigator and author of this paper.

Chapter 3, in full, is a reprint of the material as it appears in CEUR Workshop Proceedings 258, 2007, ISSN 1613-0073. Larson, Stephen D.; Martone, Maryann E. The dissertation author was the primary investigator and author of this paper.

VITA

2002	B. S. in Computer Science and Electrical Engineering, Massachusetts Institute of Technology
2003	M. Eng. in Computer Science and Electrical Engineering, Massachusetts Institute of Technology
2012	Ph. D. in Neurosciences with a Computational Specialization, University of California, San Diego

PUBLICATIONS

Ghosh, S., Larson, Stephen D, Hefzi, H., Marnoy, Z., Cutforth, T., Dokka, K., Baldwin, K. K. (2011). Sensory maps in the olfactory cortex defined by long-range viral tracing of single neurons. Nature, 472(7342), 217-20. Nature Publishing Group. doi:10.1038/nature09945

Hawrylycz, M., Baldock, R. a, Burger, A., Hashikawa, T., Johnson, G. A., Martone, Maryann, Ng, L., et al. (2011). Digital Atlasing and Standardization in the Mouse Brain. (L. J. Graham, Ed.)PLoS Computational Biology, 7(2), e1001065. doi:10.1371/journal.pcbi.1001065

Larson, Stephen D, Martone, Maryann E. (2009). Ontologies for Neuroscience: What are they and What are they Good for? Frontiers in neuroscience, 3(1), 60-7. doi:10.3389/neuro.01.007.2009

Martone, M.E., Tran, J., Wong, W. W., Sargis, J., Fong, Lisa, Larson, Stephen, Lamont, S. P., et al. (2008). The cell centered database project: an update on building community resources for managing and sharing 3D imaging data. Journal of structural biology, 161(3), 220231. Elsevier.

Bug, William J, Ascoli, G. a, Grethe, J. S., Gupta, Amarnath, Fennema-Notestine, C., Laird, A. R., Larson, Stephen D, et al. (2008). The NIFSTD and BIRNLex vocabularies: building comprehensive ontologies for neuroscience. Neuroinformatics, 6(3), 175-94. doi:10.1007/s12021-008-9032-z

Fong, LL, Larson, SD, Gupta, A., Condit, C., Bug, W.J., Chen, L., West, R., et al. (2007). An ontology-driven knowledge environment for subcellular neuroanatomy. CEUR Workshop Proceedings (Vol. 258, pp. 16130073).

Larson, S.D., Fong, L. L., Gupta, Amarnath, Condit, Christopher, Bug, W.J., Martone, M.E. (2007). A formal ontology of subcellular neuroanatomy. Frontiers in neuroinformatics, 1(November), 1-12. Frontiers Research Foundation. doi:10.3389/neuro.11/003.2007

Larson, SD, Martone, ME. (2007). Rule-based reasoning with a multi-scale neuroanatomical ontology. CEUR Workshop Proceedings (Vol. 258, pp. 16130073).

Gupta, A., Larson, S., Condit, Christopher, Gupta, S., Fong, Lisa, Chen, Li, Martone, M. (2007). Toward an ontological database for subcellular neuroanatomy. Advances in Conceptual ModelingFoundations and Applications, (8), 6473. Springer.

Larson, S. (2004). Intrinsic representation: Bootstrapping symbols from experience. Advances in Artificial Intelligence, 202216. Springer.

Larson SD, Intrinsic Representation: Bootstrapping Symbols from Experience, M.Eng Thesis, Massachusetts Institute of Technology, 2003

ABSTRACT OF THE DISSERTATION

Semantic and Spatial Multi-Scale Information Models of the Nervous System

by

Stephen David Larson

Doctor of Philosophy in Neurosciences with a specialization in Computational Neuroscience

University of California, San Diego, 2012

Professor Maryann E. Martone, Chair

Neuroscience is a discipline rich in data. In neuroscience, information about the nervous system is derived using a wide-range of experimental methodologies. While each of these methodologies can provide unique insight into nervous system function, each of them are also fundamentally tied to the subjects they measure. Because of this, no single experimental methodology can provide a universal perspective on nervous system function. It becomes sensible, therefore, to explore how we might gain greater insight into the nervous system from unifying the data derived from different experimental methodologies. In this dissertation, I apply techniques from computer science to assemble the rich sets of data in neuroscience

into data structures that allow otherwise impractical or impossible questions to be answered. Because of the nature of the nervous system, these data structures must be capable of dealing with data describing parts of the nervous system that vary in size across eight orders of magnitude, otherwise known as "multi-scale" data. These parts include molecules, parts of cells, cells and their circuits, brain regions and features of gross anatomy. In order to address this challenge, I distinguish between information about the nervous system that expresses facts or semantic relationships (the "what"), from information that expresses images or spatial information (the "where"). I apply a data structure known as an ontology to describe parts of cells relevant to neuroscience, i.e. subcellular anatomy. This ontology is applied to enable automated question answering, i.e. inference, from data sets displaying parts of cells that are derived from electron microscopy. The challenges of building these data structures as a neuroscience community are reviewed, and in response I present an online software system, NeuroLex.org, that provides a first step towards addressing them. Additional examples of automated question answering in neuroanatomy are shown using the NeuroLex knowledge base, which now spans $\tilde{1}$7,000 concepts specific to neuroscience. To address the challenge of the "where" of neuroscience, I present an online software system, the Whole Brain Catalog, that enables two-dimensional, three-dimensional, and four-dimensional (i.e. time-varying) multi-scale data sets to be spatially integrated within a common coordinate system. These data are then navigable by a user in three dimensions in real time and across spatial scales, and can be integrated to answer scientific questions. The dissertation concludes with a look towards the future of the continued convergence between computer science and neuroscience.

Chapter 1

Introduction and Background

The discipline of neuroscience is dedicated to understanding how the nervous system works and how to fix it when something goes wrong. Unfortunately, the quantity and complexity of nervous system cells immediately poses a foundational obstacle to this understanding: *what are the parts of the brain*? This question has been taken up by field of neuroanatomy, which has provided experimental methods for dividing the brain into separate regions, as well as methods for determining the shape and molecular constituents of the cells that compose them, chiefly neurons and glia. Ideally, neuroanatomy would provide a parts' list of the brain that is crisp, comprehensive and uncontroversial. In reality, when one goes beyond a superficial description of the parts of the brain, we find ambiguity, lack of consensus, and large gaps in the literature regarding basic facts of neuroanatomy. Basic questions such as how many different neuronal types make up the mammalian nervous system are still mired in controversy. While thousands of neuroscientists make progress uncovering basic mechanisms of the nervous system, a lack of common understanding about the tissue they study threatens to sharply limit the ability to integrate that progress into a unified understanding (Akil et al., 2011).

Neuroscientists rely heavily on technological advances to expand our capacity to deal with this enormous complexity. Certainly, neuroscience has been the direct beneficiary of recent revolutions in molecular biology, imaging technology and computational technology. These convergent revolutions are producing views

of the brain of increasing size and detail, as we acquire data spanning multiple scales across increasing expanses of brain tissue. With the concomitant increase in computing power now available to scientists, the increased data generation is leading to production of ever more realistic computational models (Coggan et al., 2005; Lopreore et al., 2008), allowing scientists to probe the consequences of the structural and biochemical complexity in ways not amenable to direct experimentation.

The potential power of these integrated approaches is exemplified in large-scale projects such as the Blue Brain (Markram, 2006), the Allen Brain project (Lein et al., 2007) and Genomes to Cognition (Croning et al., 2009). These projects realize huge monetary and manpower investments into the generation of large amounts of data. Because data within these projects are mainly acquired within a single framework, they are able to build powerful informatics infrastructure to serve these data along with custom analysis tools. Mining these richly integrated data sets is starting to yield new insights into how the brain is organized (Lau et al., 2008).

The vast majority of neuroscience, however, is still conducted by individual researchers, who contribute their data and unique insights through less well structured venues such as the literature and websites or the creation of smaller custom databases. Although the amount of data is increasing daily, neuroscience as a whole, with its exceptionally large scope and diverse research community, lacks a coherent community framework for bringing these data together. Because such a framework has not been readily available, each source tends to use its own terminology and is structured, reasonably so, around its own particular data needs. Such customization is a significant barrier to data integration, because it requires considerable human effort to access each resource, understand the context and content of the data, and determine the conditions under which it can be compared to other similar results. The effect of this customization is that much neuroscience data is opaque to modern computational and bioinformatics tools that can operate across vast amounts of data, but require information to be parsed by a machine in order to be accessed and utilized.

Why is the data integration problem so difficult in neuroscience? Neuroscience, unlike molecular biology, does not have a single data type like a gene sequence that is easily stored or exchanged. Nor, like the geosciences, does it have a single well-characterized spatial framework in which to place data. Without these types of hooks, it is difficult to create common tools like Google Earth that mash up data coming from diverse sources. Neuroscience, despite modern 21st century community building tools, remains largely a set of isolated communities, with data isolated in different silos. Thus, building a successful knowledge framework for neuroscience in the modern era, in which access to data and information is largely through computers, requires a multipronged approach to accommodate the diversity of data and the multiple temporal and spatial scales over which they are acquired. Essentially, a framework should be able to specify for each piece of data what, when and where and provide the means for tying them together; that is:

1. A coherent semantic framework encapsulating the concepts that neuroscientists use to communicate about the content of their data, the experimental conditions under which they were acquired and the conceptual and temporal relationships among them;

2. A spatial framework for tying data to its general location in the nervous system; and

3. A community infrastructure software base where researchers can share tools and data easily.

Building these types of frameworks is hard, but through significant national and international investments in infrastructure over the past decade, the base elements of such a framework are beginning to emerge. In this dissertation, I will describe the challenges in each area described here, the parts that I have contributed, and the pieces still left to build in the future to establish an integrated cyberinfrastructure for neuroinformatics.

Biological Driver

To illustrate the challenges in multiscale integration, Consider the following question regarding the anatomy of the cerebellum, a brain region whose structure has been extensively studied across multiple scales: "what is the synaptic organization of adjacent cerebellar glomeruli with respect to the axon terminals of afferent mossy fibers?". What kind of knowledge is needed to answer it?

First we would need the basic definitions of the terms involved. The glomeruli of the cerebellum are complexes of synapses having a mossy fiber ending as its core, synapsing with axons of Golgi type II neurons and dendrites of granule cells. The mossy fibers refer to axons that make up the primary input to the cerebellum and enter from many different nuclei of the brain stem, excluding the inferior olivary nucleus. In order to understand the synaptic organization of glomeruli, the position and orientation of those glomeruli would need to be known with respect to the context of the positions of cells in the cerebellar cortex.

Second we would need to understand the receptive fields of the mossy fibers of the cerebellum. We want to know what modalities of information are transmitted by them, as well as anatomically, where the cells that give rise to the mossy fibers are, and how their axons make their way to the cerebellar cortex. Third, we would need to understand how the mossy fibers from different cell populations distribute their projection fields throughout the cerebellar cortex. Finally we would need to understand how glomeruli are distributed within a microdomain of those projection fields.

The literature is scattered with studies that hold pieces of this puzzle. Is the answer known or does it require a new experiment? The first challenge is finding the puzzle pieces that exist in the literature. Studies that have examined the organization of mossy fibers in the cerebellar cortex generally do not track the origins of the mossy fibers that give rise to the glomeruli (e.g. Garwicz et al., 1998). Studies that track the origins of the mossy fibers to the cerebellar lobules where they project do not look at how the axon terminals are organized (e.g. Cooke et al., 1971). In order to understand the organization to answer a related question such as "do neighboring glomeruli receive projections from the same or different

mossy fibers?" we must assemble this knowledge in such a way that we can "track" mossy fiber originating cells from their somatic origins all the way to their axon terminals.

Given the existence of other projects that have tackled problems in assembling neuroscience information, (Hucka et al., 2002; Boline et al., 2006; Joshi et al., 2009) it would be desirable to be able build a multi-scale model of the cerebellum using some set of them. It is unfortunately very difficult to re-use the content of these other projects for these purposes. As with most information systems, they have been built with a single purpose in mind, generally to answer a particular set of questions in a particular domain of neuroscience. However, when attempting to answer questions that require knowledge outside of the specialty domain of a given information system, significant problems are met. Switching from information system A to information system B to pursue a question all too often results in irreconcilable gaps between the different ways system A and system B chose to represent their information.

This points to the major limitations a neuroscientist faces when trying to utilize the current set of information systems to answer questions. If the answer to a question is known, how can it be found? If the answer is known, it is frequently scattered across multiple systems. If the answer is not known, information systems do not make it obvious what are the missing pieces in the literature needed to answer the question. There is a disconnect between what neuroscientists want to know from information systems, and what information systems are being built to allow neuroscientists to accomplish.

The approach taken in this dissertation is to build a system that provides a framework and a set of tools for re-assembling the fragmented information into a unified whole irrespective of brain system or scale, rather than providing an information system that is designed only to handle a single type of data or a single scale of data. It is for this reason that focusing on abstractions like "what" and "where" knowledge is powerful, because it provides us a way to build information systems that can handle neuroscience knowledge irrespective of domain or scale.

1.1 Background

As a means of grappling with the multi-scale challenges of brain structure, it is useful to divide the knowledge of brain structure into separate knowledge frameworks. Clearly defined and delineated knowledge frameworks are helpful because they facilitate the construction of appropriate knowledge representations, which in turn facilitate the construction of appropriate information systems. A simple division can be made between the "what" and the "where". The "what" consists of the logical statements that capture relationships between parts of the brain (e.g. "mitochondria are inside neurons"). In contrast, the "where" consists of spatial relationships, which are expressed with respect to an absolute coordinate system from an atlas or with respect to the coordinates of an image.

Logical statements and semantic frameworks

Scientists regularly find themselves organizing what they know into lists and categories. In chemistry, a shining example is the periodic table of elements, with data such as atomic weight, proton and neutron count included, and categories such as noble gases and metals used to group them (Connelly and Damhus, 2005). How would the field of chemistry make progress if its definitions were ambiguous, its names were unevenly applied, and the number of elements ranged into the millions? This is analogous to the problem faced when trying to organize knowledge of the pieces and parts of biology that impact activity in the nervous system.

For a multidisciplinary science like neuroscience, some sort of formal semantic framework is critical to provide the necessary conceptual bridges to link data 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, not directly through the data types produced. 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.

As a means of resolving this disconnect, there has been a popular movement towards encoding logical statements in flexible XML-based knowledge representations such as the resource description framework (RDF; `http://www.w3.org/RDF/`) and more recently the Web Ontology Language (OWL; `http://www.w3.org/TR/owl-features/`) (Ruttenberg et al., 2009; Cheung et al., 2009; Larson and Martone, 2009). This has caused great interest in using "ontologies" to express knowledge, particularly in the life sciences. Here, an ontology means a set of logical statements represented in a formalism such as OWL capable of performing reasoning. The type of reasoning supported by a well constructed ontology ranges from simple subsumption, e.g., a Purkinje neuron is a neuron, a neuron is a cell, therefore a Purkinje neuron is a cell, to more rule-based logical deductions, e.g., a GABAergic neuron uses GABA as a neurotransmitter, a Purkinje neuron uses GABA as a neurotransmitter, therefore a Purkinje neuron is a GABAergic neuron. Ontologies also deal with temporal entities and relationships. For example, upper level ontologies usually divide the world into two branches, one encompassing the what, e.g., objects such as organisms, organs, cells and molecules, and the other encompassing the when, e.g., events and processes (Grenon, 2003; Grenon et al., 2004; Smith et al., 2005).

Previous work relevant to this dissertation has included the construction of a comprehensive ontology for neuroscience covering many disparate domains within the discipline (Bug et al., 2008; Larson and Martone, 2009). This work demonstrated that concepts in neuroscience were capable of being formalized to the degree that they could be defined in a machine processable format suitable for exchange between information systems (OWL). The resulting ontology, the NIF Standard ontology, has considerable breadth as its 20,000 concepts were derived from vocabularies from different sources in neuroscience, bridging molecules, subcellular anatomy (see chapter 2), cell types, brain regions, experimental instruments, qualities, and others. Contained within it are basic facts of categorization, such as the fact that a neuron is a type of nervous system cell, which is a type of cell, as well as human readable definitions, synonyms, relationships between brain regions such as which region is a part of which other region, etc. With the

NIF Standard, the task of constructing information systems that use neuroscience concepts as its computable parts is a practical endeavor.

In this dissertation, in order to demonstrate the process of building an ontology for neuroscience, I describe the construction of a portion of the NIF Standard dealing with subcellular entities, named the Subcellular Anatomy Ontology (Chapter 2). The subcellular anatomy ontology (SAO; Larson and Martone, 2007) specifies the relationships among these parts, e.g., neuron has compartment dendrite; dendrite is continuous with the cell soma.

To demonstrate the potential to gain new knowledge from ontologies that had not been specified in advance, chapter 3 describes a preliminary study that was conducted to explore making deductions over ontologies with rule-based reasoning systems. This chapter demonstrates deductions made by using IF-THEN rules that the presence of an axon terminal implies the presence of a neuron. Here the focus was on making inferences about the kind of structures visible in electron micrographs: parts of cells, macromolecules and supra-cellular aggregates. The latter class includes structures that span two or more cells, e.g., synapses. Figure 2 illustrates how having some knowledge about where the axon terminal is coming from can be leveraged to infer new information.

While these technologies provide valuable infrastructure for the expression of biological knowledge in a rigorous and reusable manner, another roadblock lies in the path to widespread adoption of these technologists by the most important end-users, the neuroscientists on the front lines of observation and data collection. Like learning a new mathematics, the formalisms of ontologies requires a focused investment of time and attention to use gainfully. This roadblock prevents all but the most motivated of neuroscience experts to venture into the community of ontology researchers who know how to best build ontologies. Collaborative work between biomedical ontology experts and neurobiology experts is still rare. When considering the immense size of the domain of biology and all the entities that need to be described, this lack of cross talk is a major challenge to the enterprise of neuroinformatics.

However, the Internet is defining new modes of interaction that are impact-

ing younger generations around the world. Scientists are only beginning to take advantage of collaborative tools such as Wikis and social networking software to promote easy exchange, not only of data, but of opinions, protocols and professional networks (e.g., LinkedIn.com). The advantage of these communities is that scientific discourse can occur in the public sphere, where an electronic record is left behind of individual's expertise and experience (Clark and Kinoshita, 2007). Taking inspiration from these developments, and by combining several open source technologies related to semantic wikis and the NIF Standard Ontology, I have created NeuroLex.org, the first semantic wiki for neuroscience, described in chapter 4

Coordinate systems and spatial frameworks

The "where" side of the puzzle has two major components to it. First, one can describe spatial relationships with respect to a single uniform coordinate system across the whole brain. These kinds of relationships lend themselves to formalizations and algorithms that have been devised by the geospatial information systems community, who deal with the problem of mapping the surface of the earth using the coordinate system of the global positioning system (GPS). Second, one can describe spatial relationships with respect to the coordinate system of a 2D or 3D image, such as those taken from different kinds of brain imaging, e.g. light microscopy, electron microscopy, fMRI, etc. These kinds of relationships are usually captured within the image formats themselves, describing data as pixels, voxels for 3D data, or geometric, vector-based descriptions.

It is worth noting that this is a pattern that is already being followed by a new generation of information systems. Google Maps (`http://maps.google.com`) is an example of an application that combines "where" knowledge, using GPS technology, with "what" knowledge. The current version of the tool enables users to turn on additional information that allows places of interest (the "where") to be overlayed with an icon that takes a user to the Wikipedia (`http://wikipedia.org`) article about that place (the "what"). Google Earth (`http://earth.google.com`) provides more functionality via a rich client that a user can download to their computer. These information systems are popular and useful, and provide a good

model for neuroscience to follow.

Also popular amongst the general public, particularly younger individuals, is the use of virtual 3D online environments and massively multiplayer online game engines (e.g., Second Life, World of Warcraft, RuneScape, etc). Such environments offer new and exciting possibilities of much more interactive modes of collaboration in new immersive visualization environments where multiscale data can be synthesized. Such environments would allow individuals around the world to participate in the construction of a detailed synthetic view of the brain. While these systems are all useful for their particular purposes, they fall short in providing the kind of environment needed to assemble multi-scale models of the nervous system. In chapter 5, I describe a system that directly addresses the challenge of building an immersive visualization environment like Google Earth, but for the nervous system, called the Whole Brain Catalog.

Unification

These two axes, the "what" and the "where", make up the dimensions of the space of information that are involved when trying to assemble a complete picture of a complex structure like the nervous system. The literature is filled with statements that can be turned into logical relationships, and we continue to fill digital repositories with images taken from many different scales of the brain using many different experimental protocols. With this understanding, we can therefore reformulate the problem of making sense of the data explosion in neuroscience with the problem of building an information framework that is able to integrate the "what" and "where" dimensions and allows further analysis to be done on that integrated information. In chapter 6, I summarize the contributions of the dissertation.

Chapter 2

A formal ontology of subcellular neuroanatomy

2.1 Introduction

In neuroscience, scientifically relevant complexity occurs at every spatial and temporal scale that is currently open to examination. Unfortunately, our current complement of experimental and analytical techniques generally locks an investigation into a very limited dimensional range, leading to a fragmented and incomplete view of nervous systems across scales. This fundamental multiscale problem of neuroscience is, at its core, a problem of information integration. One indication of the extreme difficulty of information integration in the neurosciences is the conspicuous lack of any widely practiced automated methods for integrating information among major classes of neuroscientific data: structural, functional, and behavioral. Many tools have been developed to provide infrastructure to organize and analyze brain data, resulting in large part from the Human Brain Project, funded through the US National Institutes of Health (Huerta et al., 1993; Koslow and Huerta, 1997). Such tools have included databases for storing primary data (e.g., CCDB; Martone et al., 2003, WebQTL; Wang et al., 2003, etc.), knowledge bases for derived information (e.g., BAMS; Bota et al., 2005 and CoCoMac; Stephan et al., 2001), tools for performing novel analyses of brain data and mining

the literature (e.g., Textpresso; Müller et al., 2004). However, the integration of diverse types of information still occurs largely through the efforts of individuals who examine the data and construct the necessary bridges between different data based on their knowledge of neuroscience.

The grand challenge of neuroinformatics is the creation of systems that seamlessly integrate data across spatial and temporal scales such that information, for example, about white matter bundles derived from diffusion tensor imaging can be analyzed in context with electrophysiological data recorded from the neurons whose axons make up the bundles. The difficulties in performing this type of integration from data alone is illustrated in Figure 2.1, which shows an intracellularly injected medium spiny neuron from the mouse nucleus accumbens, imaged using correlated light and electron microscopy. At each level, different types of visualization and analytical tools are applied to extract meaningful content, for example, the branching structure of the dendritic tree, the surface area of dendritic spines. The knowledge required to richly inter-relate these different data representations and analytical results, however, largely resides in the domain scientists with specific, detailed understanding of the links between the various data types and the biological objects from which they derive.

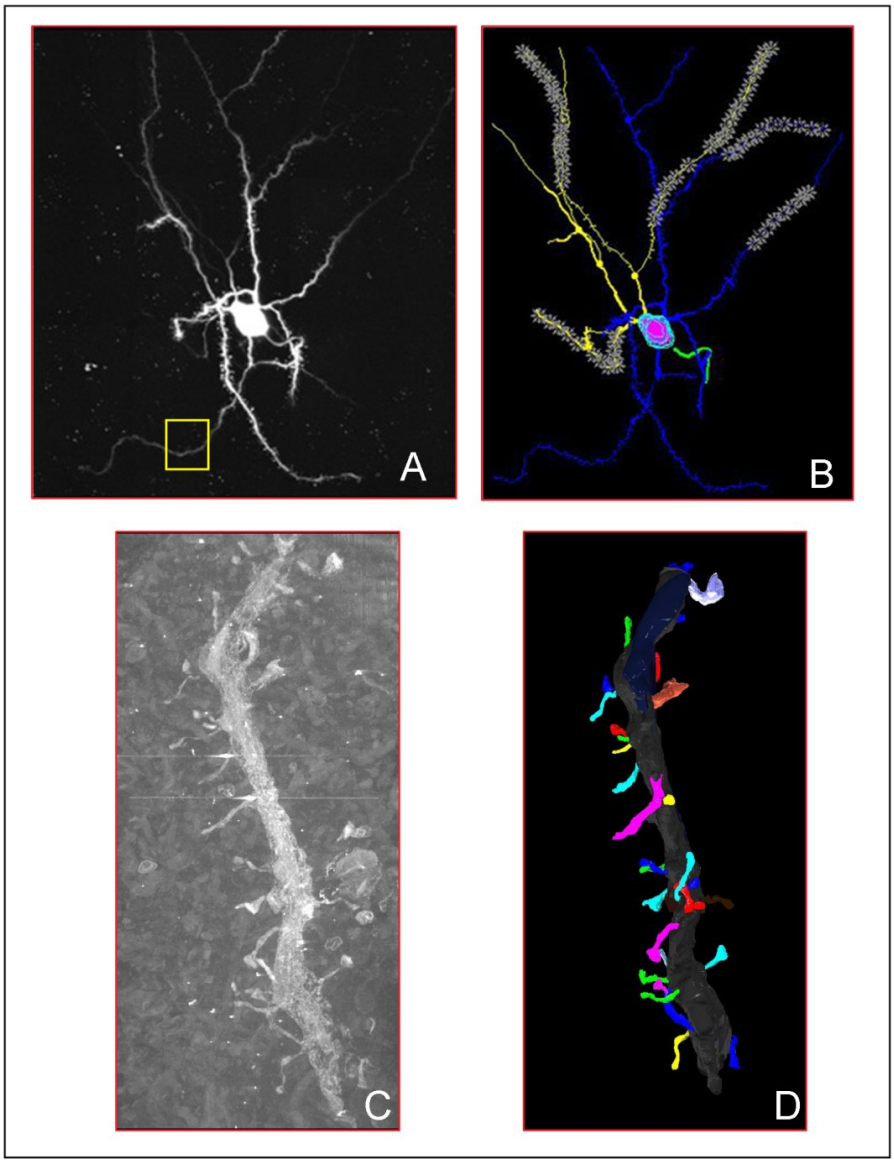

Figure 2.1: Multiple representations of the same medium spiny neuron taken from the CCDB. In (A), a light-level fill of the neuron. The yellow box shows the portion of the dendritic branch shown in (C). In (B), the Neurolucida segmentation of that neuron. In (C), the EM image of the portion of the dendrite featured in (A). In (D), the 3D reconstruction of the dendrite from (C) after segmentation.

In this chapter, we describe specific steps toward creating generic information bridges by constructing a formal ontology designed to provide the knowledge necessary to integrate data acquired across multiple scales in structural neuro-

science. An ontology is a formal representation of knowledge in a domain (Gruber, 1993). It defines the inter-related set of concepts representing a knowledge area and the common terms used to describe them, for example, neuron *is a* cell and cell *has part* plasma membrane. A critical aspect of modern ontologies is the encoding of these entities and relationships in a standard form where the semantics of the domain are machine interpretable using open source tools and software libraries. Ontologies are used by people, databases, and applications to share information in a semantically precise way within and across particular domains (Gruber, 1993).

The ontology for subcellular anatomy (SAO) focuses on the spatial scale that has come to be known as the mesoscale, roughly defined as the dimensional range encompassing macromolecular complexes, subcellular structures up to the level of cells and cellular networks. The SAO describes neurons, glia, their parts, and how these parts come together to create the dense feltwork of processes that characterizes the nervous system. The SAO was constructed through the Cell Centered Database (CCDB) project (Martone et al., 2002, 2003, 2008a), an on-line resource for disseminating data derived from light and electron microscopic imaging. The CCDB project, as its name implies, takes the view that the cell should provide the rallying point for information integration in biological tissues. Thus, the SAO starts with the cell and models how cell parts, including molecules, fit into coarser levels of anatomy. This view contrasts with the approaches of many ontologies that start at the level of gross anatomy and traverse down to the level of the cell, for example, the Foundational Model of Anatomy (FMA) (Rosse and Mejino, 2003) and BAMS (Bota et al., 2005). The SAO was built as a reference ontology with the ultimate goal of describing data, principally derived from light and electron microscopy, through the use of multiple annotation applications. It is built using the Web Ontology Language (OWL; `http://www.w3.org/TR/owl-features/`) a W3C open standard for ontologies. Version 1.0 of the SAO was presented in Fong et al. (2007), which concentrated on the use of OWL and the associated tools for its construction. In this paper, we present an updated version (1.2) of the SAO, provide considerably greater detail on the design principles from a neuroscience point of view, describe new examples of reasoning, and describe new examples of

data that are marked up using the SAO. We also briefly illustrate how it is being used as the semantic glue that binds together an environment of tools capable of annotating disparate types of structural data from imaging studies of the nervous system.

2.2 Methods

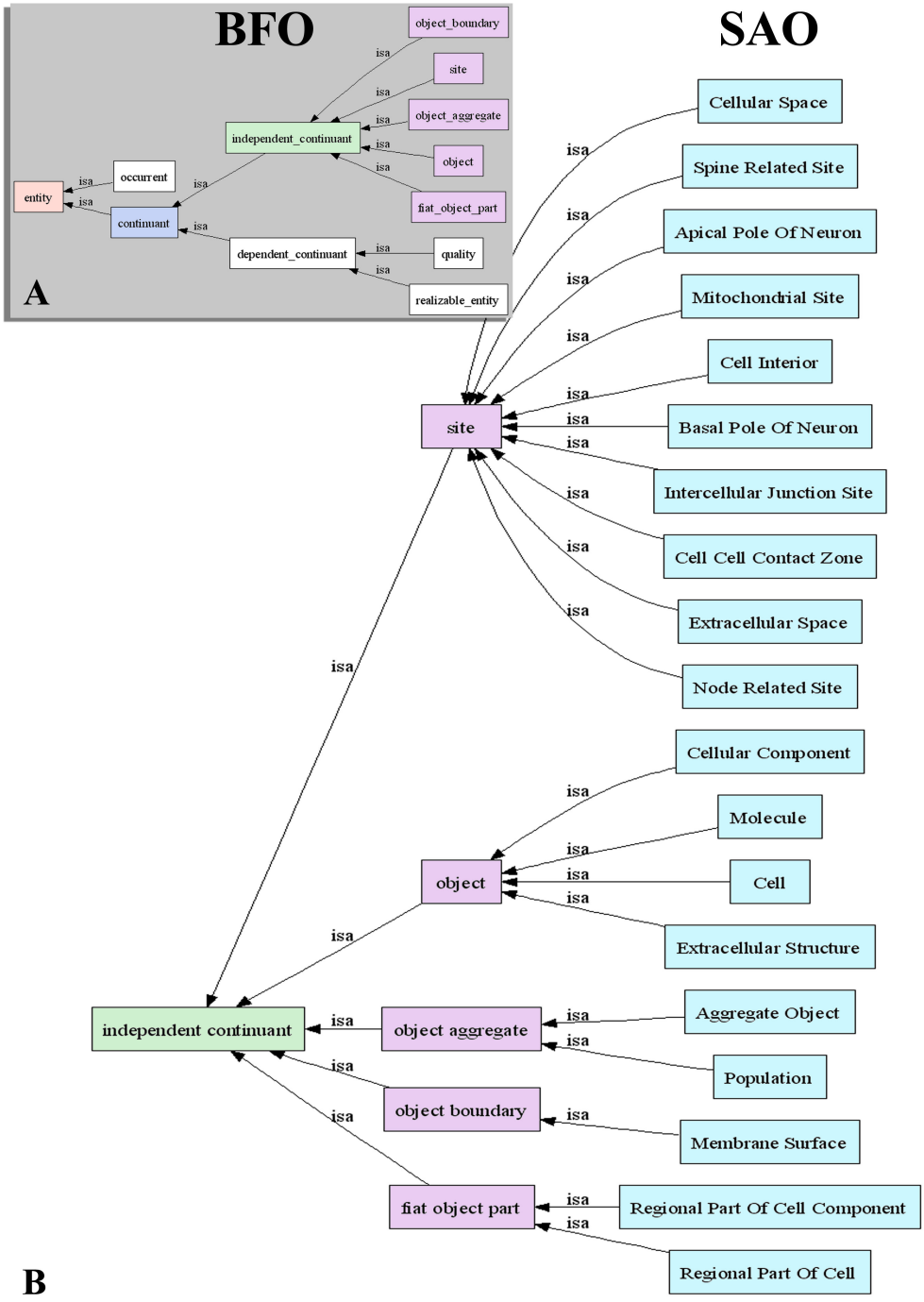

Figure 2.2: High level class structure of the SAO. The BFO entities are shown in (A) and in the green and pink boxes in (B). SAO classes that are under the BFO hierarchy are shown in blue in (B).

The primary source for subcellular anatomy used for the construction of the SAO was Peters et al. (1991) The Fine Structure of the Nervous System Ed 2, the standard reference for neuronal ultrastructure. Additions and modifications to this framework were also made from more recent literature. The source of each entity in the ontology is indicated as an annotation to the concept. As a way to keep epistemological distinctions clear, we adopted as an organizing framework the Basic Formal Ontology version 1.0 (BFO 1.0; Grenon, 2003) (Figure 2.2). The structure/function dichotomy is expressed in the BFO through the division of all possible entities into continuants (objects, qualities, sites, etc.) and occurrents (dynamic processes, temporal intervals). A continuant is an entity in the world that endures through time (Grenon et al., 2004). Examples of continuants are basic cell structures such as mitochondria and nuclei, as well as lumens and membranes. On the other hand, an occurrent refers to a process, event, activity, or change. Examples include the cell cycle phases, cell secretion, and motility. The BFO further divides continuants into dependent and independent continuants. An independent continuant is an entity that exists irrespective of its relationship to anything else, for example, cell, organism. A dependent continuant is an entity that inheres in an independent continuant, for example, color, age.

The SAO is available for download and browsing at (`http://ccdb.ucsd.edu/SAO`) and has been incorporated into the BioPortal1, a resource maintained by the National Center for Biomedical Ontologies (`http://www.bioontology.org/ncbo/faces/index.xhtml`). The SAO is expressed in OWL DL. OWL is a vocabulary extension of the Resource Description Framework (RDF) and is derived from the DAML + OIL OWL. Together with RDF and other components, these tools make up the growing semantic web community (Neumann and Prusak, 2007). One of the goals of the semantic web is to create tools for achieving highly interoperable data resources. The SAO was composed using Protègè version 3, an open source authoring tool for OWL ontologies (Noy et al., 2001; Rubin et al., 2007). The OWL standard is designed as a kind of description logic, which means that an application domain described in OWL is automatically described using formal logic-based semantics. One benefit of this is that tools like Protègè and additional reasoning

tools such as Pellet (Sirin et al., 2007) and Swoop (Kalyanpur et al., 2006) can identify statements that are logically inconsistent. It also supports machine-based inferencing to generate new knowledge and to provide classification. The other major benefit is the machine-readability of OWL, which can be expressed as an XML document. This means that arbitrary software applications can take advantage of the knowledge and data that is encoded in an ontology as their underlying data model. It also means that ontologies written in OWL can be automatically imported and cross-linked by other ontologies.

An OWL ontology contains a series of classes, properties, and annotations. The classes are simply the entities that are organized in a top-down hierarchical graph structure (Figure 2.2A). Classes contain subclasses, for example, neuron and glia are subclasses of nerve cell. Subclasses are related to superclasses through the *is a* relationship, for example, neuron *is a* nerve cell. Properties are parts or attributes a class, for example, nucleus *is a* part of a cell; age *is an* attribute of organism. Properties are typically related to a set of classes through some form of *has a* relationship, for example, cell *has part* nucleus. Properties may be related to other properties through inverse, symmetric or transitive relationships, for example, is part of is the inverse of has part. Annotations are used to record metadata about the entity, for example, definitions, abbreviations, synonyms, sources of data, comments, and references. OWL allows for the placing of restrictions on classes, defining necessary and sufficient conditions for classification, and providing constraints on what properties need to be filled in for a given class, for example, (Neuron *has regional part some* Regional Part of Neuron) is a restriction that requires that a Neuron be related with the property has regional part to the class Regional Part of Neuron.

In the OWL language, all properties are first-class entities, meaning they exist independently of classes they are used to describe. Consequently, whether using properties as attributes, or as relations, the same underlying logical mechanism is invoked. Therefore, OWL properties do not have the facility to distinguish between structural properties (i.e., attributes) and relationships between classes (i.e., relations). Instead, structural properties are defined through the use of OWL

restrictions, which we have used throughout the SAO. These can be seen in Figure 2.6, where arrows with blue text describe relationships enforced by restrictions, where arrows with black text describe relationships defined only for this particular instance.

In constructing the SAO, we have tried to adhere to best practices recommended by the OBO Foundry project (Smith et al., 2007). These practices include unique identifiers for each concept, re-use of existing ontologies where possible, provision of human-readable definitions that are consistent with the machine interpretable definitions encoded within the ontology. The SAO follows the principle of single inheritance as recommended by Smith et al. (2007). Single inheritance results in a *is a* hierarchy that is a simple tree, where children have only one parent. Through the assignation of the part of relationships, we utilize some of the features of OWL to cross-cut the *is a* hierarchy such that new hierarchies can be generated. Examples of this concept will be illustrated in the Results section. For the SAO, we incorporated several existing ontologies using the owl:imports mechanism of OWL within Protègè 3. In this way, we do not reinvent content that is already substantially covered in other ontologies. The import mechanism allows wholesale incorporation of existing ontologies into the SAO while maintaining the integrity and source of the original ontology. In addition to the BFO, we imported an extensive set of annotation properties from the BIRNLex (`http://nbirn.net/birnlex`). Entities may be added to a merged resource, but entities may not be deleted or modified nor the class structure changed. Additional resources of relevance, for example, the cell component hierarchy from Gene Ontology, that were not encoded in OWL, were imported manually and cross referenced to the appropriate identifiers.

2.3 Results

2.3.1 Structure of the SAO

Classes

The high level structure of the SAO is illustrated in Figure 2.2B. The main classes of biological independent continuants within SAO are Cell, Regional Part of Cell, Cell Component, Extracellular Structure, and Molecule. The current version primarily covers structural entities that would be observed within the adult mammalian nervous system. Each class is assigned a unique identifier. We utilize the class identifier as the class name, but also assign a commonly used human understandable label to each class, for example, sao1224657022 corresponds to the label Nerve Cell.

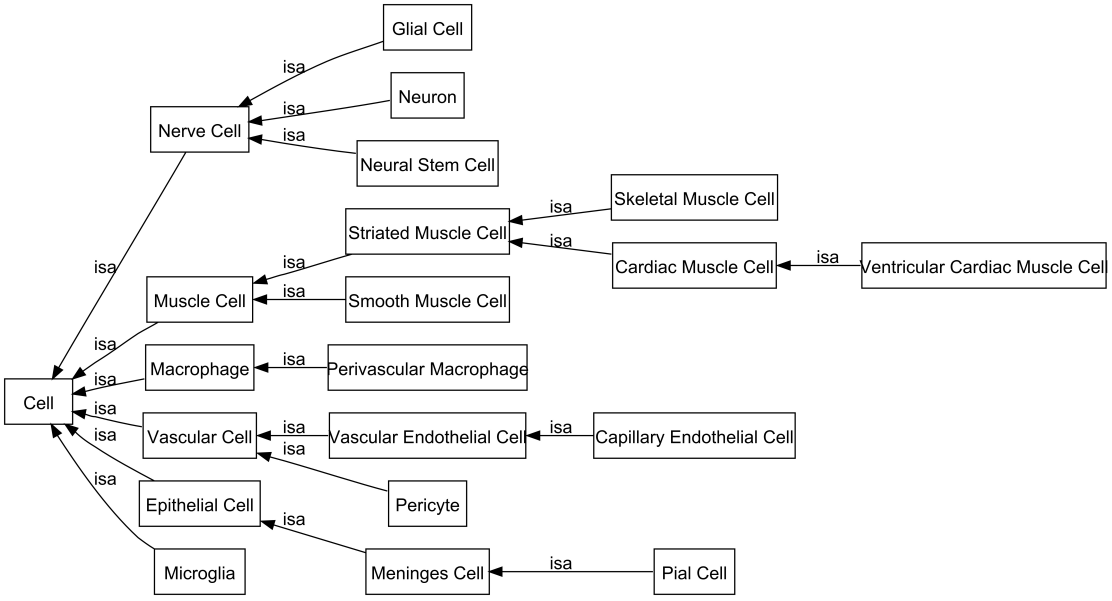

Figure 2.3: Cell Hierarchy.

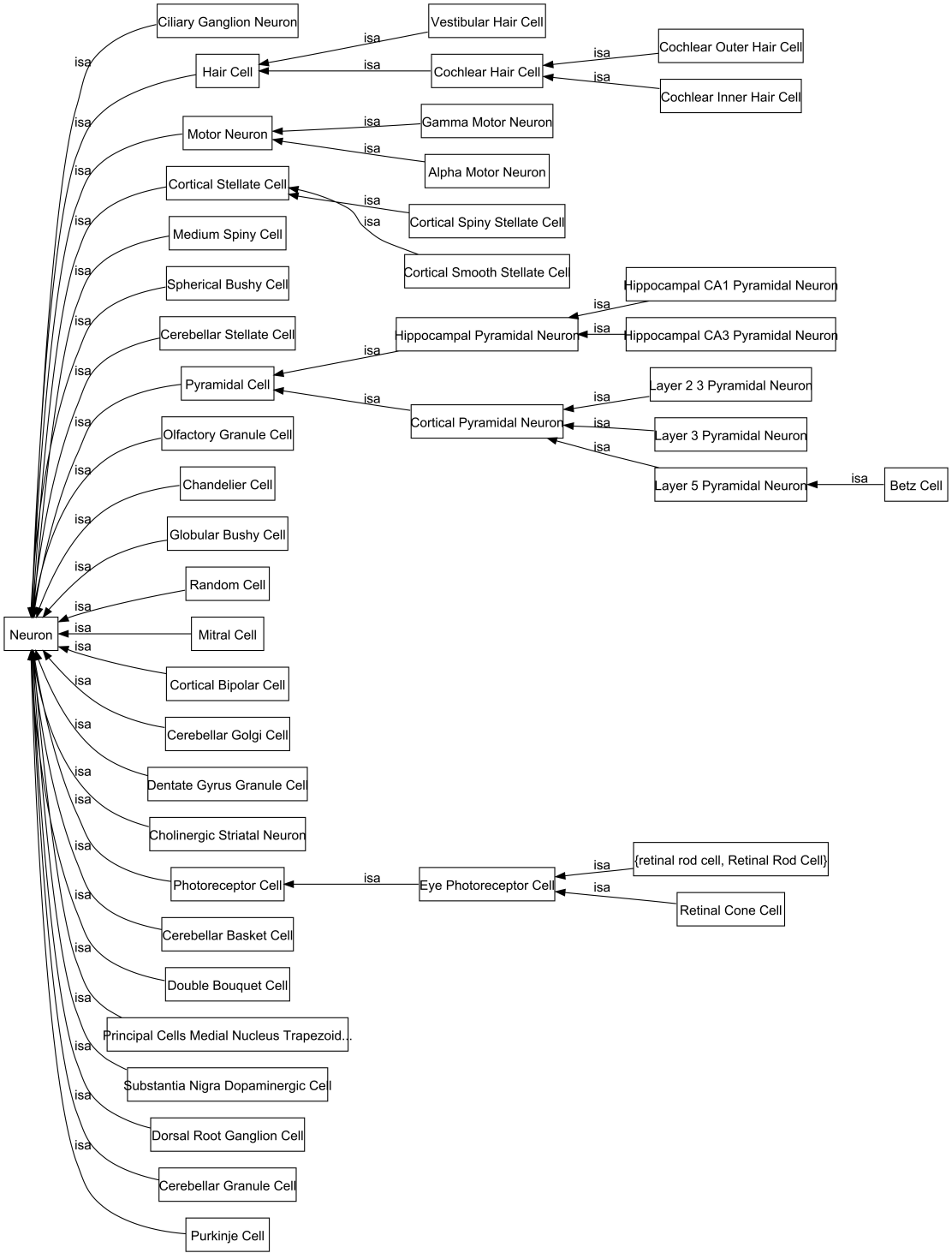

Figure 2.4: Neuron Hierarchy.

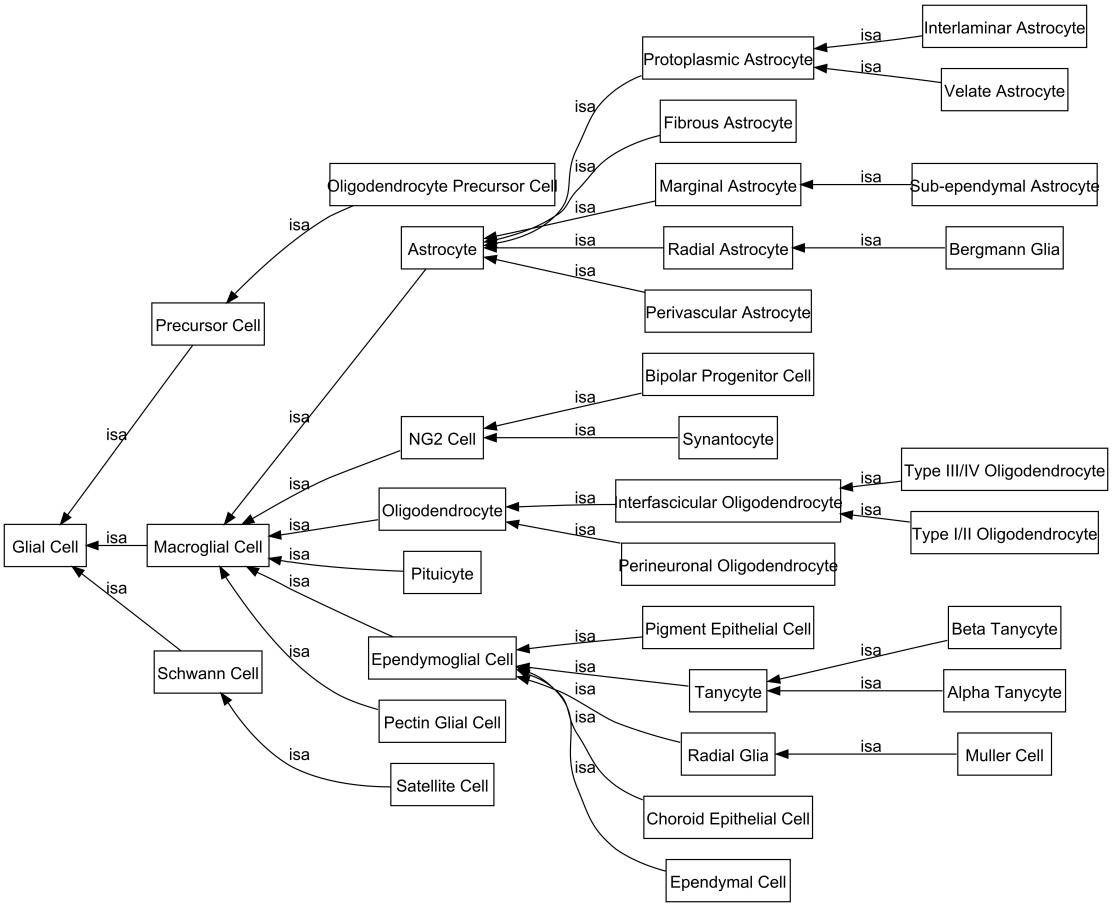

Figure 2.5: Glial Hierarchy.

Cell

We have included a set of cell types found in the nervous system (Figure 2.3) that include neurons and glial cells, as well as other classes of cells that one would encounter in structural studies of the nervous system, for example, vascular cells, endothelial cells, muscle cells, and macrophages. The class Nerve Cell contains neurons and glia, that is, cells that are derived from the neuroepithelium. We also include neuronal stem cell under this category. The SAO lists neurons (Figure 2.4) according to common names reflecting a mixture of classification criteria, for example, morphology (pyramidal neuron), proper names (Purkinje neuron). The SAO utilizes these names merely as labels that were assigned to cells and does not further classify cell types into subtrees based on these names, except in

instances where the hierarchy is fairly straightforward, for example, layer 3 cortical pyramidal neuron *is a* cortical pyramidal neuron. The name chosen is meant to have meaning to a neuroscientist and not express the importance of a particular criterion for classification. In other words, we chose the label layer 3 cortical pyramidal neuron because we believe that there is a class of cell defined by a set of properties, not because we think its location in layer 3 is its defining characteristic. We deliberately chose to keep the cell classification flat because the SAO can be used to classify neurons along multiple dimensions according to their specific properties (see Subsection User- Defined Reclassification and Query). Rather, we have focused on providing a comprehensive model of subcellular parts and how these parts relate to the parent cell. As we discuss in a later section, we utilize the relationships between cell parts and features to infer hierarchies as they are required. The SAO organizes glial cell types (Figure 2.5) from a morphological perspective rather than from a strict lineage perspective. Macroglial cells include astrocytes, ependymoglial cells, oligodendrocytes, and NG2 cells, according to classifications outlined in recent literature, for example, Reichenbach and Wolberg (2005). The reference from which a particular entity was drawn is included as an annotation property for that entity. The SAO does not aim to provide a comprehensive list of nerve cells as this domain is covered in other resources, for example, BAMS (Bota et al., 2005) and the Cell Type Ontology (Bard et al., 2005). Because the SAO is meant to be applied to data, we anticipate that users will add cell types from these resources to the SAO as they are encountered.

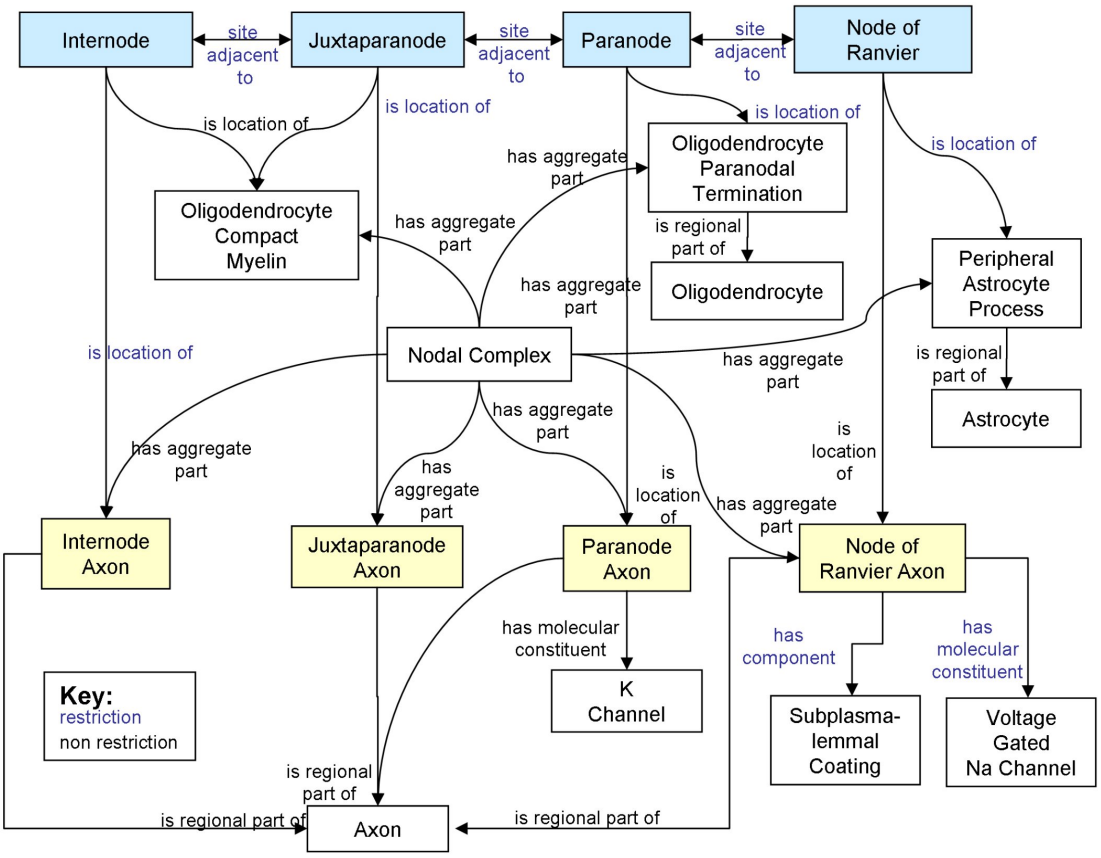

Figure 2.6: Diagram of a Node of Ranvier instance description in the SAO. The boxes indicate instances of classes that are related to one another as a description of a particular instance of a Node of Ranvier. The blue text indicates relationships that are enforced between classes through the use of OWL restrictions, while the black text indicates relationships defined for this instance alone.

Part of cell

The SAO comprises two main classes of cell parts, following the structure of the FMA: regional part and component part. Regional part of cell is elaborated under the BFO concept Fiat Object Part. A fiat object part is a part of an object that possesses at least one boundary where there is no obvious physical discontinuity or landmark structure. For example, the transition between a dendrite and the cell soma has no clear boundary. Regional parts of neurons include processes, such as dendrites and axons, the cell soma and protrusions such as dendritic spines.

Regional part of glia include the cell soma and glial processes such as astrocytic endfeet and myelinating processes. Each of these regional parts may in turn be further subdivided into finer parcellations. For example, dendrites are divided into trunk, that is, the primary dendrite emanating from the cell somata, branches, and terminal specializations. Component parts are considered to be independent objects and represent the building blocks common to all cells, for example, plasma membrane, mitochondrion. Components are largely drawn from the Gene Ontology cell component hierarchy (Gene and Consortium, 2001), with additional neuron-specific parts such as post-synaptic density added when necessary.

Molecules

Macromolecules are also elaborated within SAO under the independent continuant class. Just as with cell types, the SAO does not contain an exhaustive list of macromolecules, because we anticipate that these entities are covered in other resources. As molecules are encountered in biological data, they may be added to the SAO. Because the SAO is designed for annotation of data, we include separate entities for the RNA, DNA, and protein forms of a molecular entity. In this way, users can capture the target of a labeling study according to the molecular species localized and assign the species to the correct subcellular compartment. Properties. We have devised three major groups of properties in the SAO: part of, morphological and spatial relationships, again largely following the model of the FMA. Regional parts are assigned to each cell class using restrictions, for example, neurons may only have neuronal regional parts. The geometrical relationships among cell parts are specified by relationships such as continuous with, for example, dendrites are continuous with the cell somata; dendritic spines are continuous with dendrites. Thus, each regional part is assumed to belong to a parent cell. Although some properties are assigned at the level of cell class, for example, morphological type, most are assigned at the level of cell part. In this way, cell components and macromolecules are assigned to the particular part of the nerve cell in which they are found. Similarly, because nerve cells are large and may span many brain regions, the property has anatomical location, designed to situate the cell

within a regional part of the nervous system, is assigned separately to each part of the cell. The SAO thus differs from most anatomical ontologies, for example, BAMS (Bota et al., 2005) where anatomical location is assigned at the level of cell class.

We have employed restrictions within OWL to associate regional parts with the appropriate cell class. Thus, a neuron may only have regional parts of a neuron; an astrocyte may only have regional parts of an astrocyte. In contrast, component parts may be found in any cell. Although certain neuronal classes are distinguished by features such as a characteristic number of dendrites, the presence of spines or a myelinated axon, we have largely avoided creating many restrictions along these lines. Unlike gross anatomy, we usually have very few examples of a given class from which to infer these types of rules and there tends to be considerable variation within and across species of these parameters. We therefore have chosen to create a fairly generic model of a neuron in the SAO which can be used to describe individual instances of neuronal cell classes in a standard way. The SAO places molecules within their cellular contexts through the has molecular constituent property and its inverse is molecular constituent of. This property is defined as a special type of has part. Most of these molecules will be localized using techniques such as immunocytochemistry and in situ hybridization. Molecules may be assigned to any aspect of the cell, both regional and component parts, and at whatever level of granularity can be determined from the technique. An exception to this rule is the assignment of neurotransmitter. Because neurotransmitter has traditionally been one of the defining properties of a neuron to most neuroscientists, we included the property has neurotransmitter as a special type of has molecular constituent and assigned it at the level of cell class.

In theory, we should be able to derive the neurotransmitter from a consideration of the types of molecules located within the synaptic region, but because techniques such as immunocytochemistry often determine neurotransmitter indirectly, for example, through the localization of a synthetic or degradative enzyme for a neurotransmitter, and because determination of a neurotransmitter usually involves additional physiological or pharmacological criteria, we decided to assign

this as a simple property for now. Through the properties has anatomical location, the SAO situates cells and parts of cells into higher order brain regions. The SAO divides anatomical localization into three categories: has general anatomical location; has specific anatomical location; has atlas location. General anatomical location is assigned to the level of the cell class and is meant to encode the generally known location of a cell class. This property again was included for expediency, because neuroscientists are so used to naming individual cells as parts of anatomical regions, even though only the cell soma may be located there. The level of specification may be fairly coarse in this case, for example, Purkinje cell has general anatomical location cerebellar cortex. Specific anatomical location is meant to be assigned at the instance level and is intended to be assigned at as fine a level of granularity as possible, for example, my Purkinje cell dendrite has specific anatomical location outer third of cerebellar molecular layer. If known, anatomical location can be recorded as a set of atlas coordinates through the has atlas anatomical location property. This property type contains the atlas referenced, the coordinates, and the reference point from which the coordinates are derived, for example, bregma. Currently, the SAO assigns anatomical location in the form of free text.We are in the process of changing the anatomical location to an object property that is drawn from the BIRNLex anatomical ontology, which in turn draws its anatomical entities largely from the Neuronames hierarchy (Bowden and Dubach, 2003).

Supracellular structures

One of the biggest challenges in constructing the SAO was to provide the specification of supracellular entities like the Node of Ranvier and the synapse. Although these entities are treated by other ontologies (e.g., Zhang et al., 2007) as if they are independent entities, in fact neither of these objects exist independently within complex tissue. Rather, they represent sites where certain configurations of subcellular objects are found (e.g., neuropil, synapses, glomeruli, and the Node of Ranvier) and where certain functions are presumed to occur. Thus, although in preliminary versions of the SAO, we classified synapses and Nodes as objects, start-

ing in v1.0 we utilized the structure of the BFO to classify supracellular domains through the object aggregate and site classes.

An object aggregate in BFO 1.0 is defined as an independent continuant that is a mereological sum of separate objects and possesses non-connected boundaries. Examples: a heap of stones, a group of commuters on the subway, a collection of random bacteria, a flock of geese, the patients in a hospital. A site is defined as an independent continuant consisting of a characteristic spatial shape in relation to some arrangement of other continuants and of the medium which is enclosed in whole or in part by this characteristic spatial shape. Sites are entities that can be occupied by other continuants. The BFO further clarifies sites in this way: In BFO, site allows for a so-called relational view of space which is different from the view corresponding to the class spatial region. Space and spatial region entities are entities in their own rights which exist independently of any entities which can be located at them. This view of space is sometimes called absolutist or the container view. In BFO, the class site allows for a so-called relational view of space, that is to say, a view according to which spatiality is a matter of relative location between entities and not a matter of being tied to space. The bridge between these two views is secured through the fact that while instances of site are not spatial region entities, they are nevertheless spatial entities. (BFO 1.1; `http://www.ifomis.org/bfo/1.1`).

We considered supracellular domains as object aggregates because they represent a somewhat ad hoc grouping of cell parts into a higher order structures. However, many of these ad hoc groupings are given special designations because they are believed to be the locations at which a particular function occurs. For example, the Node of Ranvier is the site of action potential propagation down the axon; the synapse is the site at which neurotransmission occurs. The location of that function is inferred because of the presence of one of more molecules or cell components that have been demonstrated to be involved in the expression of these dynamic processes. Figure 2.6 shows the SAO structure for describing the Node of Ranvier from the central nervous system. We define the Node of Ranvier as a site on the axon in the gap between two segments of myelin. Neuroscientists have

identified different compartments of the node based on the locations of certain structural configurations and molecules such as ion channels. We thus constructed a set of entities, grouped under Node Related Sites, utilizing the parcellation described in Sosinsky et al. (2005) to describe the different sites, the cellular objects located at each site and the spatial relationships among them. Note the difference between Internode (transitively a subclass of Site) and Internode Axon (transitively a subclass of FiatObjectPart). Internode is not the parent class of Internode Axon, because they refer to distinct entities in the axon. The distinction between the two reflects the difference between material and location. If we were to ask what is the material located at the Internode site? the answer would be not only the Internode axon, but would also include compact myelin, protein channels and other macromolecules.

Conversely, if we were to ask where is the Internode Axon? in the sense of asking where the material substance of this regional part of an axon is located, the answer would be, at the site called the Internode. Similarly, asking where is there both compact myelin and a regional part of an axon? would also give the answer, at the site called the Internode. In this way, the SAO can provide a very precise specification of the different macromolecules and provides a formal basis for creating rules by which a structure can be recognized.

The synapse is modeled using the object aggregate and site classes (Figure 2.7). We created an aggregate object consisting of a pre-synaptic part, a post-synaptic part, and a junctional part, similar to the Synapse Ontology of Zhang et al. (2007) and then localize them to the synaptic site. Each of these parts have cell components, for example, synaptic vesicles, located within them that define the extents of these parts, that is, the pre-synaptic part is the part of the presynaptic structure (axon terminal, dendrite, or soma) containing synaptic vesicles. In our earlier versions of the SAO, which classified the synapse as a single material entity rather than a site, we encountered the problem that our designation of cellular structures as pre- or post-synaptic provided no way to distinguish the part that participated in the synaptic contact from the whole structure. When we say that the neuron soma is the post-synaptic structure, we are usually saying is that

there is a contact on a part of the cell body. Through the relationships encoded in the SAO, we can restrict the definition of the synapse to that part of the cellular structure where certain structures, for example, synaptic vesicles, or molecules are localized.

Anatomical qualities

Version 1.2 of the SAO has included a more extensive list of morphological qualities under the dependent continuant class that are used to modify objects within the SAO. Generic morphological qualifiers such as round or spherical are imported into SAO through the Phenotype and Trait Ontology (Gkoutos et al., 2005). However, we included a set of qualities that were specific for subcellular anatomy, for example, spine shapes (mushroom, thin, stubby), nuclear shape (round, lobulated, indented), and cell soma shape (pyramidal, fusiform). We elected in most cases not to precoordinate these terms with the independent continuants they describe, because these qualities can be assigned at the time of annotation. By pre coordination, we mean the creation of a set of independent continuants which incorporate the qualifier, for example, mushroom-shaped spine; lobulated nucleus. Precoordination was used for morphological classes that required unique identification like spine classes, where the designation of mushroom shape confers a set of unique properties to that class. We chose not to precoordinate when the qualifier was considered descriptive of an instance and not necessarily indicative of a member of a distinct class. In these cases, we apply the qualifier to the instance, for example, instance of nucleus with morphological quality indented at the time of annotation. In this way, we do not have to generate large numbers of classes that differ on what might be a superficial detail. Additional qualities that are assigned to each object are morphometric quantities such as length, surface area, etc., orientation, and polarity.

Annotation properties

Annotation properties contain information about the ontology entities. We imported the annotation properties from the BIRNLex, a lexicon developed for the

Biomedical Informatics Research Network (BIRN) project (www.nbirn.net/birnlex). These properties cover lexical entities such as definitions, synonyms, alternative spellings, and the curation status of each entity. The label assigned to the class name is also an annotation property. The BIRNLex, in turn, imported many entities from the Simple Knowledge Organization System (SKOS; `http://www.w3.org/2004/02/skos/`), a set of RDF properties and classes for describing the entities in a knowledge resource.

The definition property provides a human-readable definition for each entity in the SAO. We believe that such definitions are critical for human annotators to reference when using ontology class terms to describe data, because the equivalence between the descriptions of objects observed in an investigation and the ontology elements provides the ontology with its semantic power. Thus, a human must clearly understand the way the term is defined in the ontology in order to apply it. Because of the somewhat artificial and complicated structure imposed on some entities (see Figures 3 and 4), the definition cannot be easily extrapolated by a human from the structure of the ontology itself. Thus, following the recommendations of the OBO Foundry, we provide a human-readable definition in the form of A is a type of B which exhibits C. A *is a* B provides the location of the entity within the class hierarchy, for example, A protoplasmic astrocyte *is a* type of astrocyte, translates easily into protoplasmic astrocyte *is a* astrocyte in the SAO. Which exhibits C provides the extensional property or properties differentiating the entity from others in a class, for example, a protoplasmic astrocyte is a type of astrocyte which is characterized by many fine processes and relatively few intermediate filaments. From this definition, the property has regional part process and has component intermediate filament may be inferred. The goal is to provide a human-readable definition that is consistent with the machine-processable definition encoded in the ontology.

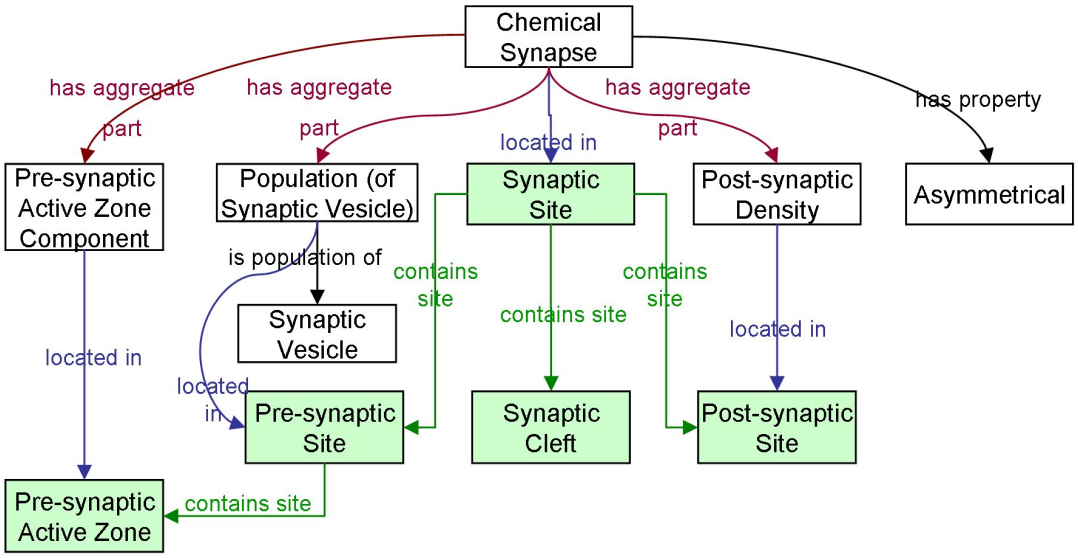

Figure 2.7: Diagram of a chemical synapse instance description in SAO. Sites are indicated by green backgrounds. The boxes indicate instances of classes that are related to one another as a description of a particular instance of a chemical synapse.

2.3.2 User-defined reclassification and query

To illustrate how properties in OWL can be used to infer additional hierarchies from the SAO, we constructed some OWL classes which reclassify the neuron cell types based on their properties assigned by the SAO. We classified neurons based on neurotransmitter, morphological type, or the presence of spines simply by defining using OWL and Protègè that these categories ought to include any cell which had the main property of that category (e.g., that the neuron was known to use glutamate or GABA as a neurotransmitter, etc). After defining these categories, we used the open source ontology reasoner Pellet (Sirin et al., 2007) to transform the flat version of the SAO neuron type hierarchy in Figure 2.8A into the inferred hierarchy in Figure 2.8B. The inferred hierarchy demonstrates that a cell like the a Medium Spiny cell is both spiny and GABAergic while a Dentate Gyrus granule cell can be classified as spiny, glutamatergic, and granule at the same time. Any arbitrary reclassification may be performed using the combinations of prop-

erties that suits the purpose of the user. Since the parent-child (is-a) relationships of the inferred hierarchy are not written back to the ontology, this allows us to maintain a hierarchy with single parents in the authored version of the ontology. However, the classes of the inferred hierarchy, Spiny Cell, Glutamatergic Neuron, Granule Cell, and GABAergic Neuron are implicitly embedded in the authored ontology as children of the class Neuron. These classes use OWL restrictions to define the kinds of children that it must logically have, and thus implicitly allows cells to exist in multiple inferred categories.

2.3.3 SAO as semantic glue

In order to use the standard names of the SAO to annotate images in different data formats, the SAO is itself used as a data exchange format between three image annotation software applications. To apply the ontology to actual data, we have incorporated annotation with the SAO into our routine segmentation tools for light and electron microscopy. We have created a programmatic interface to the OWL ontology that may be called by Jinx, our 3D segmentation tool for electron tomography data. Through Jinx, users describe the objects contained in electron microscopic volumes of neural tissue as instances of the SAO, rather than as a set of user-defined objects with no relationship among them. The application of SAO captures each object and allows the definition of related objects. Instances of the SAO are then stored in a large instance store, which we call the Cellular Knowledge Base (Fong et al., 2007), where they can be queried (Chen et al., 2006). The data files used to generate the instances are stored in the CCDB which tracks their experimental and data provenance. We are in the process of incorporating SAO into additional analysis tools for analyzing neuronal branching patterns and for annotation of spatially varying signals using our GIS-based brain atlas, the SMART Atlas (Martone et al., 2008b).

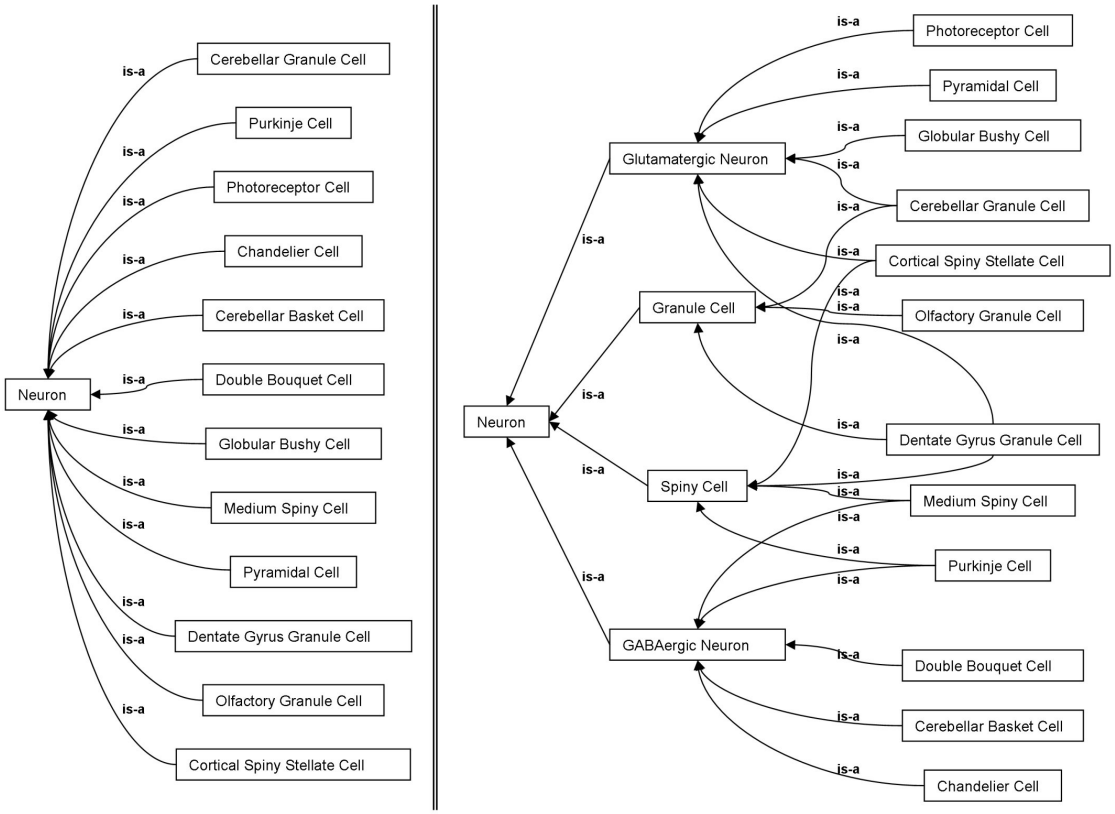

Figure 2.8: Inferred hierarchies using OWL. On the left, a subset of the hierarchy under the Neuron class prior to inference. On the right, the automatic reclassification of that subset under four user-defined groupings, Glutamatergic Neuron, GABAergic Neuron, Spiny Cell, Granule Cell, based on the properties of the cells alone.

The SAO and Cellular Knowledge Base architecture enable us to integrate these different data types through the shared semantic representation of biologically significant elements. For example, the image of a dendritic tree generated with two-photon fluorescent microscopy (Figure 2.1A), is annotated as an instance of sao:Dendritic Tree, which is part of medium spiny neuron, and has part dendrite. The instance of dendrite has regional part dendritic segment. This same instance of dendritic segment is visible in the correlated electron microscopic volume of the same medium spiny neuron (Figure 2.1C), where we can further assign has regional part Dendritic Spine to this dendrite. An algorithm with access to the SAO infer

that the dendritic spine is part of the dendritic tree, and apply properties derived from the electron tomography study to that acquired from the light microscopic imaging. Without this common interlingua and the codified knowledge explicitly declaring the shared semantic context, programmatic combination and cross query of these images and data types is much more difficult and requires customized algorithms to encode the semantic information.

By structuring the SAO in OWL, we have made its encoded knowledge available to OWL reasoners and RDF query engines. Consequently, we use instances stored in the Cellular Knowledge Base and the knowledge encoded in the ontology to determine what molecular constituents are found in the Node of Ranvier, and which sites on the Node are they respectively found in. We can also query about the glial cell types associated with the Node, and how the parts of the glial cells relate to the different parts of the Node.

2.4 Discussion

We created an OWL ontology representing the subcellular anatomy of the nervous system to provide the necessary scaffold for integrating molecular and anatomical data through accurate description of mesoscale anatomy. By codifying it in OWL, we have enabled algorithmic query and analysis of that knowledge. Moreover, we have enabled the use of formalized knowledge as a standard for making connections between data formats, making connections between other ontologies, and as a data exchange format for image annotation tools. This scaffold is amenable both to tool development and to semantically driven information exchange across the field. It also provides individual researchers a means to perform reasonerbased quality control and inferential analysis of annotated neuroimages. Applying formal semantic representation techniques to neuroanatomical structure has been preliminarily addressed in the macroscopic domain (Martin et al., 2001; Mechouche et al., 2006); little exists in the mesoscopic neuroanatomical domain as yet. A Synapse Ontology was recently constructed (Zhang et al., 2007), but it does not situate synapses in their cellular and tissue contexts, nor is it built on

top of community-shared foundational ontologies. Our motivation for creating the SAO was to provide the necessary tools for describing the types of subcellular and supracellular entities located in the dimensional range now termed the mesoscale. The SAO is designed as a reference ontology, defined by Brinkley et al. (2006) in the following way: Unlike application ontologies, reference ontologies are not designed for any specific application, but are intended to be re-used in multiple application contexts [. . .] Reference ontologies are designed according to strict ontological principles, whereas application ontologies are designed according to the viewpoint of an end-user in a particular domain. We elected to tackle the more difficult task of creating a reference ontology with formal semantics, because we believe that such resources are needed to build models of mesoscale structures that combine information from multiple domains and to be able to utilize information obtained at the mesoscale at coarser and finer scales of granularity. Through application of the ontology, researchers can work in a narrow dimensional range, but their observations are immediately linked across scales. For example, a researcher segmenting a reconstruction derived from electron tomography may make the observation that an endoplasmic reticulum of a dendritic spine from a Purkinje cells expresses the IP3 receptor. Through the SAO, the following inferences can be made: There exists a Purkinje cell dendrite that expresses the IP3 receptor; the cell class Purkinje cell expresses the IP3 receptor; the cerebellar cortex expresses the IP3 receptor; and the cerebellum expresses the IP3 receptor.

The SAO is meant to describe structure, not function nor dynamic processes, following the parcellation of biomedical reality established by the BFO. However, although we try to adhere as much as possible to this distinction within the formal class structure of the ontology, as can be seen by the labels assigned to SAO classes, many labels that are applied to our SAO entities have a functional flavor to them, for example, chemical synapse. Where possible, we tried to remove entities that mixed a structure with a function, for example, myelinating oligodendrocyte or with a physiological state, for example, activated microglia. However, we also felt in some cases that it was important to assign the labels that are commonly employed by the community. Although these labels appear in the figures

and text provided in this paper, SAO classes are actually identified using semantically neutral numeric labels (e.g., SAO class sao1507566336 has the preferred label Post-synaptic Component). The human-readable preferred label is assigned as an annotation property, as are a variety of lexical term variants, such as alternate labels, abbreviations, synonyms, acronyms, and so on. This practice is standard in the ontology community, and although it makes working with the ontology at times cumbersome for humans because of the need to associate the label with the class, we find it philosophically appealing. The entity is the same entity regardless of what we call it, that is, a rose by any other name would smell as sweet. So the fact that our neuron labels reflect mixtures of classification schemes does not impact the class structure of the SAO; rather, the class of neuron to which the label is applied is defined by the set of properties assigned to it.

Ultimately, the goal of anatomy is to provide the structural substrate for mapping function and understanding the structural constraints on dynamic processes. Anatomy is a mature discipline with a rich history. Many structures have been described, and continue to be described, particularly in electron microscopy, for which no functional property is known. The classic view of structure-function relationships assumes that structural differences reflect functional differences as well. However, mapping function onto structure is a complex issue that is currently beyond the domain of the SAO. We chose to adhere to a strict structural approach to keep the SAO scope tractable. We also, however, believe that by not mixing structural and functional classes together, it will be easier in the future to utilize the SAO within a functional ontology. As an example, the term synapse, as is recounted in all introductory textbooks, was a functional concept introduced by Sherrington to describe the transmission of information between cells. The morphological correlate of the synapse was described by Palay and colleagues using electron microscopy in the 1950s, and is also familiar to beginning students of neuroscience. SAO currently provides a formal description of the set of entities to describe the morphological correlates of what are assumed to be the sites and machinery for synaptic transmission in the nervous system. Although the labels employed, pre-synaptic and post-synaptic compartment, do have functional

significance, the precise mapping of the functional aspects onto the morphological correlate is not straightforward. Though these familiar functional labels date back to work on the cellular physiological correlate of Sherringtons synapse first described by Katz and colleagues in the 1940s, as a recent paper indicating evidence for ectopic release from the chick ciliary ganglion synapse illustrates (Coggan et al., 2005), our understanding of neural signaling at the cellular level continues to evolve. If release of neurotransmitter can occur at sites other than the active zone visualized in electron micrographs, then the functions associated with a synapse cannot be restricted to this domain. However, by modeling a synapse as a site where objects, and eventually dynamic processes, are located, the definition of a synapse can expand as our functional understanding of synaptic transmission expands. We believe that mapping of function onto structure will be one of the greatest challenges faced by those who are creating ontologies for biomedical science.

2.4.1 Reasoning and inference with OWL

Biological objects are complex entities that do not fit neatly into single hierarchies. We have chosen to follow the recommended practice of single inheritance for all SAO classes, even when that means providing a very flat hierarchy with minimal utility for classification purposes. However, the power of OWL as an ontology formalism is that it not only enables us to explicitly express the complex qualities and inter-relatedness of entities, the standard tools built around the OWL formalism allows us to automatically infer multiple valid hierarchies for an entity, depending on what is required. For complex entities such as neuronal classes, we can use the OWL inference engine to infer hierarchies based on neurotransmitter, morphological properties, anatomical location, or circuit type (Figure 2.8). The same can be done for other classes of subcellular structures, for example, dendritic spines. This approach provides maximal flexibility to the end user and allows us to begin to cluster and define neurons based on a set of properties rather than along a single dimension (Migliore and Shepherd, 2005).

We have only begun to experiment with the power of OWL to infer new

knowledge about objects that is not explicitly encoded in the ontology that allows information to be inferred across scales. In chapter 3, we provide an example of this cross scale reasoning using OWL and rules about how cell parts relate to cells and brain regions. In this example, we showed how annotation of a synapse between a terminal of a thalamocortical axon and the dendritic spine of a cortical neuron observed through axonal tracing and electron microscopy could be used to infer knowledge about regional brain connectivity. Through relationships encoded in SAO, we inferred from the presence of a labeled axon terminal that there must be a neuron in the thalamus that has an axon projecting to the cortex. From the presence of a spine, we inferred that there existed a neuron to which the spine belonged in cortex. From the local observation that an axon terminal synapsed on a dendritic spine, we could infer that thalamic cells synapse with cortical cells, and that thalamus projects to cortex. While the reasoning itself does not provide new insight about brain function, we showhere that a computational algorithm was able to infer the same logical cross-scale consequences of the subcellular arrangement of cell parts as would a neuroscientist without our having to write custom code to embed that knowledge in the program.

2.4.2 Application of the ontology

In construction of the SAO, we have attempted to provide a formal structure for describing data, balancing the needs for a top-down versus a bottom-up approach. By top-down, we mean that the biological theory governing a domain is used to classify data products; by bottom-up, we mean that we do not impose prior knowledge constraints on interpreting data but let the data speak for themselves (Murphy, 2005). OWL classes are essentially descriptive templates that constrain the possible properties and relationships which instances may have. As such, we only encode knowledge into the class level when we are sure that it ought to constrain all further instances that may be seen. This criterion enforces a certain amount of rigor when describing the properties of biological entities. What are those things that must always be true of a biological entity? Unlike the case of gross anatomy, where we can be reasonably certain of the canonical form taken

by the human body, for example, we do not believe that we are at the stage with subcellular anatomy where we can comfortably define such canonical forms. Thus, although we sacrifice some of the reasoning power of OWL through the minimal placement of restrictions on the classes, we designed version 1.2 of the SAO to serve as the basis by which such rules can be derived from the instances.

When describing data, we apply the ontology only down to the level of granularity of which we are reasonably certain. For example, if we know the type of neuron we are describing, we can assign instances of properties to that specific class; if we do not, we can assign the observed properties to the class neuron. Using the reasoning power of OWL, it may turn out that the properties of this unidentified neuron are equivalent to a known class, but that can be inferred from the actual instance. In this way, the structure of the OWL standard forces the SAO to make careful and conservative descriptions about subcellular anatomy while still allowing a place for uncertainty.

Instances within the SAO also serve another important function by allowing us to annotate the biological description of a piece of data with the data and experimental properties from which it was derived. Entities within SAO are not directly observable by humans but must be imaged through a device such as a microscope and recorded in some form on a particular medium. Biologists are well aware that how a specimen was prepared, imaged, and analyzed will impact the types of observations that are made. In many cases, subcellular structures that are observed under certain conditions, for example, chemical fixation, are determined to be artifactual when recorded under different conditions. Most experimentalists are uncomfortable with knowledge management systems that attempt to divorce the biological reality from the methods used for acquisition, visualization, and analysis, because these methods largely determine the form that the reality will take. We must recognize, however, that the entities that we are attempting to describe in the SAO are assumed to transcend any technique. That is, we are assuming that there is such as thing as a dendrite, even though its properties can only be described in a specific experimental context. So, although the SAO itself does not assign technique or data type to the biological entity, for each instance of

the entity, we provide a link to the experimental evidentiary context and the data type from which it was derived (e.g., this instance of dendrite was stained with a Golgi stain and imaged in a light microscope).

Through the construction of the SAO, we have made progress toward the goals of building information bridges in neuroscience in three broad areas: formalization, externalization, and standardization. By formalization, we mean the process of describing concepts in a fully explicit manner in order to clarify and sharpen the meanings of the terms being used. The lengths that we have gone to either find or impose structure on implicit concepts in subcellular anatomy reflect the absence of prior efforts to bring them into a single cohesive framework. Such a framework is important for the growing community interested in producing detailed computational models of structure and function in the nervous system. It is vitally important that experimental neuroscientists be able to communicate with this community and provide increased levels of explanation of their experimental systems. Providing a formal way of communicating, these explanations make it much easier to begin the modeling process. Ontologies in general, and the SAO in particular, is crucial connective tissue to help place these goals within reach for neuroscience.

In order for formalized information to be used by software applications, the information must be capable of externalization. By externalization, we mean to draw attention to the ability to transform the information into code, as opposed to the translation of abstract concepts into a human-only readable explicit representation. Once knowledge has been formalized and subsequently codified into a computer-readable form, that knowledge becomes externalized as an entity that is capable to programmatically interact with other knowledge. This makes information much more flexible than if it resided on the printed page, and it allows algorithms to answer questions for us, saving time and effort. The process of constructing an OWL ontology formalizes the knowledge it contains, but encoding it in OWL and saving it on a computer in its underlying RDF/XML format externalizes the information for other systems to digest and manipulate via standard open source code frameworks.

Through externalization, we are able to remix knowledge into other forms. It allows us to generate diagrams, to view it in different software interfaces (e.g., Jinx), to reclassify hierarchies on demand, and to run rule-based reasoning or other automatic inferencing mechanisms. The benefits of this are obvious in the context of the goals of data sharing and model construction. Externalization is also needed in order to construct algorithms that are capable of assisting neuroscientists do their own work, such as to guide them in a literature search or to suggest the name of a structure they are segmenting.

Once an information bridge has been formalized, and also externalized, it can be used for the final important purpose of standardization. In this context, the aspect of standardization that we focus on is the ability for OWL ontologies to serve as semantic glue which allow disparate data, ontologies, and applications to interoperate. The strategy we have employed in our knowledge environment is to leverage the externalized knowledge in the SAO by embedding it in tools that have first contact with primary data. By embedding the SAO in these tools, we enable the user not only easy access to SAO terms to use in annotating their data, but also we make the tools more intelligent to minimize the amount of implied knowledge that a user must contribute.

Chapter 2, in full, is a reprint of the material as it appears in Frontiers in Neuroinformatics 2007, 1:3. Larson, Stephen D.; Fong, Lisa L.; Gupta, Amarnath; Condit, Christopher; Martone, Maryann E. The dissertation author was the primary investigator and author of this paper.

Chapter 3

Rule-based reasoning with a multi-scale neuroanatomical ontology

3.1 Introduction

Neuroscience data are complex. Whereas the field of bioinformatics has relatively straightforward data structures like lists of gene and protein sequences, the field of neuroscience has to manage data about anatomy, physiology, behavior and more. At least one reason for the complexity and expansiveness of the domain is that neuroscientists are still unsure of the most significant biological factors underlying the brain's ability to plan, coordinate and execute actions based on external and internal information. Consequently, neuroscientists are engaged in activities that span multiple temporal and spatial scales simultaneously. As the cell remains the major structural unit of biological tissues, clearly the neuron plays an important role in nervous system function, but in order to fully understand the neuron, many details about its cell biology and molecular dynamics must be known. Similarly, much needs to be known about the role that a neuron plays in the larger networks that they form. Capturing information at the spatial scale of large-scale networks, small-scale networks, single cells, and molecules is therefore required

to recognize how events across these scales come together in the processes that span them. All of these issues are equally valid in the temporal scale, as important events happen at scales of microseconds (ion channel dynamics), milliseconds (synaptic events), seconds (long-term synaptic events) and across animal lifetimes. To complicate matters further, data collected at each scale is recorded using diverse instrumentation and have no inherent means of being unified with other forms of data. As a result, the problem of neuroscientific data can easily be called the most complex data management problem yet seen by science. The significant consequences of this are twofold: 1) the complexity of the data leads to a reduced ability to share data; 2) the reduced ability to share data harms efforts to produce a synthetic understanding of the brain(Edi, 2006). A synthetic understanding of neuroscientific data should unify across physical and temporal scales and across data sources.

This study examines ways in which neuroscientific data can be synthesized through the use of information technologies. Previously we have shown how to take advantage of the W3C's OWL specification2 to construct an ontology of subcellular anatomy (Fong et al., submitted). After explaining the significance of this format for describing data, we go into more detail on the nature of the multiple scales in our model and discuss how it was aggregated from multiple sources. Then we demonstrate how logic-based rules can generate new knowledge from multiscale data. Multiscale integration required that we merge multiple independent ontologies covering different spatial scales. Examples of inferences made on the data are explored in detail.

3.2 OWL Format Enables Ontologies To Be True Data Models

The W3C's OWL 1.0 specification2 is an extension of RDF3, which is itself an extension of XML4. As reported in Martone et al. (2004), this format is ideal for the construction of biomedical ontologies designed to bridge the gaps between different scales and modalities in neuroscience. In a companion paper

to this submission (Fong et al., 2007), we describe the creation of an ontology for subcellular anatomy of the nervous system (SAO; `http://ccdb.ucsd.edu/SAO/1.0/SAO.owl`). Unlike most ontologies developed around gross anatomy, e.g., FMA (Martin et al., 2001) the SAO takes a cell-centered view. This consists of developing a model of the nerve cell that encapsulates its cell parts and their relationships both within and across individual cells. The SAO then combines these subcellular elements with those from molecular and gross anatomical scales. In the SAO, a neuron has a range of associated properties that help it unify information across scales and across experimental modalities. A list of some of the most important properties appears in table 3.1. Many of the relationships between elements in the data are explicitly described, such as in table 3.1, but many other relationships that could result from description in the SAO are implicit. For example, the observation that a synapse exists in a micrograph means that two neurons are structurally connected. If these neurons have their origins in different brain nuclei, than these nuclei have a structural connection between them. Even in the absence of an explicit statement of connection between brain areas, this information is there implicitly. Not only would we like to make this information explicit, but ideally a program would extract these relationships for us. In order to achieve this, we employed rule-based reasoning.

The key properties relating parts of cells to the cell class are has compartment, i.e., regional part, and has component, e.g., cellular components. In the SAO, properties can be assigned to parts of cells as well as the cell class itself. Table 3.2 shows an outline of entities that are valid for the range of the property "has_Compartment". These entities in turn have compartments and other properties that can be assigned to them, and so on. As a result, the SAO is intended to be used to compose a description of neuroanatomical entities from its parts as one might construct a structure out of tinkertoys. For example, an axon terminal named Axon_Terminal_1 has the following properties: 1) its origin is the posterior complex of the thalamus, 2) glutamate is an associated molecular constituent, 3) its location is in the primary somatosensory area's

Table 3.1: Select Properties of Neuron in SAO.

has_Neurotransmitter
is_Spiny
number_Of_Axons
number_Primary_Dendrites_Min
projects_To
has_Compartment
has_Component
has_Dimension
has_Inherent_3D_Shape
anatomical_Location_Atlas
anatomical_Location_Specific
functional_State
location_CNS_PNS
morphological_Type
is_continuous_with
observation_Conditions
has_parent_cell
species_of_origin

barrel field, 4) it is a subcompartment of an axon named Axon_1, and 5) it is found in a rat. Its origin, Posterior_complex_of_the_thalamus, and its location Primary_somatosensory_area_barrel_field are terms that reference entities in the Brain Architecture Management System (BAMS; Bota et al., 2005) a comprehensive terminological resource for gross brain anatomy that includes both anatomical regions and connectivity between brain regions. Its other properties are defined in the SAO. The SAO assigns anatomical location to different parts of nerve cells, reflecting the fact that nerve cells may have very long processes that project through and to multiple brain regions. We detail below how we have incorporated data from BAMS with the subcellular data in the SAO.

Table 3.2: Valid Neuron Compartments.

- Axon
  - Axon_Collateral
  - Main_Axon
- Cell_Body
- Dendrite
  - Dendritic_Branch
  - Dendritic_Segment
  - Dendritic_Subtree
  - Dendritic_Tree
- Spine
  - Axonal_Spine
  - Dendritic_Spine
  - Somatic_Spine

3.3 Disparate Data Can Be Interfaced Once Converted To OWL

In order to make inferences across physical scales, data about each scale must be integrated. To define the location of cell parts within regional brain parts, we utilized the brain part hierarchy of the Brain Architecture Management System (BAMS). To collect data on the level of brain anatomy and connectivity between brain regions, we turned to the BAMS website (`http://brancusi.usc.edu/bkms/`). A Perl script using the CPAN libraries WWW:Mechanize and HTML::TokeParser was written to read successive web pages and encode anatomical, connectivity, and reference data into the OWL format. Approximately 720

brain parts and 14,000 connectivity statements were taken from the Swanson-98 atlas of the rat brain. The result of this encoding is referred to as the BAMS ontology.

The BAMS ontology has only two major categories - BrainPart and ConnectionStatement. BrainPart gathers information about the name of the structure, the type of the structure (grey matter, white matter), and its place in the containment hierarchy defined by the brain atlas. ConnectionStatement is written anytime the web page of a brain structure indicated a reference that demonstrated a projection entering or leaving this brain area. The strength of the connection is noted. The paper reference and link to the PubMed webpage is also noted. An important point to note here is that this ontology is constructed via an algorithm, and that it can also be updated in the future by the same algorithm. In this way, information that is added to the BAMS website can be dynamically included in the ontology reasoning system.

After creating the BAMS ontology, there was a need to create an ontology that expanded on its terms. The expansion was necessary because several levels of structural granularity were missing between brain structures represented in BAMS and the cells that compose them. To close this conceptual gap, we created the BAMS+ ontology, which uses the OWL import functionality to import the BAMS ontology. BAMS+ adds more fine grained parcellations of brain regions, e.g., cytoarchitectural regions such as cortical layers. Although for the reasoning work demonstrated in this paper we did not take advantage of these constructs for simplicity, this ontology will play an important role in more sophisticated multi-scale reasoning tasks involving additional anatomical scales.

The SAO-BAMS+ ontology uses OWL imports to import both the SAO and BAMS+ ontologies. This merged ontology allows for statements to be created that involve both ontologies. Important relations are defined here such as has_Origin_In_BAMS_Location, has_BAMS_Location, has_Post-synaptic_Neuron, and has_Pre-synaptic_Neuron. The instances that are created in the examples presented and the rules that are presented are all run with this ontology loaded into Protege.

Figure 3.1 illustrates the structure of imports for the SAO merged with BAMS.

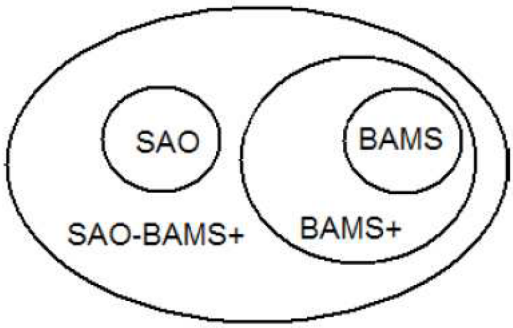

Figure 3.1: A Venn diagram of the structure of the merged ontology, labelled "SAO-BAMS+". SAO and BAMS indicate the subcellular anatomy ontology and the Brain Architecture Management System ontology generated from our Perl script. The BAMS ontology is imported within the BAMS+ ontology, which extends its semantics. The SAO-BAMS+ ontology imports both SAO and BAMS+ to enable elements of both to be used together.

3.4 Rule-based Reasoning Allows Inferences To Generate New Knowledge

In the following, we show how we utilized the SAO-BAMS+ ontology and rule based reasoning to bridge the gap between the cell-centered ultrastructural view of the SAO and the regional connectivity view of the BAMS.

Logic-based rules are IF-THEN statements that are machine readable. They can be constructed to make inferences about knowledge in an ontology (Golbreich et al., 2005). In the case of the SAO, we are interested in making inferences about the kind of structures visible in electron micrographs: parts of cells, macromolecules and supracellular aggregates. The latter class includes structures that span two or more cells, e.g., synapses. The SAO specifies the relationships among these parts, e.g., neuron has compartment dendrite; dendrite is continuous with

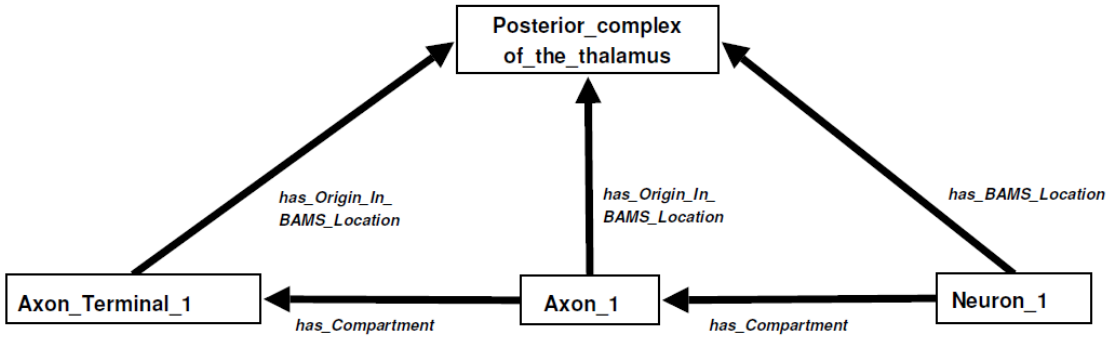

Figure 3.2: An illustration of how reasoning enables the creation of new information from old information through logical inference.

Table 3.3: Rule 1.1

```
((:instance "sao:Axon_Terminal" ?x)
 (has_Origin_In_BAMS_Location ?x ?y)
 (:add-instance (?a "sao:Axon")(:name ?a "")
   (has_Origin_In_BAMS_Location ?a ?y)
   ("sao:has_Compartment" ?a ?x)))
```

the cell soma. Almost all electron microscopic data involves the analysis of partial structures, i.e., isolated parts of dendrites, axons, cell bodies. Only rarely is a complete cell including processes examined at the ultrastructural level. The SAO was constructed so that parts of neurons that might be observed in an electron micrograph can be placed in their cellular and tissue contexts. For example, an axon terminal is a part of an axon that participates in a synaptic connection with a part of another cell. Figure 3.2 illustrates how having some knowledge about where the axon terminal is coming from can be leveraged to infer new information.

Commonsense and experience from observation implies that where there is an axon terminal there must also be an axon. Encoding this experience into a rule that acts on our data model enables the augmentation of knowledge (Fig. 3.3). Here we demonstrate schematically the result of applying Algernon7 rules 1.1 and 1.2

Table 3.4: Rule 1.2

```
((:instance "sao:Axon" ?x)
 (has_Origin_In_BAMS_Location ?x ?y)
 (:add-instance (?a "sao:Neuron")(:name ?a "")
   (has_BAMS_Location ?a ?y)
   ("sao:has_Compartment" ?a ?x)))
```

to our knowledge base. At first, the knowledge base begins with Axon_Terminal_1 known to have the relation has_Origin_In_BAMS_Location to the Posterior_complex _of_the_thalamus. Applying Algernon rule 1.1, we generate an instance Axon_1 , with the property that it shares the same origin as the axon terminal, and that it has the axon terminal as a compartment of itself. Axon_1 may correspond to a visible feature of the image, or it may be outside of its boundaries and thus only exist as an inferred entity. We can use the same logic in Algernon rule 1.2 to infer the presence of a neuron, Neuron_1 (actually the neuron cell soma) whose location, has_BAMS_Location is the origin of Axon_1 and Axon_Terminal_1. This kind of logical inference can be run on all instances of axon terminals present in our data set and build up a corpus of inferred axons and neurons that can then be used for further analysis.

Figure 3.3 illustrates a second example of reasoning that merges across three scales of data. Prior to reasoning, Synapse_1 has two intercellular junction compartments, Pre-synaptic_compartment_1 and Post-synaptic_compartment_1. Applying Algernon rules 2.1 and 2.2, it is discovered that the pre-synaptic compartment is situated in Axon_Terminal_1 and our post-synaptic compartment is situated in Dendritic_Spine_1. Using the rules illustrated in figure 2 to infer the presence of neurons from axon terminals, and an additional set of rules for dendritic spines, the synapse can be directly associated with the two neurons that participate in that synapse, through the properties Pre-synaptic_Neuron and Post-synaptic_Neuron. If two neurons share a synapse, then there is a connection between those neurons. If those neurons are in different brain areas, then those areas have a connection between them. Since the neurons that participated in Synapse_1 have already been identified, and their locations are known through

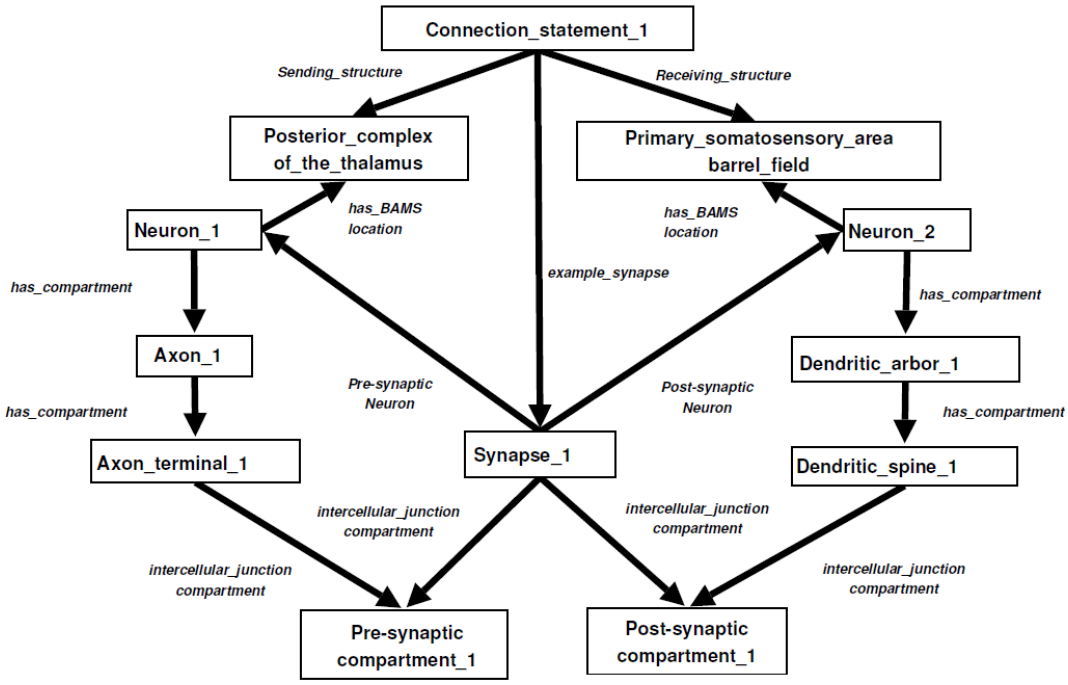

Figure 3.3: Reasoning across three levels of anatomical scale, merging the cellular, supracellular and gross anatomy.

has_BAMS_location relations, Algernon rule 2.3 can make an explicit connection statement about the brain areas in which the neurons are located. Through this reasoning, we can infer statements about connectivity between gross brain regions from local interactions that happen on the subcellular scale captured by electron microscopy.

3.5 Future Directions

We forsee two immediate directions for this project. First, the merging of ontologies across scales will continue. The current version of the BAMS+ ontology needs expansion to capture the existing concepts of neural systems that lie between the cellular level and the gross anatomical level. Additional synthesis of neuroanatomy into formal entities in this mesoscale will aid understanding greatly.

Table 3.5: Rule 2.1

```
((:instance "sao:Chemical_Synapse" ?a)
 (:instance "sao:Post-synaptic_Compartment" ?b)
 ("sao:intercellular_Junction_Compartment" ?a ?b)
 (:instance "sao:Neuron_Compartment" ?c)
 ("sao:is_Intracellular_Junction_Compartment_Of" ?b ?c)
 (:instance "sao:Neuron" ?d)
 ("sao:is_Compartment_Of" ?c ?d)
 (has_Post-synaptic_Neuron ?a ?d))
```

Table 3.6: Rule 2.2

```
((:instance "sao:Chemical_Synapse" ?a)
 (:instance "sao:Pre-synaptic_Compartment" ?b)
 ("sao:intercellular_Junction_Compartment" ?a ?b)
 (:instance "sao:Axon_Terminal" ?c)
 ("sao:is_Intracellular_Junction_Compartment_Of" ?b ?c)
 (:instance "sao:Axon" ?d)
 ("sao:has_Compartment" ?d ?c)
 (:instance "sao:Neuron" ?e)
 ("sao:has_Compartment" ?e ?d)
 (has_Pre-synaptic_Neuron ?a ?e))
```

Table 3.7: Rule 2.3

```
((:instance "sao:Chemical_Synapse" ?a)
 (has_Pre-synaptic_Neuron ?a ?b)
 (has_Post-synaptic_Neuron ?a ?c)
 (has_BAMS_Location ?b ?d)
 (has_BAMS_Location ?c ?e)
 (:add-instance (?f "bams:Connection_Statement")
     (:name ?f "") ("bams:reference" ?f "Inferred")
     ("bams:sending_Structure" ?f ?d)
     ("bams:receiving_Structure" ?f ?e)
     (example_Synapse ?f ?a)))
```

The main challenge of this effort will be careful knowledge engineering of these entities that creates definitions that are not too general but also not too specific.

The second important direction is to forge a tighter coupling between these ontologies and spatial and structural representations. While our team has made excellent progress creating systems that interoperate with such representations (see Fong et al., 2007), the next challenge will be to fuse them together into a seamless whole. This will necessitate a system of "reasoning" that is not just limited to inference or rule-based chaining, but which also accommodates calculations of 3D geometry and space, and can merge the results of these diverse processes together.

3.6 Contributions

In this paper, we have demonstrated that in order to perform inferences across physical scales in neuroscience, one requires data organized in a representation that enables integration. OWL-based ontologies are idea common grounds for many neuroscientific data because they allow information from different sources and on different topics to be analyzed using standard tools. We described one example of consolidating existing information on the web into the OWL format by writing a program that reads the web pages at the Brain Architecture Management System website and generates an OWL ontology. Having data from different scales in the OWL format then allowed us to describe one data set in terms of the other. Furthermore, we have shown that common sense about the way that data in neuroscience fits together can be encoded into logical rules that, through inference performed by simple rule based tools, can generate new knowledge from old knowledge. We have demonstrated examples of this by first inferring the presence of axons and neurons from axon terminals and secondly by inferring connections between brain areas by the presence of synapses.

3.7 Methods

In order to perform the rule based inferences on the data, we used the Protege-OWL editor version 3.2.1 with the Algernon plugin (Hewett), developed by Michael Hewett. It provides a simple syntax for querying and updating the data model and is capable of inserting and deleting instances. However, one disadvantage is that over large data sets it is inefficient.

Additional materials described in the paper such as ontologies and rule files are available at `http://ccdb.ucsd.edu/SAO`

Chapter 3, in full, is a reprint of the material as it appears in CEUR Workshop Proceedings 258, 2007, ISSN 1613-0073. Larson, Stephen D.; Martone, Maryann E. The dissertation author was the primary investigator and author of this paper.

Chapter 4

Neurolex.org: An online parts list for neuroscience

4.1 Introduction

Imagine that every time a neurobiologist identified a new piece of the nervous system, he could record information about it on a magic index card. This magic index card would list different features of that piece of the nervous system, and whatever was written on it would immediately be visible to the entire world. From anywhere in the world, the magic index card could be sorted into any list of nervous system parts to which it belonged. If the index card described a neuron that used glutamate to signal other neurons, then it would appear on a list of glutamatergic neurons. If the same card asserted that the neuron was found in the hippocampus, then it would also appear on a list of neurons in the hippocampus. This card could be written for any piece of the nervous system, whether it was an individual ion channel, or a whole chunk of the brain. Every time a new card was added, it would enrich a global framework for the understanding of the nervous system rather than add a disconnected puzzle piece to an ever expanding sea of data. Such a system would be an important start of a global semantic framework for neuroscience data integration, and is necessary to overcome the challenges of reaching broad understanding of a system as complex as the nervous system. In

order to tackle these challenges, it is crucial that we define the major concepts in neuroscience, e.g., neuron, brain, and the relationships among these concepts in a machine-readable form, but also make this machine-readable form accessible and easy to improve collaboratively over the internet.

There has been a great deal of work in the study of "knowledge representation" in computer science (Davis et al., 1993). Knowledge representation concerns itself with how to capture the meaning of statements in machine processable forms. Examples of statements of interest to neuroscience include "mitochondria are part of neurons" and "Purkinje cells are located in the cerebellar cortex". In these examples, "mitochondria", "Purkinje cells", neurons and cerebellar cortex are the entities, "part of" and "located in" are properties (sometimes referred to as relations or relationships),. By splitting knowledge into these atoms, computational systems can do a better job of analyzing the relationships expressed within them. This makes individual statements available for search, query, and reuse into other information systems, which is still currently difficult with unstructured prose. When done over a large knowledge base, this approach enables computational systems to keep large amounts of complex information well-organized and easily accessible, enables search across distributed databases and allows data-minded scientists to rapidly pose and answer questions about existing knowledge via automated logical deductions (Martone et al., 2004; Larson and Martone, 2009). Computer science is increasingly producing tools to find patterns in large corpuses of data that have been structured this way. However, a well structured knowledge base that has been comprehensively populated and has built up significant consensus from the neuroscience community is far from completion.

The lack of well structured knowledge in the biosciences has been acknowledged in the past and there is a significant history of effort to try and resolve it. Seeing this challenge, in 1965, the College of American Pathologists created the systematized nomenclature of pathology (SNOP), which evolved over the next forty years into the Systematized Nomenclature of Medicine, Clinical Terms (SNOMED CT; Cornet and de Keizer, 2008). In 1986, the United States National Library of Medicine began a large project researching techniques to create computer pro-

grams that "understood the biomedical meaning in user inquiries" and "use this understanding to retrieve and integrate relevant machine-readable information for users" (The Unified Medical Language System; UMLS; Lindberg et al., 1993). The Foundational Model of Anatomy (FMA) was initially developed as an enhancement of the anatomical content of UMLS (Rosse et al., 1998b,a; Rosse and Mejino, 2003). In 2000, the Gene Ontology Consortium began work on an unambiguous representation of knowledge to deal with the problem of connecting genes to their gene products (Gene ontology; Ashburner et al., 2000), which was then subsequently mapped into UMLS (Lomax and McCray, 2004). These developments brought about an increased usage of a technology known as "ontologies", intended to enable knowledge representation to go beyond what was currently possible with databases.

4.1.1 Ontologies

The word "ontology" takes its meaning from the branch of philosophy concerned with categories of things that exist in the world, and the relationships of similarity and difference between them. The more specialized sense of the word "ontology" has been applied by computer and information science, as computer systems have provided a means to use digital logic to enforce constraints of rigor on descriptions of entities (Gruber, 1993; Antoniou and Harmelen, 2009). Computer systems have also been developed to allow for automated query, inference, and reasoning from properly formalized knowledge. These developments have enabled the construction of "expert systems" (Russell and Norvig, 2003), which are capable of organizing and processing complex logical information and helping humans find unintuitive connections between facts.

Why work with ontologies instead of relying on databases? While databases are extremely powerful means of capturing and organizing data, one of the challenges of their usage for open ended discovery is that the relationships between data types may change rapidly as new information comes out. Due to the fact that database columns are separate data objects from rows, there is no explicit,

strongly-typed[1] relationship between data entities as there are in schemes such as the Resource Description Framework (RDF) (Spyns et al., 2002). The disadvantage of keeping relationships implicit within databases is it causes practical management burdens to the creation of a flexible information management structure where the domain knowledge model is rapidly evolving. Because of the practical problems using standard relational databases, some researchers organizing neurobiological information turned to entity-attribute-value databases (Miller et al., 2005) or ontologies. For a more detailed review of ontologies and their applications in the neurosciences, please see Larson and Martone (2009).

4.1.2 Computer-assisted knowledge management in the neurosciences

Efforts to use information management technologies and ontologies to help organize the domain knowledge of neuroscience specifically have been underway for over two decades. Earliest work used systems such as HyperCard (Brinkley et al., 1989), an application sold with the Apple Macintosh in 1988 that combined database capabilities with a graphical, flexible, user-modifiable interface. Known as the NeuroNames brain nomenclature (Martin et al., 1990), it evolved into a more complete brain hierarchy built according to the standards of the UMLS (Bowden and Martin, 1995) and was eventually incorporated into the BrainInfo resource (Bowden et al., 2007). In part due to a mandate of the Human Brain Project (Huerta et al., 1993), several projects were supported and emerged that were translating neuroscience specific data into digital forms. For example, the University of Southern Californias Brain Architecture Management system (BAMS; Bota et al., 2005) collated a significant database of neuroanatomical terms and collections into an ontology (Bota and Swanson, 2008). Projects such as the Cell-centered Database (Martone et al., 2002), SenseLab (Crasto et al.,

[1]A term from computer science referring to the explicit management of data type within the text of the code, as opposed to allowing data type to be implicit and managed by the compiler. Declaring a variable to be a String in your code is evidence of a strongly-typed language. Languages that do not require the declaration of a variable as a string are referred to as "weakly-typed"

2007), CoCoMac (Stephan et al., 2001), and Neurome (Bloom and Young, 2005) are just some of the many information systems that found solutions to cope with the multi-dimensional nature of neuroscience knowledge within their targeted domains(Bloom, 1996). However, as more data were being made available in digital form in the neurosciences in separate databases, there became an increased call for neuroscience-specific data integration across domains with ontologies as the main backbone (Martone et al., 2004).

Based on the recognition of a lack of a consistent semantic framework for the neurosciences, the National Institutes of Health's Blueprint for Neuroscience Research project created The Neuroscience Information Framework (NIF; http://neuinfo.org; Gardner et al., 2008) in 2005. The project was conceived both to provide a current inventory of resources (tools, materials, data) relevant for neuroscience and to provide the means by which such resources could be effectively searched. The NIF was also tasked with providing the necessary framework for constructing and annotating such resources to promote their discovery and utilization. The NIF has been available in production since Fall of 2008. It is supported by an expansive lexicon and ontology, built through the synthesis of open access community ontologies, covering the broad domains of neuroscience and an infrastructure for bringing together diverse data sets into a single portal (Bug et al., 2008). The current NIF lists over 4000 individual resources of relevance to neuroscience and its virtual data federation brings together millions of records from independently maintained databases.

One of the mandates of the NIF was to apply an ontological approach to the organization of databases and knowledge in the neurosciences. Prior to NIF's establishment, the Biomedical Informatics Resource Network produced a comprehensive cross-disciplinary ontology for the neurosciences. This ontology, called BIRNLex, evolved into the Neuroscience Information Framework Standard ontology (NIF Standard ontology; Bug et al., 2008). The list of topics it covers includes behavioral activity, behavioral paradigms, brain regions, cells, diseases, molecules, nervous system function, subcellular components (see chapter 2), information resources, resource types, and qualities. Demonstrations of automated reasoning

and processing enabled by organizing information in the neurosciences into the NIF Standard Ontology helped to validate the importance of this approach Fong et al., 2007; Cheung et al., 2009.

In addition to the activities of the NIF, a consortium dedicated to improving the quality of biomedical ontologies focused, among other things, on a consistent set of relations has been operating since 2005 called the OBO foundry (Smith et al., 2005, 2007). Most recently, many of these ontologies have been collected into the National Center for Biomedical Ontology's BioPortal resource (Noy et al., 2009). Taken together, these activities represent an effort by the biomedical ontology community to build a foundation of machine-processable descriptions of biology reality.

4.2 Challenges and motivation

The aim of making knowledge about biology, and neurobiology in particular, machine-processable through the use of ontologies comes with several fundamental challenges. First, the domain is a poor candidate. The domain of all entities relevant to neurobiological function is extremely large, highly fragmented into separate subdisciplines, and riddled with lack of consensus. These characteristics make neurosciences challenging to describe using ontologies (Shirky, 2005). The nature of neurosciences suggested that while small, well defined pieces of biological reality could be encoded into ontologies, scaling the process up to encompass the whole domain would require a different approach.

A second challenge is the need for human curation. The need for humans to structure knowledge limits the rate that knowledge can be ingested into machine-processable systems. While efforts such as Texpresso (Müller et al., 2004) have made progress applying automated text-mining techniques to this problem, a completely automated solution has not yet emerged. As a result, for now, human beings must continue to be a significant part of this effort. All the projects listed above have taken the approach of hiring curators and in some case ontologists dedicated to the task of knowledge engineering, data entry and data processing within the

specific information system they had built. Moreover, in order to be effective at curation and knowledge engineering, an individual must learn to follow a logical rigor that is akin to calculus and differential geometry in its formal and arcane nature. In short, curation and knowledge engineering in the biosciences is still very manual, highly technical, and therefore costly.

A third challenge is that the tools used to create and maintain biomedical ontologies require a lot of specialized knowledge and have not been inherently collaborativethough some are moving that direction. The most popular and functional ontology editors, such as Protègè (Rubin et al., 2007) or OBO-edit (Day-Richter et al., 2007), were originally designed for single user interaction. This presented a challenge to teams that wished to collaboratively edit an ontology. Only recently have collaborative editing tools begun to emerge with full support for ontologies (Tudorache et al., 2010, 2011). Moreover, Protg and OBO-edit were designed as stand-alone applications not suitable for display on the World Wide Web. Only with the emergence of the Bioportal (Noy et al., 2009) and Web Protg (Tudorache et al., 2011) have web-accessible interfaces to ontologies been made available. Another challenge for working with ontologies is that as they grow larger and larger, software applications for editing an ontology have difficulty efficiently processing the ontology, and this in turn may cause programmatic errors that go undetected when the ontology size is small.

A final challenge is the disconnect between the communities concerned with making primary observations of biology knowledge and those concerned with creating machine-processable representations of that knowledge. Following directly from the high costs of describing biological observations in machine processable forms, both in terms of skills and technologies, it has historically been difficult for working scientists to derive value from ontologies. Ontologies have been difficult to find, difficult to examine, and even more difficult for domain scientists to contribute to in order to correct errors, make suggestions, or augment the data model. For these reasons, the barrier to entry is high for a biological scientist who is engaged in the observation of living systems — also a highly technical endeavor — to cross over into the area of biomedical ontologies and easily make useful con-

tributions. This high barrier to entry has provided a disincentive that has kept the communities of neuroscience apart from the community of biological ontology, preventing the necessary crosstalk between these disciplines. These factors have also led to a superficial representation of neural structures in many community ontologies, e.g., the Gene Ontology.

As a result of the challenges enumerated above, we began to look within the NIF project to emerging technologies that might offer alternative ways to collaborate in producing and editing structured knowledge. The wiki technology, prominently on display at Wikipedia, brought with it advantages of rapid editing of web pages by any author, and provided a system well suited for collaborative management of knowledge about expansive domains (Neumann and Prusak, 2007; Spinellis and Louridas, 2008). The semantic extension to the Wiki made it possible to formalize categories and the relationships between them, the cornerstone of knowledge representation. Moreover, the open source nature of the software underlying Wikipedia made it an ideal platform for others to build additional websites. Biomedical communities such as the BiomedGT wiki

(Solbrig and Jiang, 2009), a project to reconcile terms related to cancer, run by the National Cancer Institute and the Mayo Clinic. They had been advised to express the NCI Thesaurus (Sioutos et al., 2007) in OWL and to better adhere to ontological principles (Ceusters et al., 2005). Additionally they were advised to capture all dialog and discussion in the editing and curation process in order to capture authors intent and identity. Around the same time, researchers in the field of the "Semantic Web" were exploring how distributed networks of people could collaboratively structure knowledge out of raw data and information (Neumann and Prusak, 2007) and began building tools bridging ontologies and wiki technologies to enable this Solbrig and Jiang (2009); Krotzsch et al. (2006); Vrandecic and Krötzsch (2006).

In this article, we describe NeuroLex.org, a semantic wiki-based website and knowledge management system, the goal of which is to bring the complex frontier of knowledge within neurobiology into a framework that allows neuroscientists to review the facts of neuroanatomy, aggregate their personal understanding with

that other scientists, and expose facts that are still controversial or missing on the parts list of the nervous system. To date, the site is tracking $\tilde{1}$7,300 unique things in neurobiology spanning experimental techniques, behavioral paradigms, anatomical nomenclature, genes, proteins and molecules. Here we show how the structuring of information about these parts in the nervous system can be reused to answer multiple questions neuroscientists would want to know, such as displaying all known GABAergic neurons, and displaying all known brain regions that send axons into the cerebellar cortex.

4.3 Methods

Neurolex is built using the Semantic Mediawiki (Krotzsch et al., 2006) platform. To provide some basic structure to the Wiki suitable for neuroscience content, we utilized the Semantic Forms extension. This extension allows a set of customizable forms to be constructed, whose fields can be filled in, via autocomplete, by a subset of the terms within the wiki customized to each field (see Fig. 4.3 for an example of a form). Because of this, content that is added to the wiki is immediately available to select from when constructing the knowledge of a page. Additionally, the forms shield users from relying on sometimes arcane conventions of wiki text.

After installing Semantic Forms, the properties, templates, and forms were designed and created to assist users in providing structured information, without having to learn the Semantic Media Wiki syntax. We created forms for 1) neurons, 2) brain regions, and 3) resources. In addition, a generic template input form was created for all other content that was not covered by these. These forms can be seen when adding a new term on the landing page (Fig. 4.1), or when editing a term that used the form in its creation.

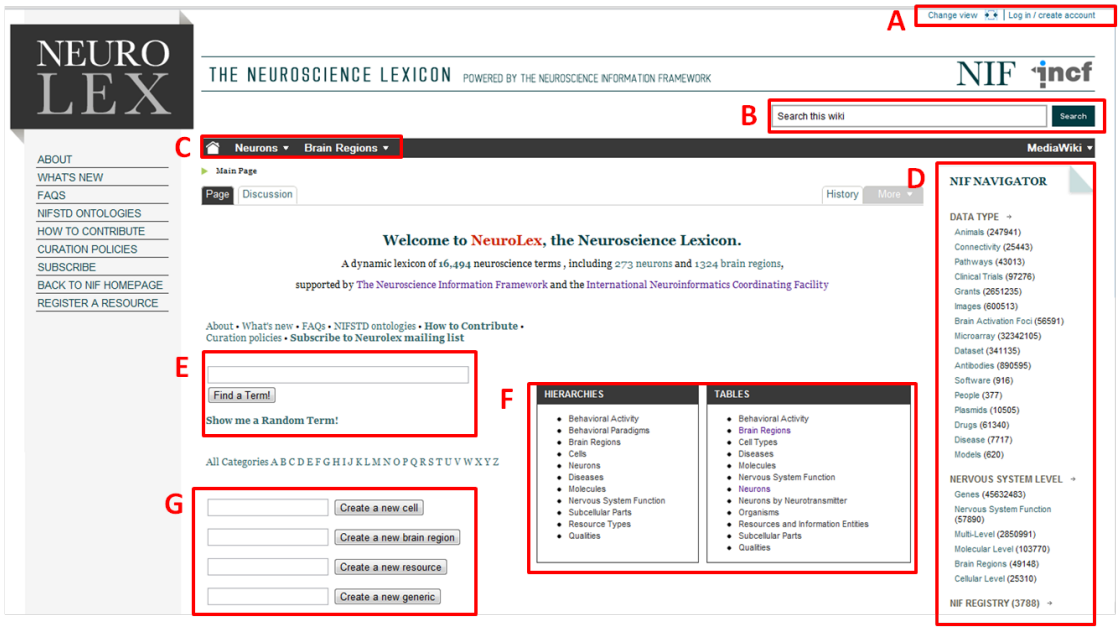

Figure 4.1: Landing page for NeuroLex.org. Several features are highlighted. A) Login / user management controls. B) Global site search bar. C) Quick navigation to neuron or brain region information. D) NIF Navigator, connecting the Neuroscience Information Frameworks federated resources to each NeuroLex page. E) Global site search bar. F) Quick navigation to hierarchies or tables containing detailed information about diverse entities in Neuroscience. G) Quick creation forms for cells, brain regions, resources, and generic page contents.

In addition to extensions that were already available in the public domain, we constructed some custom extensions of our own to enable automatic generation of identifiers for each category/concept, and a tool to allow categories to be uploaded via a comma separated value (CSV) text file. A full list of the extensions used in Neurolex can be found online[2].

In order to allow search engines such as Google to index our page content, we customized the information in the "description" meta-tag in the header of each page to display text that was specific to each page.

[2]http://neurolex.org/wiki/Special:Version

4.3.1 Population and semantic query of NeuroLex

NeuroLex.org was originally populated with the NIF Standard ontology Bug et al. (2008) in an effort to improve the process of editing and maintaining it. The initial update process, written with custom scripts using the PyWikipedia bot python framework, created one wiki page from each ontological class in the ontology. Content that is added or updated in NeuroLex.org is contributed back to the NIF Standard ontology. This content is not directly added to NIF Standard ontology, but is incorporated into the NIF Standard ontology OWL file by a knowledge engineer after curation by the NIF ontology group. This is done to enable the NIF Standard ontology to be authoritative, while allowing the content in NeuroLex to be more fluid.

To demonstrate the additional knowledge aggregation features of the wiki, we constructed several pages defined by queries that are automatically updated to reflect the content of the wiki at that time (e.g. Fig. 4.6 & Fig. 4.7). Examples include the overview of brain regions, neuron types with definitions, and neuron by neurotransmitter.

We have set NeuroLex to export a complete RDF graph of the entire contents of the wiki and upload them to a free triple store host provided by the N2 platform by Talis once every hour.

To validate the system, and to test whether it could be used to answer meaningful questions in neuroscience, we created a test case around projections to the cerebellum by augmenting the structures imported from NIFSTD with connectivity information, largely derived from the rodent and feline cerebellum, as summarized in (Altman and Bayer, 1997). While many questions can be asked using Semantic MediaWikis native query language, we initially found ourselves limited by an inability to pose queries that could handle an arbitrary number of transitive operations in a single query[3]. To address this, we imported the RDF graph from NeuroLex into an instance of the OWL-IM semantic repository[4] in-

[3]http://j.mp/A8V3YJ
[4]http://www.ontotext.com/owlim

stalled on Amazon EC2[5]. We then utilized SPARQL 1.1[6] queries to explore the knowledge base and answer specific questions about connectivity, as well as to pose questions about the completeness of the descriptions of neurons.

[5]http://neurolex.org/wiki/Reasoning_With_SPARQL_1.1
[6]http://www.w3.org/TR/sparql11-query/#propertypaths

4.4 Results

Table 4.1: Overview of key contents in NeuroLex.

Category	Definition	Number of terms
Behavioral activity	Behavior; the actions or reactions of an object or organism, usually in relation to the environment or surrounding world of stimuli. (NCI) [...]	27
Brain Regions	Anatomical divisions of the brain. Regions of brain across multiple species are contained within this category.	1324
Cell Types	The basic structural and functional unit of all organisms. Includes the plasma membrane and any external encapsulating structures such as the cell wall and cell envelope (Gene Ontology).	340
Diseases	A disorder of structure or function in a human, animal, or plant, esp. one that produces specific signs or symptoms or that affects a specific location and is not simply a direct result of physical injury.	369
Molecules	A biomolecule is any organic molecule that is produced by a living organism, including large polymeric molecules such as proteins, polysaccharides, and nucleic acids as well as small molecules such as primary metabolites, secondary metabolites, and natural products. (Wikipedia)	569
Nervous System Function	An activity or purpose natural to or intended for part of or a whole nervous system.	37
Organisms	A living thing	902
Protein	A biological macromolecule that is composed of amino acids linked in a linear sequence (a polypeptide chain) and is genetically encoded. [..]	1354
Resources	An entity that provides access (either in the open community or within an organization) to material, intellectual, financial, technological, or electronic means of carrying out research and development. (BRO)	4353
Subcellular parts	Cell part which has a definable shape, bounded predominantly by bonafide boundaries and is countable (FMA).	148
Qualities	A quality is an attribute or a property.	1379

NeuroLex.org has made progress towards establishing and enabling the creation of a comprehensive corpus of machine-processable, multi-scale neuroscience knowledge that is editable collaboratively online and is discoverable on the internet by search engine queries. We were motivated to create a system for neuroscience that had the following features:

- Every concept needed to be a web page
- Every concept needed to be editable individually
- Editing a concept needed to be easy
- Linking concepts needed to be easy

Semantic MediaWiki is an extension on the MediaWiki software that is the foundation of Wikipedia and supports millions of users a day. Semantic MediaWiki extends MediaWiki by allowing users to formalize knowledge through the use of special tags within wikitext. This means that a page in Semantic MediaWiki can be marked up to reveal knowledge within it in a structured way. For example, a wiki page about an apple can indicate that its color is red and its flavor is sweet through the use of special properties such as "has_color" or "has_flavor". This feature allows a wiki page to serve essentially as a database record for the topic that it covers, going beyond a simple text entry and allowing software to be written to analyze and synthesize content across pages. Limited support for online reasoning has been incorporated into Semantic MediaWiki using inline queries, though more sophisticated reasoning processes have been incorporated into other wiki software (Kuhn, 2008).

NeuroLex currently hosts 17,388 active concepts including 275 neurons and 1366 brain regions, an increase of 10,257 concepts since its original launch[7]. Table 4.1 shows a high-level overview of the contents of NeuroLex.org, broken down by the high-level categories spanning scales and domains relevant to neuroscience.

An example page within the NeuroLex is shown in figure 4.2. In this example, structured knowledge about the cerebellum is assembled and displayed on a

[7]http://j.mp/xJnXej

web page, which allows the user to bookmark or share this content by the unique URL. As NeuroLex is built on top of MediaWiki software, an edit button is present that allows any user to modify the contents on this page (B). Boxes corresponding to (D), (E), and (F) demonstrate the ability of the NeuroLex.org infrastructure to assemble knowledge related to the cerebellum automatically. On a standard Wikipedia page, all the knowledge on a page must be manually entered within the single text box provided. Because NeuroLex leverages a semantic backend to structure knowledge, its pages can dynamically call information from other pages in when they are relevant. For example, in (D), all neurons reporting their somas to be within the cerebellum, or within any other brain region that is defined as *Is part of* the cerebellum (as shown in E), is listed here. The list is assembled via Semantic MediaWikis inline query functionality, which allows structured content from other pages to be organized and reported. Additionally, any cell that reported that its axon passes through the cerebellum or its parts is listed separately.

The ability to structure knowledge within the Semantic MediaWiki platform also allows classes to be rendered using tables, trees, and lists that combine the asserted content of the class with queried content derived from other classes. For example, (E) in figure 4.2 displays a dynamic tree that lists the brain region classes that have been asserted as *Is part of* the Cerebellum. Lastly, (F) shows tracts that have been defined as going into the cerebellum or going out of it.

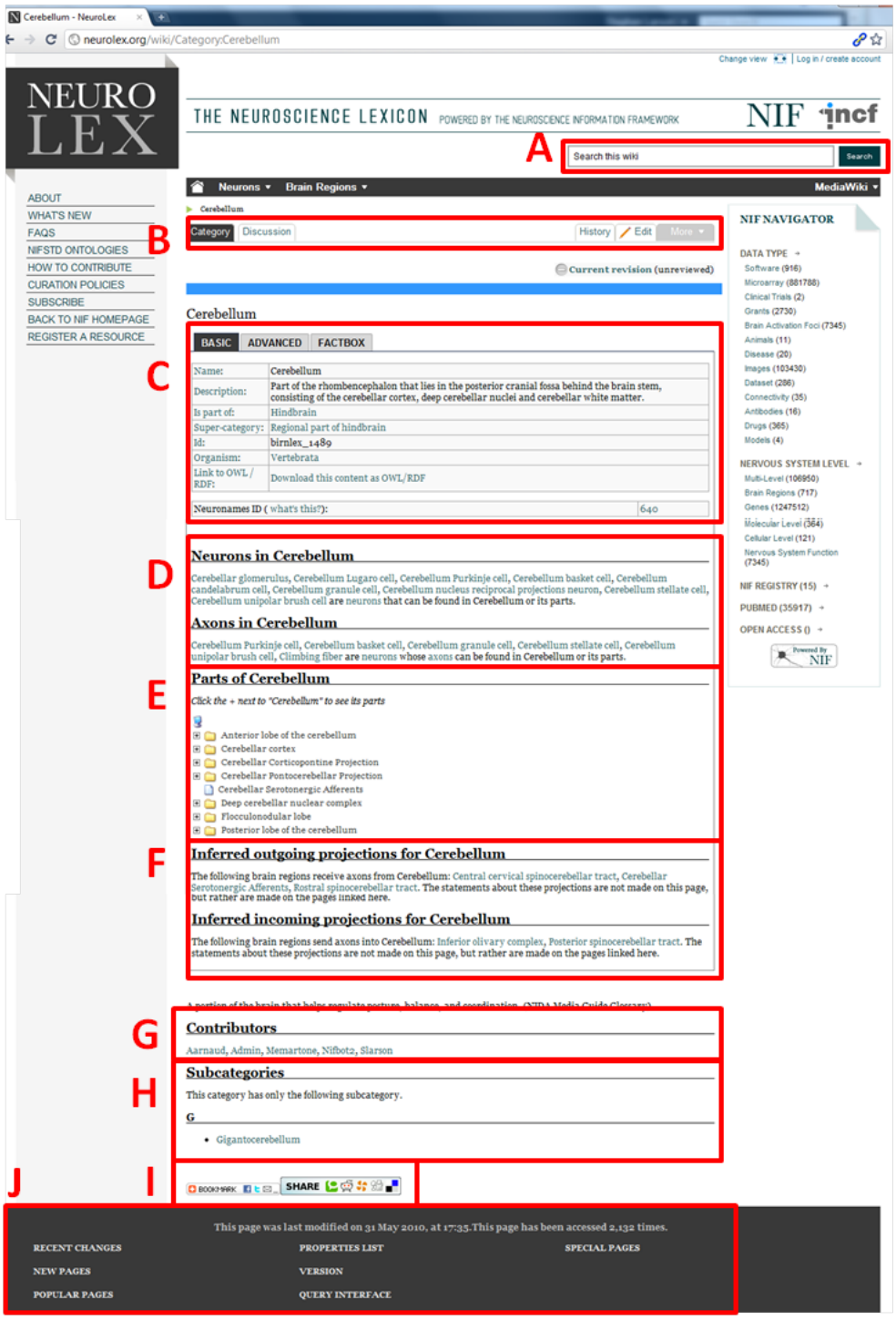

Figure 4.2: Example category page for the concept "Cerebellum."

Figure 4.2 shows an example category page for the concept "Cerebellum." The elements of the page include: A) Global site search bar. B) Wiki controls for

this page, including link to a discussion page, page edit history, and edit controls for this page. C) Basic facts for this concept, including text description, super category and more. Tabbed interface also contains additional advanced facts. D) Advanced auto-generated report for neurons whose somas or axons are located in the Cerebellum. E) Advanced auto-generated report of other brain regions that are listed as being a part of the cerebellum. F) Advanced auto-generated report of outgoing and incoming projections for the cerebellum G) List of users that have made edits to this page. H) List of subcategories for this concept, i.e. concepts that are more specific than this current concept. I) A widget that allows users to share this page with their social networks. J) A global footer that contains last modified information, as well as site-wide information like recent changes, a list of new pages, special reports, and version information.

4.4.1 Editing the NeuroLex

To enable users to make modifications to knowledge that has been structured within NeuroLex.org, we implemented a form-based edit system as opposed to the standard free text and markup system used by many wiki sites, including Wikipedia. Fig 4.3 shows an example of editing the page for a cerebellum granule cell within NeuroLex[8]. Any user may edit a NeuroLex page, with or without an account, though the abilities to delete and to move pages are restricted to users with accounts. This permissive approach to edits is balanced by a built-in history and change tracking mechanism that makes edits transparent to curators, who can easily undo changes that negatively impact the quality of the knowledge base. Change history is available both at the level of individual pages[9] or at the level of the entire site[10] .

[8]http://j.mp/xzUwKR
[9]http://j.mp/zC8Aci
[10]http://neurolex.org/wiki/Special:RecentChanges

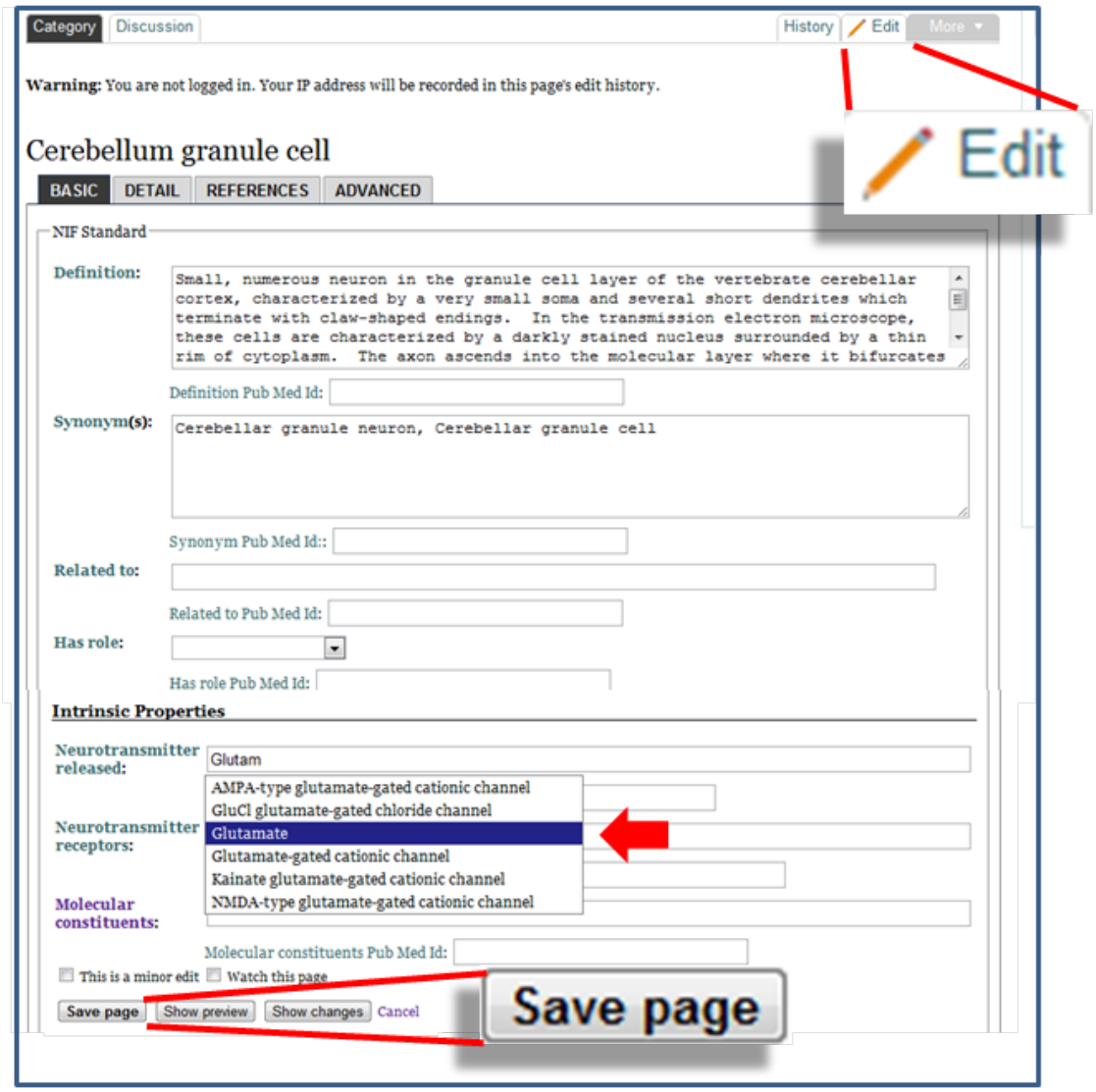

Figure 4.3: The edit form for the Cerebellum granule cell page. A user has pressed the edit button, enlarged in the upper right hand corner to get here. Text boxes enable the user to make edits to the fields of information on the page. Towards the bottom of the page, a user is in the process of typing in "Glutamate" into the field "Neurotransmitter released". As the user hits each key, an autocomplete interface helps to choose the appropriate term from a drop down list, indicated by the solid arrowhead. After the user is done, the save button at the bottom of the page is clicked.

NeuroLex.org has been online since December 2008. In that time it has

received (at time of this writing) 203,796 absolute unique visits and 568,972 page views from 185 countries and territories. Currently NeuroLex is receiving $\tilde{6}$00 hits per weekday. One hundred and three users have made edits, 12 of them have been active in the process of editing in the last 2 months. NeuroLex has recorded 188,304 edits providing a ratio of visits to edits of $\tilde{1}$.1:1 and a ratio of edits to content of 11:1 (Wikipedia has 14:1) (Spinellis and Louridas, 2008).

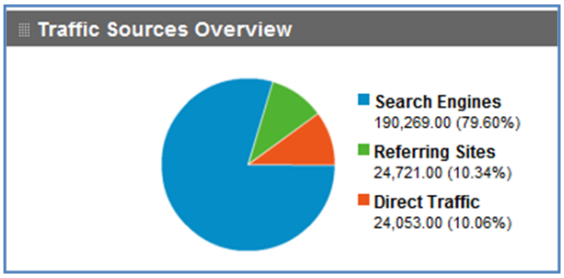

Figure 4.4: Traffic sources to NeuroLex.org since December 2008. Direct traffic refers to a user typing "neurolex.org" into the browser or following a personal bookmark. Referring sites are visits where a user started at another site and clicked a link to arrive at NeuroLex.org. Search engines refers to any user that came to NeuroLex.org from a web search. Google searches made up 95% of the search engine traffic.

A significant amount of traffic comes to NeuroLex.org from Google (figure 4.4). As shown in the usage graph in figure 4.5, Google search has had a strong impact on traffic flow to NeuroLex.org. Modifications to the way that NeuroLex reports the contents of each page to Google resulted both in a reduction of average traffic at the end of 2009 and a sharp increase of average traffic at the end of 2010.

At time of writing 523 distinct terms when searched in Google returned a NeuroLex page in the first 10 results. Over the course of its history, the top 10 Google searches that have pointed to NeuroLex.org have led to 7, 228 hits and include terms such as "dorsal root ganglion", "telencephalon", "nervous system function", "cholinergic neurons", "mni atlas", "mitral cells", "primary olfactory cortex", "lateral septum", "movement quality", and "oddball paradigm". How-

ever, those top 10 hits make up only 3.8% of the 190,000 visits from Google. In fact, 96.2

Figure 4.5: A graph of visits to Neurolex.org over time since December 2008. Hits in 2010 were depressed by modifications in the presentation of metadata for search engines. This was corrected at the end of 2010, which led to increased traffic seen in 2011.

4.4.2 Usage and adoption

NeuroLex.org has been cited in a series of articles commentaries and reviews as a tool that can help address informatics challenges in neuroimaging (Nielsen, 2009b; Turner et al., 2010; Mejino et al., 2010), autism (Young et al., 2009), event-related brain potentials (Frishkoff et al., 2009), and cognitive science (Miller et al., 2010; Derom et al., 2010; Yarkoni et al., 2010). The neuroinformatics community has cited it as a development encouraging integration between tools (Nielsen, 2009a; Akil et al., 2011; Ascoli, 2010; Katz et al., 2010; French et al., 2009; Hamilton and Ascoli, 2010). It has also been cited by the semantic web community as an example of a new development in distributed collaborative creation of biological ontologies (He et al., 2009; Cheung et al., 2009; Alquier et al., 2010). In order to enable broad community contribution to the NIF Standard Ontology (Bug et al., 2008), the Neuroscience Information Framework adopted NeuroLex.org and made it available as an easy entry point for the community (Grethe, 2009; figure 4.1). As described in Imam et al., (2011), the NeuroLex has become the community facing front-end to the NIF Standard Ontology as well as the platform on its comprehensive registry of neuroscience resources is hosted. As content is modified and updated on NeuroLex.org, an ontology engineer tracks the latest changes, reformulates them into a stricter formalism if necessary, and updates the OWL

representation of the NIF Standard Ontology. The NIF Standard Ontology then serves as the back-end for concept-based search that is deployed to the NIF search engine (Akil et al., 2011).

4.4.3 Neurons and their properties

NeuroLex.org has also served as a test bed for the efforts of the Program of Ontologies of Neural System (PONS), an activity of the International Neuroinformatics Coordinating Facility (INCF) (Martone et al., 2010). In the last year, content and edits have been contributed by several members of the task forces in this activity. Task force members can indicate their ownership of edits by creating a user account with their name and logging in before making edits. NeuroLex also allows individuals to make edits anonymously, after completing a test that distinguishes human users from computer programs.

Through the activities of the NIF and the INCF, we used the NeuroLex as a platform to aggregate information about the cells of the nervous system. NeuroLex can tally the number of subtypes of a given concept as users make updates– something that is not possible in Wikipedia. Table 4.2, generated by a query in the NeuroLex (`http://neurolex.org/wiki/Contents_Overview`), shows the cell types contained in the NeuroLex. Of the 270 neurons and 41 glial cell types listed, the vast majority are identified in the mammalian nervous system.

Table 4.3 shows the breakdown of neurons that have defined what neurotransmitter they use. Currently 43% of all neurons listed in NeuroLex have been associated with some neurotransmitter.

Table 4.2: Nervous system cells in the NeuroLex.

Category	Definition	Number of terms
Neurons	The basic cellular units of nervous tissue. Neurons are polarized cells with defined regions consisting of the cell body, an axon, and dendrites, although some types of neurons lack axons or dendrites.[...]	270
Glial cells	A non-neuronal cell of the nervous system. They not only provide physical support, but also respond to injury, regulate the ionic and chemical composition of the extracellular milieu. Guide neuronal migration during development, and exchange metabolites with neurons.	45

Table 4.3: Neurons grouped by neurotransmitter in the NeuroLex. Not listed here are Serotonergic neurons (3) and Norepinephrine neurons (1).

Category	Definition	Number of terms
GABAergic Neurons	A neuron that uses GABA as a neurotransmitter	63
Glutamatergic Neurons	A neuron that uses glutamate as a neurotransmitter	28
Cholinergic Neurons	A neuron that uses Acetylcholine as a neurotransmitter	15
Dopaminergic Neurons	A neuron that uses dopamine as a neurotransmitter	6

Table 4.4: Overview of key relations.

Name	Usage count
Synonym	38785
CurationStatus	10603
DefiningCitation	7959
Is part of	5355
Abbrev	2661
Species	1290
DefinitionSource	524
ExampleImage	400
LocationOfAxonArborization	167
DendriteLocation	71
DefiningCriteria	67
CellularSynapticTarget	60
AfferentProjections	51
EfferentProjections	43

Tables available online[11][12] show examples of a subset of the properties of some of the neurons in NeuroLex.org. Since its inception, the editable input form that has been used to define a neuron has evolved based on feedback from several members of the neuroscience and neuroinformatics community. The current set of properties represented on the input form are visible at NeuroLex.org[13].

The relationships mentioned thus far (e.g. *Is part of*, tracts projecting in and out of brain regions, the presence of a cell soma in a brain region, are explicitly defined properties in NeuroLex.org. A significant amount of planning and effort has gone into defining these and other additional relationships that are most appropriate to the domain of neuroscience (cite Gordon and Giorgios paper). These relationships provide the underlying machine processable link between concepts and data and concepts to other concepts (see table 4.4 for a partial list of relations used).

[11] http://neurolex.org/wiki/Neurons_of_the_cerebellum
[12] http://neurolex.org/wiki/Neurons_of_the_hippocampus
[13] http://neurolex.org/wiki/Form:Petilla_neuron

In order to determine how complete the knowledge is within NeuroLex, we queried the system to see how many properties that had been were assigned values (see Supplemental Table 1). With 250 neurons and 30 properties, 7500 facts would be needed to fully flesh out the descriptions of the current list of neurons in Neurolex. Of that total, we found 33% of the fields were completed. Some of the neuron types have been filled in more than others. For example, neurons of the olfactory system have received the most complete description, on average, compared to the rest of the neurons. Because additional properties may be added in the future, the columns may continue to expand as necessary. This report reveals that while there is still more to be done to come to a complete description of the neurons in NeuroLex, we have been able to create a framework that places a quantifiable upper bound on the amount of description necessary to fill in knowledge about neurons.

4.4.4 Custom list functionality

Figure 4.6: The page for all Glutamatergic neurons.

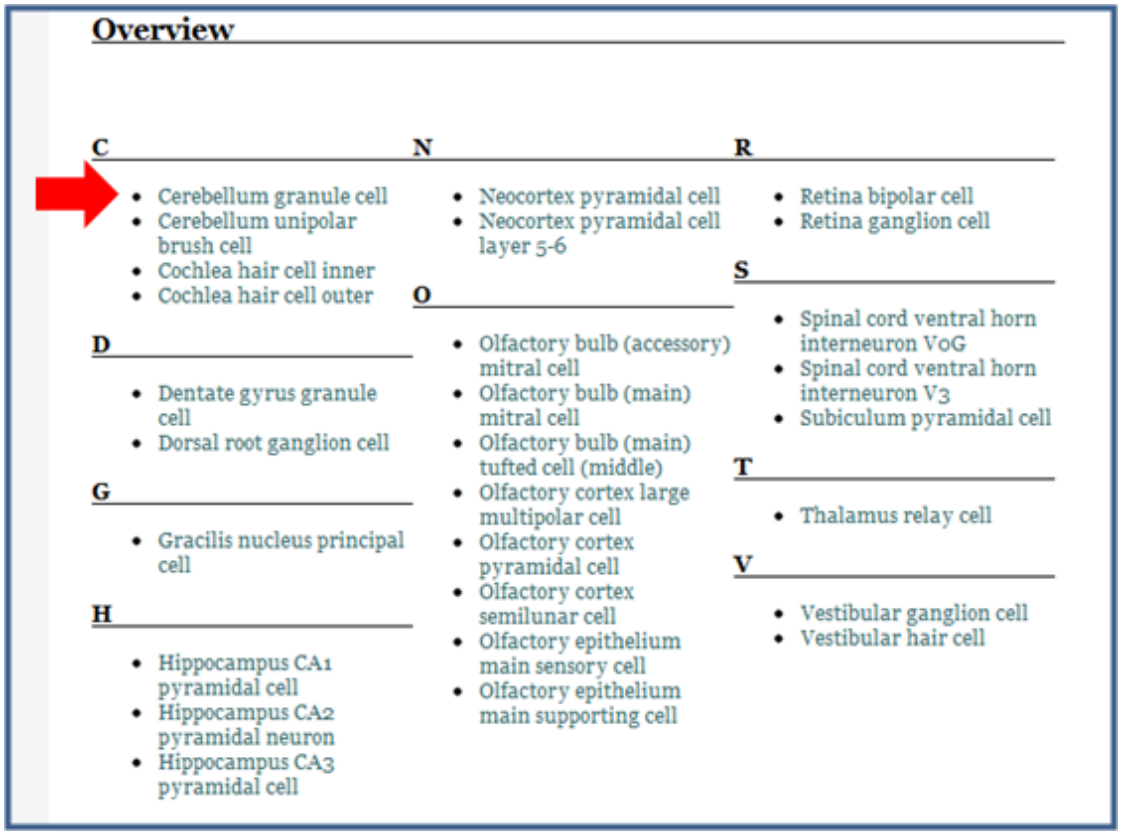

Figure 4.7: The modified overview section of the Glutamatergic neuron page. After having entered Glutamate as the Neurotransmitter released in the Cerebellum granule cell page, this neuron now appears in the list when it did not before (compare with open arrowhead in fig. 4.6 above)

Figure 4.6 illustrates a page in NeuroLex that is designed to keep track of neurons based on a particular property. This page is a defined neuron class, which means that it is a collection of neurons that come together as a result of a shared property, the presence of glutamate as its neurotransmitter. The neurons within NeuroLex for which this is true are listed in the Overview section, which starts with the open arrowhead. In this case, the page only lists neurons that are defined to have glutamate as their main neurotransmitter. Rather than editing this page directly in order to maintain this list, the page is updated automatically by the system (figure 4.7) when a user edits the page of an individual neuron and indicates that its neurotransmitter is glutamate (figure 4.3, filled arrowhead).

Table 4.5: Wiki text that creates a list of Glumatergic neurons

```
{{#ask: [[Category:Neuron]]
[[Neurotransmitter::Category:Glutamate]]
| format=category
}}
```

Table 4.5 shows the small amount of wiki-text that is used to create this automatic list. This can be copied and pasted by users to create custom lists of other types of specialized categories of items. The results of this automatically generated list also becomes visible to search engines, meaning that a list that is useful to an individual may also become content that is reused by others. For example, at time of writing, the custom list on NeuroLex for cholinergic neurons appears on the first page of results in Google.

4.4.5 Cerebellum Reasoning Example

In order to validate the ability of NeuroLex to structure information in a form that enabled the answering of significant questions about the nervous system, we explored the capacity of the knowledge base to answer the question: What are all the brain regions that send projections into the cerebellum or any of its parts? To address this question, we included statements derived from the cerebellar anatomical literature in order to give the system a means of understanding 1) the parts of the cerebellum, 2) what it means to send projections, and 3) what mossy fibers are. In this case, mossy fibers are implied by any axons that enter the cerebellum, or any projection whose destination is the cerebellum.

Table 4.6 demonstrates the result of the query that we issued to answer this question. The query itself is shown in table 4.5. The significance of these results are discussed below.

Table 4.6: SPARQL 1.1 query to return the brain regions that project into the cerebellum

```
## Prefixes
PREFIX rdfs:<http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd:<http://www.w3.org/2001/XMLSchema#>
PREFIX owl:<http://www.w3.org/2002/07/owl#>
PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX property:<http://neurolex.org/wiki/Special:URIResolver/Property-3A>
#############################################################################
## Get me all brain regions that project to any Regional part of cerebellum,
## using the Efferent/Afferent projection properties
## AND the LocationOfAxonArborization and SomaLocation properties
## Report results by starting region, pathway taken, ending region
#############################################################################
select DISTINCT ?startName ?pathName ?endName where
?x property:Id "birnlex_1571"^^xsd:string .
?end rdfs:subClassOf* ?x .
OPTIONAL
?path property:LocationOfAxonArborization ?end .
?path property:SomaLocation ?start .
?start property:Label ?startName .
?path property:Label ?pathName .
?end property:Label ?endName  .
OPTIONAL
?path property:EfferentProjections ?end .
?path property:AfferentProjections ?start .
?start property:Label ?startName.
?path property:Label ?pathName .
?end property:Label ?endName
```

Table 4.7: Results from performing the query in table 4.6 to answer the question what brain regions project to some part of the cerebellum. *The final column on cell type has been added manually.

Structure that cell soma is in	Pathway taken	Structure that axon arborizes in	Cell type*
Vestibular ganglion	Cerebellar Primary Vestibular Afferents	Flocculus	Vestibular ganglion cell
Vestibular ganglion	Cerebellar Primary Vestibular Afferents	Hemispheric Lobule VIII	Vestibular ganglion cell
Vestibular ganglion	Cerebellar Primary Vestibular Afferents	Vermic Lobule I	Vestibular ganglion cell
Vestibular ganglion	Cerebellar Primary Vestibular Afferents	Vermic Lobule IX	Vestibular ganglion cell
Vestibular ganglion	Cerebellar Primary Vestibular Afferents	Vermic Lobule X	Vestibular ganglion cell
Trigeminal nucleus	Trigeminal Mossy Fibers	Hemispheric Lobule VI	Trigeminal nucleus principal cell
Trigeminal nucleus	Trigeminal Mossy Fibers	Hemispheric Lobule VIIA	Trigeminal nucleus principal cell
Trigeminal nucleus	Trigeminal Mossy Fibers	Hemispheric Lobule VIIBi	Trigeminal nucleus principal cell
Trigeminal nucleus	Trigeminal Mossy Fibers	Hemispheric Lobule VIII	Trigeminal nucleus principal cell
Trigeminal nucleus	Trigeminal Mossy Fibers	Vermic Lobule IX	Trigeminal nucleus principal cell

continued on next page

continued from previous page

Structure that cell soma is in	Pathway taken	Structure that axon arborizes in	Cell type*
Pontine reticular formation	Cerebellar Afferents From The Pontine Reticulotegmental Nucleus	Anterior lobe of the cerebellum	Serotonergic cell group B8
Pontine reticular formation	Cerebellar Afferents From The Pontine Reticulotegmental Nucleus	Hemispheric Lobule VII	Serotonergic cell group B8
Pontine reticular formation	Cerebellar Afferents From The Pontine Reticulotegmental Nucleus	Paravermic Lobule VII	Serotonergic cell group B8
Pontine reticular formation	Cerebellar Afferents From The Pontine Reticulotegmental Nucleus	Paravermic Lobule VIII	Serotonergic cell group B8
Pontine reticular formation	Cerebellar Afferents From The Pontine Reticulotegmental Nucleus	Vermic Lobule VII	Serotonergic cell group B8
Pontine reticular formation	Cerebellar Afferents From The Pontine Reticulotegmental Nucleus	Vermic Lobule VIII	Serotonergic cell group B8
Cervical spinal cord	Central cervical spinocerebellar tract	Vermic Lobule II	N/A
Cervical spinal cord	Central cervical spinocerebellar tract	Vermic Lobule III	N/A

continued on next page

continued from previous page

Structure that cell soma is in	Pathway taken	Structure that axon arborizes in	Cell type*
Cervical spinal cord	Central cervical spinocerebellar tract	Vermic Lobule IV	N/A
Cervical spinal cord	Central cervical spinocerebellar tract	Vermic Lobule V	N/A
Cervical spinal cord	Central cervical spinocerebellar tract	Vermic Lobule VI	N/A
Cuneate nucleus	Posterior spinocerebellar tract	Paravermic Lobule IV	N/A
Cuneate nucleus	Posterior spinocerebellar tract	Paravermic Lobule V	N/A
Cuneate nucleus	Posterior spinocerebellar tract	Paravermic Lobule VI	N/A
Cuneate nucleus	Posterior spinocerebellar tract	Vermic Lobule IV	N/A
Cuneate nucleus	Posterior spinocerebellar tract	Vermic Lobule V	N/A
Cuneate nucleus	Posterior spinocerebellar tract	Vermic Lobule VI	N/A
Lateral reticular nucleus	Cerebellar Afferents From The Lateral Reticular Nucleus	Anterior lobe of the cerebellum	N/A
Lateral reticular nucleus	Cerebellar Afferents From The Lateral Reticular Nucleus	Hemispheric Lobule VIIBii	N/A

continued on next page

continued from previous page

Structure that cell soma is in	Pathway taken	Structure that axon arborizes in	Cell type*
Nucleus prepositus	Cerebellar Afferents From The Prepositus Nuclear Complex	Cerebellar hemisphere	N/A
Nucleus prepositus	Cerebellar Afferents From The Prepositus Nuclear Complex	Cerebellar Paravermis	N/A
Nucleus prepositus	Cerebellar Afferents From The Prepositus Nuclear Complex	Hemispheric Lobule IX	N/A
Nucleus prepositus	Cerebellar Afferents From The Prepositus Nuclear Complex	Vermis	N/A
Pontine nuclear complex	Cerebellar Pontocerebellar Projection	Hemispheric Lobule VI	N/A
Pontine nuclear complex	Cerebellar Pontocerebellar Projection	Hemispheric Lobule VIIBi	N/A
Pontine nuclear complex	Cerebellar Pontocerebellar Projection	Hemispheric Lobule VIII	N/A
Pontine reticular formation	Cerebellar Afferents From The Pontine Reticulotegmental Nucleus	Hemispheric Lobule VIII	N/A
Vestibular ganglion	Cerebellar Primary Vestibular Afferents	Hemispheric Lobule IX	N/A
Vestibular nuclear complex	Cerebellar Secondary Vestibular Afferents	Flocculus	N/A

continued on next page

continued from previous page			
Structure that cell soma is in	Pathway taken	Structure that axon arborizes in	Cell type*
Vestibular nuclear complex	Cerebellar Secondary Vestibular Afferents	Hemispheric Lobule IX	N/A
Vestibular nuclear complex	Cerebellar Secondary Vestibular Afferents	Hemispheric Lobule VIII	N/A
Vestibular nuclear complex	Cerebellar Secondary Vestibular Afferents	Vermic Lobule I	N/A
Vestibular nuclear complex	Cerebellar Secondary Vestibular Afferents	Vermic Lobule IX	N/A
Vestibular nuclear complex	Cerebellar Secondary Vestibular Afferents	Vermic Lobule X	N/A

As a comparison, we also performed similar queries against other on-line databases that contain information about connectivity: BAMS, etc., The equivalent query to find the incoming projections to the cerebellum in the BAMS database reveals a list of 3 brain regions that project into the cerebellar cortex and 27 brain regions that project into the flocculus[14]. BAMS reports 5 regions that project into the cerebellum as a whole[15]. While NeuroLex can report the pathway or cell type that is responsible for the projection, this kind of report is not currently available from BAMS. The CoCoMac database does not store connections to the cerebellum, so the equivalent query could not be run.

The system used to create these queries, with the version of the data used to call for them, is available for inspection as an Amazon EC2 instance[16].

4.5 Discussion

The production of a well structured and comprehensive "parts list" or knowledge base that is machine processable would be a key asset to the field

[14]http://brancusi.usc.edu/bkms/brain/show-infconaf.php?aidi=789&publi=1
[15]http://brancusi.usc.edu/bkms/brain/show-conaf.php?aidi=23&publi=1
[16]http://neurolex.org/wiki/Reasoning_With_SPARQL_1.1

of neuroscience as it would drive hypothesis generation across subdisciplines. This endeavor has been the ultimate goal of many efforts throughout the last fifty years starting with the systematized nomenclature of pathology Cornet and de Keizer, 2008. The advent of widespread usage of the internet as the primary means of conducting research online has made it possible to explore entirely new means of building both broad, deep, and organized corpuses of knowledge in a distributed collaborative manner (Huss et al., 2008; Neumann and Prusak, 2007; Spinellis and Louridas, 2008). The great magnitude of increase in potential interaction of large numbers of individuals over the internet provides hope in solving the even greater challenge of building a detailed corpus of knowledge about the nervous system.

In order to tap into this potential, the exchange of knowledge in the biological sciences in the age of the internet will increasingly demand tools that allow the organization, presentation, and dissemination of the complex relationships of living systems through interfaces that are easy to update and easy to use. Maintaining a careful balance between complexity and simplicity is a multi-disciplinary challenge. Addressing this challenge requires as much attention to the interests of biological scientists who do not have deep experience with information systems as it does to the interests of logicians and ontological experts, who have experience structuring knowledge in ways to allow automated query and reasoning. These interests, frequently in competition, make up two sides of a coin. If biological scientists cannot easily get knowledge out of an information system, they cannot benefit from it. If biological scientists cannot easily put knowledge into an information system, the system will be uninteresting for lack of content. At the same time, if ontological experts cannot structure queries and reason over domain knowledge, an information system will not be able to return interesting results or reveal non-obvious knowledge connections and will also suffer from disuse.

With NeuroLex.org, we have demonstrated success in building a community platform where neuroscientists, ontology engineers and knowledge managers can structure knowledge in a wiki form in an online repository that is accessible to everybody. Usability and how users can contribute to this graph has been considered for the purpose of reducing the barrier to entry to allow and encourage

a greater number of people to contribute. Additionally, we have exposed structured knowledge about neuroscience to the world via search engines, which unlocks the potential for many others to find and learn from the knowledge base we have created.

We have also demonstrated success in producing an online information artifact that is useful for the discovery and organization of neuroanatomical facts. We invested effort into creating properties that allow knowledge to be appropriately interlinked for the purpose of creating a machine queryable semantic graph-based knowledge base that can connect facts at the micro-scale in neuroscience, to facts at a macro-scale. The system enables any user to look at different relationships between the entities through a built-in reasoning system. The knowledge base is still a work in progressmore facts are needed to fill in the significant number of gaps. However, as we have shown in the example of reasoning over facts of the cerebellum, the system is enabling us to ask much more powerful questions about facts that have been recorded in the literature. Of equal importance, the Wiki platform allows users with a minimal knowledge of ontology languages or programming skills to create custom reports, something that typically requires an ontology expert or database programmer.

4.5.1 Contribution model, usability and interface

The issues surrounding the neuron input form made it clear that issues related to usability and interface, not commonly given first priority in scientific disciplines of informatics, are increasingly becoming crucial to enable online community interaction around scientific subjects. If we are serious about pursuing distributed collaboration, also called "crowdsourcing" Howe (2006) as a means of assembling complex knowledge bases in the biological sciences, then we must immediately view other successful crowdsourcing and social media ventures such as Wikipedia, Twitter and Facebook as models to follow. The organizations behind these sites make significant investments in interface and usability in order to produce easy-to-use experiences for their target audience. This acknowledges the reality of the contribution model on the internet where sites like NeuroLex must

compete for attention and for the dedication of its targeted users. These resources are difficult to get with clunky, cluttered, unintuitive interfaces that put a large cognitive burden or have a significant learning curve to use.

The NeuroLex, though having a relatively small number of active users who make edits, has an active user for every 1400 pages. To reach the ratio enjoyed by Wikipedia, which has an active user for every 210 pages would require adding an additional 68 active users, or in other words, persuading 1 out of every 300 unique visitors to the site to make an edit over the course of two months (the time window of user activity for them to be considered "active"). Companies like Google offer free tools that enable a website to present different interfaces to different users and measure the average activity of a user given an interface. This is used by websites throughout the world to test out what encourages user contribution and what does not. While NeuroLex.org has not yet employed such tools, it is clear that such an approach is only now possible with flexible web-based interfaces for community ontology creation and management such as the one developed here.

The current NeuroLex interface has many features for encouraging community interaction (see fig. 4.1 and 4.2). For example, the ease of bookmarking, linking to, and sharing content via unique resource locator (URL), unique resource identifier (URI) or unique id lowers the barriers to individuals discovering the site and returning to it. These unique identifiers are also important means for referencing these same concepts within other information systems, thereby facilitating interlinking and integration of knowledge.

The advantage of wikis is that the edit features provide immediate gratification to users when making a contribution because they can see their edits are immediately visible globally. This approach is directly opposite to the tightly controlled nature of edits made to other ontologies as proposed by the OBO Foundry, another prominent group involved in structuring knowledge for the biosciences (Smith et al., 2007). For the OBO Foundary, only a small group of editors is allowed to make any changes, and a request system is used to make updates. OBO Foundry exposes their ontologies via the BioPortal (Noy et al., 2009), which has a mechanism for commenting or requesting modificiations, rather than making ed-

its directly to the underlying contents. This approach has been taken to ensure rigourous consistency of the knowledge-base with a gatekeeper model limiting the number of individuals allowed to make changes. In contrast, the NeuroLex wiki-based approach is to allow anyone to edit, and to deal with consistency after edits have occurred with a "Recent changes" list that makes it transparent exactly what edits have occured when and by whom (see the link in (J) on figure 4.2).

Of the $\tilde{3}0$ ontologies managed by the OBO Foundry at their SourceForge repository, the combined number of requests on all trackers for all ontologies at time of writing was 3402, which makes for 55 NeuroLex edits (over two years on line) to every OBO Foundry request for modification (over 6 years of operation at SourceForge.com). Put another way, the OBO Foundry ontologies have been requested to be edited on average 1.5 times a day while the NeuroLex ontology has been edited on average 258 times a day.

The difference in approach has an impact on the number of edits made to the underlying content. While more edits does not immediately suggest a higher quality of content, a difference of two orders of magnitude suggests a difference in the scalability of the contribution model.

4.5.2 The impact of a Google searchable ontology for neuroscience

Neuroscientists increasingly use Google as a primary source of research. The traffic patterns Neurolex has received from Google shows there is a broad worldwide demand for information about a variety of terms in neuroscience. This is powerful because with more than 4000 different online resources relevant to neuroscience on the internet (Bug et al., 2008), it would be natural to assume that NeuroLex.org would rarely appear near the top of a Google search. However, as we discovered with the traffic pattern dropping in 2010 and resurging in 2011 (figure 4.5), optimizing a page for Google's search engine can make a big difference in the usage of an online resource. If Wikipedia is any guide, the more that scientists searching for scientific terms land on pages containing information that is useful to them, the more likely they are to begin contributing their own knowledge

back to it (Spinellis and Louridas, 2008), thereby moving towards the conditions necessary to drive the vision of a community-vetted, collaboratively editing corpus of structured knowledge in the neurosciences to become a reality.

4.5.3 Knowledge Base Quality

As we have considered NeuroLex both as a platform for helping a community and as an information artifact, the question of knowledge base quality has arisen. The open nature of NeuroLex raises potential concerns about the completeness and accuracy of its content. In addition to building the NeuroLex on the software stack of Wikipedia, we have also built the contribution model on the foundation that underlies the success of Wikipedia in ensuring the breadth and accuracy of its articles Giles (2005). The idea is simple, the easier it is to edit and the more transparent those edits are to others, the more possible it is for the readership of the resource to also act as the reviewers and editors of content. However, since the concept pages are constantly changing and not rigorously peer-reviewed, they are not intended to be cited in journals as authoritative. Rather, wherever possible, we have included the ability to cite a publication elsewhere via a Digital Object Identifier (DOI) or a PubMed ID, or to otherwise indicate its original source from another resource, such as BrainInfo, BAMS, or published books. This is intended to enable and encourage the end user to find the publication from which the facts have been derived, confirm their veracity, and use these references as the authoritative, citable reference. Nonetheless, we believe that the integration of information within NeuroLex is its major strength, much in the way that the index of a book enables a quick lookup of concepts. The value of an index is derived from its ordering of concepts and the correctness of its pointers to pages. The value of NeuroLex is derived from organizing concepts into a data model and its accurate pointers to external references.

4.5.4 Relations / Properties

Understanding the importance of well-defined relationships is crucial to the mission of creating computer frameworks to grapple with the complexity of biological systems (Smith and Rosse, 2004; Smith et al., 2005). Indeed, these relationships are the glue that hold the knowledge base together – they are the edges that connect the vertices of the complex web of interactions that must exist between the biological entities playing out their roles within biological systems. For example, once you define a neuron as an entity of interest in a computer system, you are presented with the challenge of defining the set of relationships that this neuron should have to other entities – essentially what are the properties of a neuron? Despite more than 100 years of investigation, this is a challenge that is mostly unrecognized by the neuroscience community, and for which no consensus has yet been established by those in the biomedical ontology community (Migliore and Shepherd, 2005; Ascoli et al., 2008).

The challenges in defining the proper relationships a neuron should have for a complete description fall into multiple areas of concern: simplicity, exhaustive completeness, computability. The concern for simplicity is that the set of relationships a neuron have should not be so large that no neuroscientist is able to contribute knowledge about a neuron because the number of things they need to fill out is daunting and overwhelming. Additionally, if the relations are too simplistic they may not fully describe the aspect of reality they intend to, much like describing a number as being between one and a hundred when the value 55 is really what is needed to be able to sum it together with another value. On the other hand is the concern for exhaustive completeness by those whose concerns are for the utility of the framework – without all the details about a neuron, they worry there will be unnecessary gaps that will prevent valuable insight to be gained. Related to this concern, but still separate is the concern over computability. Here the concern is over the formality of the relations – if formal statements can be used, then more expressive questions can be asked via first order logic operations. For example, those in favor of more formality may take a relationship like has_soma_location and decompose it into a more formal and verbose form that included a has_part

relationship between a soma and a neuron as well as a has_location relationship between the soma and a brain region.

In the results section we have presented the current state of the form NeuroLex presents to users when defining the properties of a neuron. This is the result of several revisions in collaboration with the NIF project and the INCF Program on Neuronal Ontologies.(Grethe, 2009; Martone et al., 2010). Several considerations went into the development of the current list. An initial version adopted the more extensive recommended property list proposed by the Petilla convention for interneuron properties (Ascoli et al., 2008). While this provided a fairly complete overview of relevant properties, users reported difficulty making contributions to the NeuroLex with the number of properties was too high or when the values to be supplied were unclear. This highlighted the reality, as demonstrated by the Petilla convention, that even when defining a single neuron, different potential contributors to the knowledge base will disagree on the importance of certain properties, and the values they should have. In pursuit of a more parsimonious set of properties, the current list was produced to include more multiple choice options and fewer numerical values (e.g. Cell soma size, Firing rate). While the NeuroLex has taken the approach of setting the template for all users, a new effort to create a Neuron Registry has explored the possibility of allowing users to flexibly define and use properties as needed (Hamilton and Ascoli, 2010). While both approaches have their pros and cons, data on which approach is best is still limited.

4.5.5 Organizing structured knowledge online

Wikipedias approach to exposing structured knowledge has been through info-boxes that are implemented via mediawiki templates. The DBPedia project has mined these info-boxes to package the knowledge within them into RDF. With NeuroLex, we have circumvented the need for this two step process, first to an info-box, and second to RDF, by building on top of Semantic MediaWiki. The use of Semantic Forms by NeuroLex eliminates the need for the user to learn wiki-text syntax, as Wikipedia editors must. Users are always presented with a form rather than a block of text when they click the edit button in NeuroLex, unless they

specifically request otherwise .

In addition to being simpler to learn, NeuroLex facilitates a more interconnected knowledge base via two key features: redlinks and autocomplete. When a page is edited and saved, values associated with the concept on that page may appear in one of three colors, black, blue or red. Black text indicates a value that doesnt link this is usually reserved for values or qualities that are provided in free text. Blue text indicates an active link to another concept also stored within NeuroLex clicking this link will take the user to this concept, thus enabling discovery of related concepts. Red text, referred to as redlinks, show a value that has the potential to link to another concept within NeuroLex, but for which that concept does not currently exist. A redlink can occur as a result of a misspelling the concepts name as supplied by the user, or the concept could be missing from NeuroLex altogether. In the first case, the red text alerts the user that they may have made an error, and may need to edit the value further to cause it turn blue. In the second case, the user may click on the red text to arrive at a blank page where they can define the missing concept.

In order to reduce the frequency of a user misspelling the name of a concept, we utilized auto-complete features of Semantic Forms that will present the user with a list of possible options drawn from the NeuroLex knowledge base that match the sequence of letters the user has typed into a field. In addition to enabling the user to fill values in with potentially fewer key strokes, this enables a user to have confidence that the value they have entered will link to another value in the knowledge base, and thus be presented with blue text once the content is saved. This provides immediate feedback to a user that they have improved the knowledge base – rather than have their bit of knowledge sit alone and by itself, they have simultaneously added greater organization to the framework as a whole.

4.5.6 Cerebellum Reasoning Example

In order to demonstrate that the structuring of knowledge in the neurosciences could help to provide answers to questions that were not obvious, we exported the RDF graph of NeuroLex into a triple store environment that allowed

us to pose a specific query (Table 4.6) and retrieve results that were pulled from the pages of the wiki (Table 4.7). We have exported this content outside of the Semantic MediaWiki framework in order to take advantage of advanced query operations and additional computational performance only available in a separate installation of a SPARQL 1.1 compliant triple store (OWLIM).

Three efforts to enable users to aggregate and investigate connectivity relationships are related to this investigation: the CoCoMac resource Kötter (2004), the Brain Architecture Management System (BAMS Bota et al. (2005)), and the ConnectomeWiki . These resources can be compared from the perspectives of access control, style of representation, and ability for external programmatic interface. Both the CoCoMac and the BAMS approaches use carefully curated and restricted access approaches to including new connectivity statements in their database. Neither of them allows users to contribute to the knowledge base without a vetting process. In contrast, the ConnectomeWiki project took a similar approach to NeuroLex in terms of access control. Both allow anyone with a user account, available for free without special permission, to make edits to their database of connectivity statement. In addition, the NeuroLex allows anonymous users to make edits as well. One different between NeuroLex and the ConnectomeWiki is that the presence of separate pages that capture information about connections in the ConnectomeWiki allowing connections to be individually named and referenced. This is an extremely useful addition that has not yet been incorporated into NeuroLex, although as the example demonstrates this is not a pre-requisite for reasoning about connectivity.

CoCoMac and BAMS have both built their knowledge bases as database schema models, which limit the ability to create open and flexible linked data models (Ruttenberg et al., 2009). NeuroLex and the ConnectomeWiki use a semantic data model. A database schema model is a series of tables with rows and columns, while a semantic data model is a directed graph. Queries on directed graphs can be written using the logic of the properties that make up the edges of the graph and the question in mind, rather than the logic of a database schema via SQL. Using the logic of the semantic data model is useful for building a query

up through an interactive process of query and exploration by the user. Graph queries also make some operations more intuitive to work with, such as the ability to recursively search through all subclasses of regional part of cerebelluman operation that is much more verbose to express using SQL. Using the same formalisms, we have also been able to answer questions about how complete the knowledge base is (Supplemental Table 1) for neurons and to reclassify neurons by their neurotransmitter.

Here we have asked for the names of all brain regions that send axons into the cerebellum. The query takes advantage of the "AfferentProjections" and "EfferentProjections" properties, the "SomaLocation" and "LocationOfAxonArborization" property, as well as more standard relationships such as "subClassOf" and "isPartOf". We specified "the cerebellum" as all the parts of the cerebellum that can be reached via the "isPartOf" relationship, which numbered 65 different areas. We found two major advantages to executing this query on NeuroLex as opposed to BAMS:

1. **Explicit representations of axons as connections between brain regions**: The system has reported to us the names of the pathways that conduct the axons into the cerebellum. BAMS considers connections between brain regions as point-to-point. BAMS contains lists of cells that reside in brain regions, but does not contain a place for knowledge that a cell may have its cell body in one brain region and its axon in a different one. NeuroLex does allow these kinds of statements to be used, and they contribute to the query answer here.

2. **Accessibility**: The BAMS connection graph is not publicly exported and cannot be imported into any tools and explored flexibly the way that SPARQL enables. The BAMS database has been made accessible to outside users by permission with specific groups but not publicly available. Some of its content has been exposed as OWL/XML (Ruttenberg et al., 2009) but this leaves out the connectivity statements. The CoCoMac database is available for use upon request of the site maintainers. The ConnectomeWiki is no

longer online, but its contents were available as an RDF dump (private communication with S. Gerhard). Of these other sites, NeuroLex alone exposes 100% of the data contained within its wiki pages in an RDF file and a triple store that are anonymously available to any user who wishes to use them[17]. The accessibility of this content to external interface is important because of the desire of others to re-analyze the underlying data, or display it in a way that is not possible by the given interface.

This example demonstrates the underlying motivation for the effort we have put into NeuroLex. The purpose of aggregating knowledge in a structured manner is to make it possible to ask questions that shed new light on facts we have already acquired. Because of the inherent complexity of the nervous system, it is difficult for any single investigator to keep the totality of its understanding in mind at once. This has necessarily led to specialization in the neurosciences and a balkanization of knowledge, which makes questions that cross between specialties hard to answer. With the NeuroLex, we have provided a system that can allow many different investigators to be aggregate their specialized knowledge into a form where questions that cross their specialties can be answered. While each individual fact may already be known, in the aggregate, and with advanced query mechanisms, it is possible to put into the hands of curious investigators the tools to ask and answer precise questions they are looking for, rapidly and without the work of aggregating the knowledge themselves.

This kind of built-in organization based on adding knowledge necessary in the biosciences because of the complexity of the entities we are trying to describe. Due to the facility to interlink knowledge into a graph, this kind of system can act as a counterweight to complexity– every piece of knowledge does not stand on its own, rather it creates a more parsimonious explanation. The features we have described here appear to us to approach the qualities of the "magic index card" system described in the introduction.

[17] http://neurolex.org/wiki/NeuroLex_SPARQL_endpoint

Chapter 5

The Whole Brain Catalog

5.1 Introduction

The function of the brain is intimately tied to the precise details of its structure. From the gross anatomical level, to the network level, to the cellular level, to the subcellular level, to the molecular level, the details of nervous system structural organization (i.e. neuroanatomy), are crucial to forming new hypotheses and solving the mysteries of the brain. A wide array of imaging technologies have been developed that give scientists powerful ways to see the detailed structure of the nervous system. While this has led to images that reveal aspects of the organization of the nervous system from nanometers to centimeters a full seven orders of magnitudea challenge has emerged: how to gainfully combine the products of the wide array of imaging technologies that depict the nervous system across scales in a manner that increases scientific understanding?

Before we can begin to answer this question, it is important to understand the way neuroanatomical atlases have been used in this area. Because images in neuroscience are derived from brain tissue, neuroanatomical atlases have provided a natural framework for putting image-based data into the context of brain regions. These atlases were traditionally composed of 2D plates of Nissl stained tissue overlaid with delineations for major brain regions and often provided a standard coordinate system (e.g. Watson, 2005) The publication of atlases in this manner allowed other studies to describe their data sets using these coordinate

systems, whether they were using magnetic resonance imaging (Cai et al., 2006; Cao, 2006; Valdés-Hernández, 2010), block face serial sectioning (Gustafson et al., 2007; MacKenzie-Graham et al., 2004), or light and electron microscopy (Martone et al., 2004).

The revolution of digital computers has allowed the scope of the brain atlases to expand beyond a reference system limited by a physical publication to desktop or web-based interfaces (e.g. MacKenzie-Graham et al., 2004; Ma et al., 2005; Lein et al., 2007) . The approach of using computers to capture knowledge in digital spatial frameworks is one of the aims of the field of neuroinformatics (Roysam et al., 2009) and the methods that go with them are collectively called digital neuroanatomy (Maye et al., 2006). Several examples of computerized atlas systems for digital neuroanatomy include a computerized atlas of rat brain precerebellar system (Brevik et al., 2001), MBAT (Boline et al., 2006), Rodent Workbench (Hjornevik et al., 2007), High-resolution 3D surgical atlas (Chan et al., 2007), and the Allen Brain Institute's BrainExplorer (Lein et al., 2007). Many of these tools both display an authoritative reference brain atlas and enable users to register their own data into the context of that atlas. Template or coordinate-based registration can be used and enables the atlas system to be used as a database and computational tool for comparing structural and functional features. (Fox et al., 2005; Lein et al., 2007; Boline et al., 2006; Martone et al., 2008a; Bertrand and Nissanov, 2008; Rybak et al., 2010; Gerhard et al., 2011). Much of these data are in the form of 2D and 3D images, showing the topography of cell distributions, gene expression patterns or physiological signals.

The above efforts start at the level of gross anatomy, defining the nuclei and white matter tracts of the nervous system. While these delineations are often based on cellular stains, these atlases have not traditionally captured the details of cellular morphology. A different approach to addressing the challenge of multi-scale integration in digital neuroanatomy has centered methods around the nervous system cell rather than with the delineations of brain regions as in traditional atlases. These include L-Neuron and ArborVitae for dealing with morphologies of neurons (Ascoli, 2002), NeuroConstruct Gleeson et al.

(2007) for integrating morphologies with computational simulations, NEOBASE, a database for neocortical microcircuits (Muhammad and Markram, 2005), Neurolucida (`http://www.mbfbioscience.com/neurolucida`) for neuronal reconstruction from primary data, and the LONI visualization environment (Dinov et al., 2006). In programs like these, the neuron stands in isolation or in reference to a few synaptic partners. Thus in large part, digital atlases are divorced from tools and repositories focused on the morphologies of neurons. Notable exceptions to this general rule can be found in the invertebrate community where researchers can identify individual neurons and there are fewer of them, e.g. the Honey Bee Standard brain (Maye et al., 2006; Kußet al., 2008; Rybak et al., 2010).

In order to make relationships between images explicit and machine- processable, and therefore available for increasing scientific understanding by asking deeper questions via structured queries, several projects have used methods to connect image data within digital atlases with ontologies or less formal semantic technologies. Work done in conjunction with the Cell-centered database (CCDB; Martone et al., 2007) has revealed several methods to connect the image data it warehouses with ontologies such as the NIF standard ontology (NIFSTD; Bug et al., 2008. The CCDB is an online repository that focuses on organizing complex image data types from neuroscience in order to make those images easily discoverable. Linking annotation of images in the CCDB with a formal ontology enables ontology-driven knowledge environments to be built that allow images to be referenced by the biological entities that are depicted within them (Fong et al., 2007; Gupta et al., 2007). These knowledge environments include tools for annotating images with terms from the ontology, which facilitates data discovery and synthesis of those images in an information system. Kußet al. (2008) has explored the use of ontologies in brain atlasing of the Honey Bee for the purposes of visualizations in hierarchical brain structures. Other projects have incorporated ontologies or terminologies of a related heritage to the NIFSTD, including the BIRNLex(Bertrand and Nissanov, 2008) and NeuroNames (Bowden et al., 2007). Richardson et al. (2010) has implemented their own extensive EMAP anatomy ontology and (Gerhard et al., 2011) draws from data stored by connectome.ch, but does use the

practice of referencing a URI.

Several efforts across computational neurobiology have looked towards methods in digital anatomy to help achieve their vision of richer, more detailed, and more spatially realistic computational models (Hines and Carnevale, 1997; Maye et al., 2006; Czech et al., 2009; Gleeson et al., 2010). This kind of modeling, sometimes referred to as multiscale modeling in the neurosciences, seeks the integration of features of biological reality such as a voltage-gated channel can be integrated with neuronal morphologies as well as gross anatomy (Hucka et al., 2002; Markram, 2006) in order to build predictive models of neuronal function. This practice is familiar to the computational physical sciences where molecules are represented at different levels of resolution in order to make predictions of about their function as individuals or in aggregates (Horstemeyer, 2009; de Pablo, 2011). In the computational neurosciences, neurons are also represented at different levels of resolution, allowing some models to focus on the activity of internal cellular processes (e.g. Coggan et al., 2005; Lopreore et al., 2008 and others to focus on the activity of networks (e.g. Gleeson et al., 2010). This approach is also used in dynamic models of the heart, which use models of heart cells at different resolutions in concert to make predictions of how function at a macro scale is affected by changing parameters at a micro scale (Campbell and McCulloch, 2011).

5.2 Challenges

Insofar as predictive models aid scientific understanding Hartmann (1996), the production of multi-scale computational models in the neurosciences provides a strong motivation to integrate image data across scales. Multi-scale models like the ones described above, however, are frequently challenging to build due to their typically large number of parameters, whose initial values must be constrained in a manner that can be justified by measurements from the tissue being modeled (e.g. Coggan et al., 2005). Faced with this challenge, it is not unimaginable that sufficiently sophisticated tools in digital neuroanatomy could greatly accelerate the production of such models by enabling scientists to mine relationships out of image

datasets that can only be realized when multiple modalities and scales of data are juxtaposed in a common reference space (MacKenzie-Graham et al., 2004).

As the representative array of tools relevant to digital neuroanatomy in Supplemental Table 2 demonstrates, there are now a significant number of tools in this area with overlapping features and practices. As the table also shows, there are several gaps in current tools between features that make it more difficult for a researcher to use them together for building detailed models. For example, the ability to interrogate connections between brain regions is essential for an information system managing information of the nervous system, but rarely overlaps with the support of display of 2D histological or MRI sections, (except for Bakker et al., 2010; Rybak et al., 2010). Another example is the absence of crossover between tools that incorporate 3D neuronal morphologies in the context of brain region boundaries with tools that allow simulation results to be incorporated from a computational engine. Gene expression is now available for incorporation into other tools via the Allen Institutes Gene Expression Atlas (Lein et al., 2007), among others, but there are no tools that are able to incorporate these gene expression patterns with 2D or 3D neuronal morphologies. Thus, for example, if a researcher wanted to determine how the terminal fields of hippocampal CA1 pyramidal cells overlapped with the gene expression pattern for axon guidance molecules such as ephrin ligands, there is no easy way to perform this operation. In each of these cases, there are rich sets of data within one scale or modality that have not yet been integrated across those modalities to allow novel analyses or to form the basis for computational models.

Moreover, while many of these tools provide rich data sets or functionality, only 7 out of 18 of them expose web service APIs that could be used to aggregate their data or functionality into integrated interfaces (Fielding and Taylor, 2002). Fewer still2 out of 18expose anonymous access to their source code. This creates a significant barrier to entry to future investigators who wish to pull these technologies together or reuse code methodologies in distinct systems.

In order to address the gaps in the current array of methods available for building multi-scale data integration into predictive models in the neurosciences,

we have built a computational system, the Whole Brain Catalog [1], whose purpose is to provide a general framework in which to aggregate observations of structure and function of the nervous system across scales. We have built it to take advantage of the connectedness of the internet and enable data to be exchanged in an open collaborative framework that is designed for data reuse. These aggregated observations, in the forms of images and models situated in a common spatial and semantic reference space, can then be made available to computational methods and algorithms suitable for the production of more spatially realistic multi-scale predictive models in neuroscience.

5.3 Methods

The overall architecture of the current WBC (Fig 5.2) has been designed in an object-oriented manner according to the best practices recommended for the development of scientific software (Baxter et al., 2006), both in terms of the design of the initial object model, and through the application of existing data standards where possible.

Fig 5.1 shows a simplified introduction to the underlying object model. The object model was implemented in an XML schema (the WBC master schema), which was then cross-compiled into Java objects that were used in both the WBC client and server, and were also made available for Python scripting via Jython. The schema was also compiled into a MySQL database schema that became the basis for the database used by the server.

The open source 3D game engine JMonkeyEngine (JME[2]) was used as a foundational library for 3D visualization and rendering. JME was chosen because it is Java-based and has a large and active open source development community. It is built using well-known design principles for 3D game engines (Eberly, 2000). Geometrical meshes from the Allen Brain Institutes Brain Explorer tool (Lau et al., 2008) were converted into .OBJ format. JMEs native support for popular formats for 3D meshes such as.OBJ enabled us to render these easily. JMEs application

[1] http://wholebraincatalog.org
[2] http://jmonkeyengine.org

programming interface (API) was used to write new code to render neuroscience-relevant data in 3D in a variety of formats. An optimized NeuronCloud renderer for the NeuroML format was designed, tested and implemented that is capable of graceful degradation of neuron morphological detail, enabling the simultaneous rendering of high detail views of 1000 neurons and low detail views of 100,000 neurons simultaneously. JMEs native support for raster-based images such as .JPG and .PNG was used as a foundation to build an importer for 2D images, either stored on the local machine, or served by a tile-based image server. Two dimensional images are displayed as textures rendered onto a plane, which can be independently positioned in the 3D space.

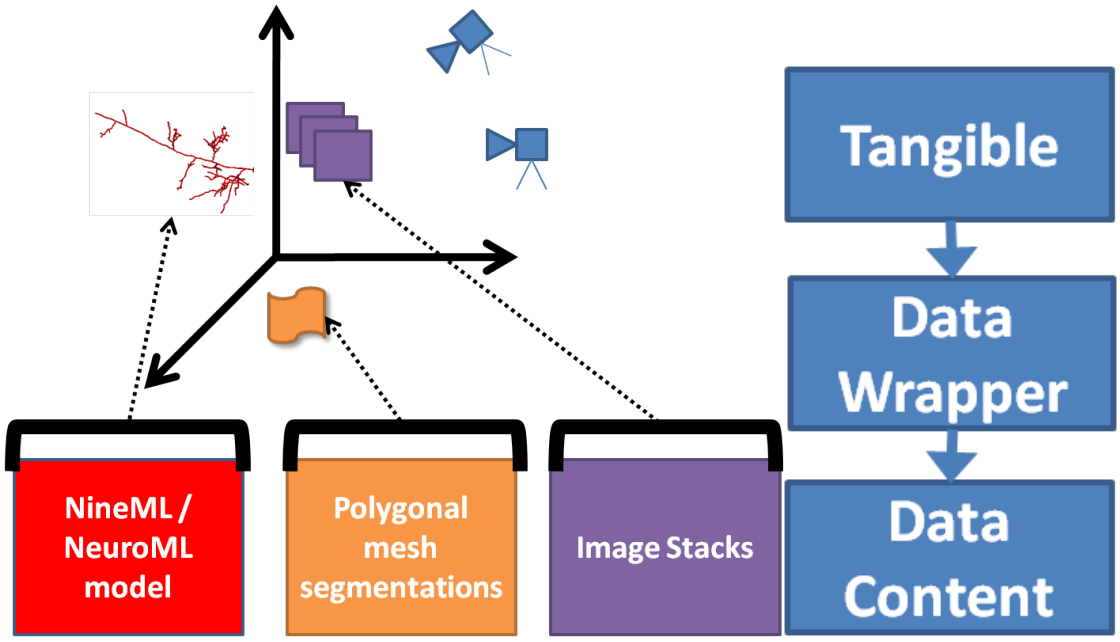

Figure 5.1: A conceptual overview of multi-scale data integration with respect to a coordinate system and three fundamental abstract object types in the Whole Brain Catalog.

Figure 5.1 (left) shows a conceptual overview of the multi-scale data integration enabled by the Whole Brain Catalog. The three axes define a space in three dimensions, into which we want to be able to view different types of data in a unified way. A 3D game engine allows the developer to define a three dimensional

space, populate it with objects, and then draw two-dimensional projections of the 3D space to the users screen based on the variable position of a camera object (in blue). Different data types need only be converted into a form that the game engine knows how to draw into the space in order to integrate them. Once this is established and the objects have been put into comon register in the space, the bounds and shape of the data can be ingested by methods from computational geometry to measure and analyze their spatial relationship.

Figure 5.1 (right) shows the relationship between the core abstractions of the system. The Tangible object holds the x,y,z coordinates of an entity that can be rendered on the screen. Tangible objects are kept separate from Data Wrapper objects, which serve as a level of abstraction between the raw content of a particular data set and its location in the integrated space. The Data Wrapper acts much like an envelope that can accept many different kinds of documents (Data Content), but presents a uniform external appearance to the post office (Tangible objects). Both Tangible objects and Data Wrapper objects are designed to reference metadata. These objects are defined in an XML schema that is then compiled into Java objects.

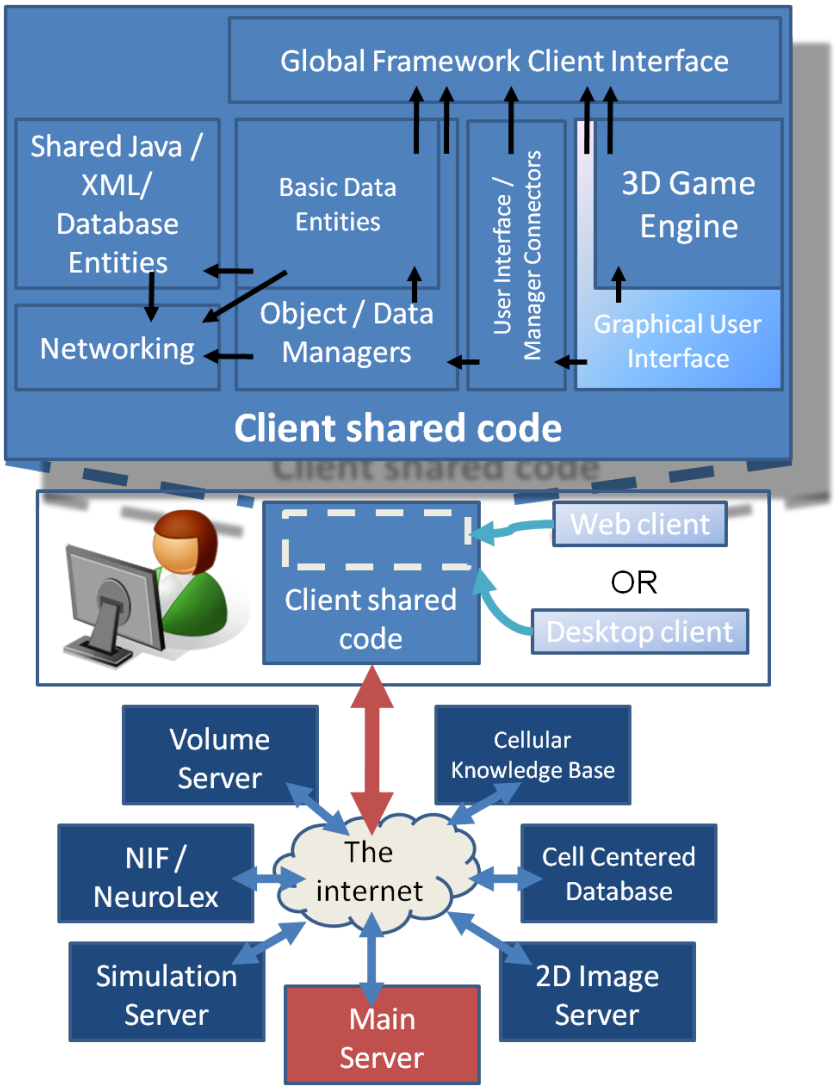

Figure 5.2: A system diagram for the Whole Brain Catalog.

Figure 5.2 shows an overview of the WBC architecture. The box at the top shows the components of the client. A 3D graphics engine renders the 3D spatial data. A reference atlas and coordinate system manages the locations of data. A tangible manager is responsible for keeping track of the visible objects, managing their visibility and selection. An interface manager is responsible for transforming user input into appropriate actions, such as selecting and manipulating objects in the environment. The box at the bottom shows the components of the server. A spatial registry is responsible for keeping track all objects that the client can

possibly make visible. A metadata component is responsible for tying the visualizations of data together with references to their sources, their logical relationships, and their positions. A digital resource repository is responsible for warehousing synthetic views of data that may be produced by users. The client also communicates with several external services including NeuroLex.org, the cellular knowledge base (Fong et al., 2007), and the image warehouse at the Cell centered database (CCDB)

Two different user-facing clients were built reusing the same modular architecture based on the model-view-controller design pattern that deliberately separates code libraries devoted to presentation from code libraries devoted to application logic. A desktop client was built as a Java webstart application, meaning it acts like a fully installed desktop application but can be loaded from the web browser. The desktop client has the JME engine embedded into a panel making use of the Java Swing libraries for 2D menus, textboxes, and tree widgets. A web-based client was built making use of the Google Web Toolkit-based Vaadin framework (`http://vaadin.com`), replacing the view of the desktop version but leaving the model and controller mostly intact. This avoided the need to rewrite the logic that interacted with the server or which managed the client interface.

Client logic was implemented that managed the creation of an option menu available upon right-click of any object in the 3D engine scene, a primary way by which users make updates to visible objects. Logic and interface elements were developed that allows visible objects to have their metadata as well as their position, rotation, scale modified through the graphical interface. Object-oriented modules were developed using a manager pattern that kept Tangible objects in lists and ensured that they were kept in correct states despite user interaction and network loading latency. The client was enabled to pull data sets directly from the CCDB, (`http://ccdb.ucsd.edu`), a web accessible database hosted by the National Center for Microscopy and Imaging Research (NCMIR), through the use of a special code library that exposes its data via a web service. The client was enabled to pull time series datasets from a simulation engine service by means of a custom data library, in a format incorporated into the master schema. By implementing a

3D movable cursor and by reusing the Allen Brain Explorers geometric meshes of brain regions, a facility was added to launch the Allen Institutes Gene Expression Atlas web application with the 3D coordinates asked for within the WBC.

The environment has been engineered in a client-server paradigm, supported by a set of REST web services(Fielding and Taylor, 2002). Both the client and the server have had their networking component built on top of the Restlet library for client-server interaction, implementing handlers to respond to GET, PUT, POST and DELETE commands for various resources including those handling data for Tangible objects and its subtypes, DataWrapper objects, Content objects, and User objects. The client and server were implemented to pass data in an XML format defined by the master schema containing the WBC object model.

The Whole Brain Catalog main metadata server was designed, tested, and implemented. Its logic was implemented to enable retrieval, update and deletion of metadata related to the entities that are situated into space together. The server was implemented to store the information about these entities in a MySQL database via the Hibernate library for server-database interaction. The set of tables, columns, and relationships within the database were autogenerated from the master WBC schema using the HyperJAXB3 open source library.

To demonstrate the utility of the framework, publicly available datasets were situated within it. The prototype has been populated with, 40 2D images, 1000 cell models, 13 subcellular EM reconstructions, and one molecular model obtained from publicly available data collections within the Cell Centered Database (CCDB), NeuroMorpho.org and the Protein Data Bank (PDB).

A client to the INCF common reference space for spatial integration of mouse brain data, known as the Waxholm space (WHS; `http://waxholm.incf.org`) was implemented to allow the exchange of spatial data and for translating among different atlasing systems developed for the mouse brain.

A client to the open source MapServer platform employed in the geospatial community was developed. The client was built to request small chunks of large images hosted on a MapServer client and a special data loader was built to reassemble and render them in the 3D engine. To enable this, the DataWrapper

object was extended to reference an image service, rather than a static file. A workflow was developed that allows the NeuroML to be pre-processed and converted into a format that the NEURON simulation engine (Hines and Carnevale, 1997) can use, transforming it so that it can take advantage of multiple CPUs (Fig 5.3). The model arrives at the supercomputer, is placed into a queue, and returns time series results after calculation. These results are then packaged using a time series standard format we have created, shipped back to the client system, and visualized.

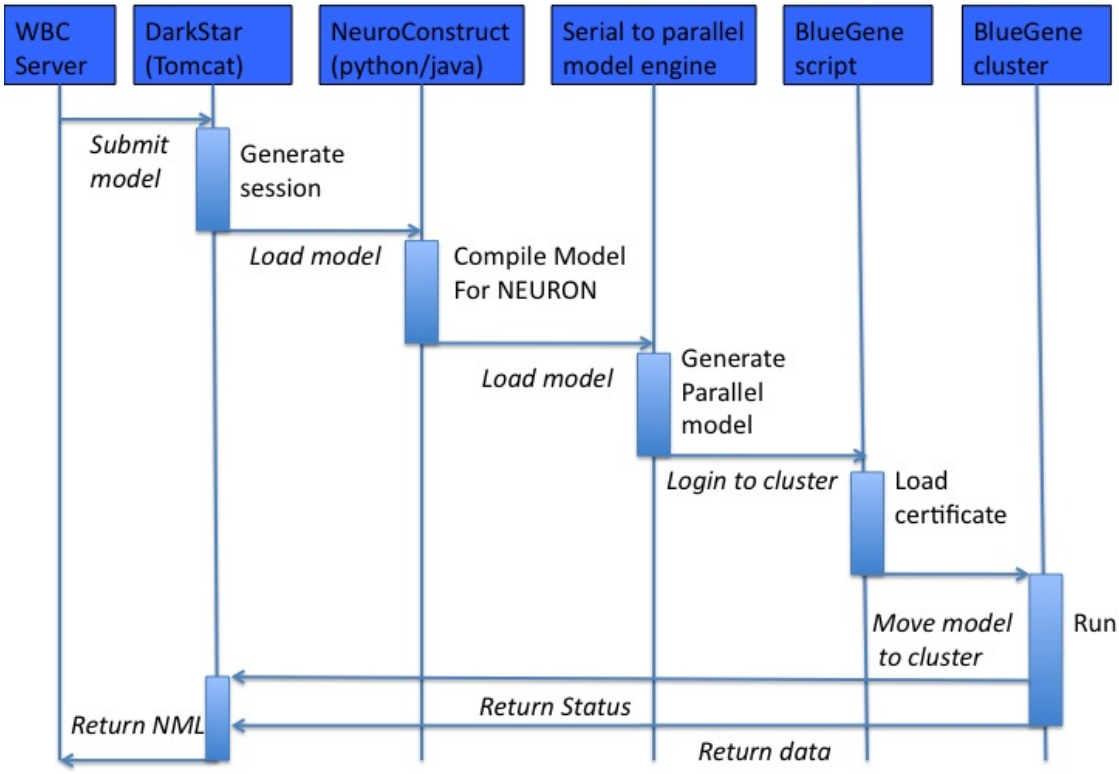

Figure 5.3: An interaction diagram of the automatic workflow between several applications and servers to take a simulation from the WBC through processing on shared resources and return the results. The WBC server submits a model according to user specification to a server we call DarkStar. That generates a session and uses the NeuroConstruct tool (to transform the NeuroML based model into a form that NEURON can use. The single CPU model is then parallelized. Finally the parallelized, NEURON ready model is passed to the Blue Gene supercomputer. While in queue or in execution, status can be checked. Finished results are returned in our time series format back to the WBC server for visualization.

To enable extra analysis on top of the Whole Brain Catalog object model (Fig 5.1), the core object model used by the client was released as a separate library. This library was combined with the Jython scripting language to enable cross- Java / Python code interface. This allows analyses that can wish to take advantage of mathematical or scientific libraries written in either Java or Python

to be performed.

The Whole Brain Catalog has been developed using an open source development model (Fogel, 2005). All of the source code is hosted at a free Google Code revision control repository and this repository serves as the authoritative storage place for the entire code base. This site also hosts a task list of features to be developed and bugs to fix that are referenced upon commits to the version control repository. To minimize errors that would prevent the source code from building appropriately, a continuous build integration system (Atlassian Bamboo) checks out, builds, and verifies a test suite of code performs correctly upon each update the code base. Documentation has been posted for users of the Whole Brain Catalog on a MediaWiki-based wiki site hosted on the web[3]. To engage the questions of potential users and developers of the system, a contact e-mail address was created, an open forum hosted at Google Groups has been created, a Twitter account has been created (@BrainCatalog) and a Facebook page has been established.

To manage the creation of the Whole Brain Catalog software, the Eclipse integrated development environment has been used. The Maven build system has been used for managing builds and dependencies on other code libraries. Unit tests have been implemented using JUnit.

[3]http://wiki.wholebraincatalog.org

5.4 Results

5.4.1 Overview of the system features

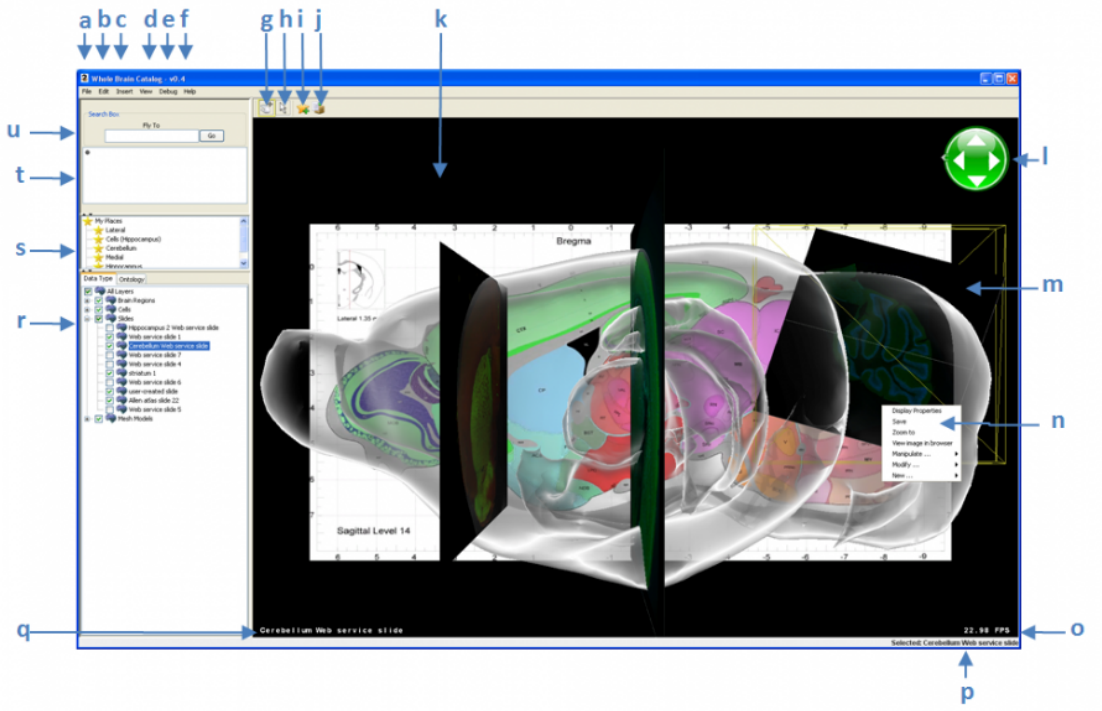

Figure 5.4: The current state of the desktop version of the Whole Brain Catalog, which incorporates the 3D geometry of the Allen Reference Atlas structures of the mouse brain with a few images taken from the Cell centered database (CCDB) and derived from imagery produced at the National Center for Microscopy and Imaging Research (NCMIR). The brain model can be easily rotated, panned and zoomed into to explore its contents. Zooming into the brain reveals further content contained within such as 3D representations of neurons, 3D representations of membranes found in EM tomographs of neuropil, and even molecular models. Letters refer to different features (see online documentation at http://wiki.wholebraincatalog.org for a guide)

The Whole Brain Catalog is an on-line 3D collaborative virtual environment for exploring, searching, and integrating neuroscience data. It includes a graphical user interface through which users may upload, view and manipulate data that is

spatially referenced by default to a reference mouse brain atlas. The 3D visualization interface is responsible for enabling data to be displayed within a 3D space. Because neuroscientists generate data in many different ways, the WBC supports the simultaneous visualization of different datatypes through the interface. The interface is capable of rendering 3D brain regions, 2D images, 3D neuron reconstructions, 3D EM reconstructions, and 3D molecular models, all within the same space at the same time. Users can bring more data into the WBC interface through upload or query of remote databases, display different data types registered to the same space and annotate data using the NIF ontologies. The WBC can also import and display the results of simulation studies performed on geometrical models using simulation platforms such as the NEURON simulation engine.

Two versions of the Whole Brain Catalog client have been created. The first is a downloadable application and runs on Windows, Mac, and Linux machines via Java Webstart. The second is a web-based application that runs directly in the browser, making use of Java applets. It works on all modern web browsers, eliminating a previous need to download the WBC client application and greatly reducing the loading time. This client allows the creation of unique URL links that take a user directly to each data set within the spatial anatomical framework. The web client shares 80

5.4.2 Dynamic change of viewpoint position and magnification

Within the 3D space, the user is visually located at a position, and can change this position through the use of the mouse and keyboard. The default view places the user outside of the brain region models that represent the brain as though it were made of frosted glass.

Users can approach the data via controls on a heads up display or by using first person mouse controls. If a user wishes to zoom in to a part of the brain, data referenced there at multiple scales will appear. Preset locations of interest can be "zoomed to" which gives the visual appearance of "flying through" the brain. For example, in the current system the hippocampus is populated with cell models

of dentate gyrus granule cells and CA1 pyramidal neurons. As the user zooms in, the models display the full dendritic and arbors of these cells, when available, simulating the effect of increasing magnification on a microscope but within a 3D virtual world. In the current version, the cerebellum is populated with a cell model of the Purkinje cell available through Neuromorpho, a mesh model of cerebellar neuropil reconstructed from electron tomography available through CCDB, and a potassium channel model from the PDB. These very different types of geometries are all rendered simultaneously, and appear sequentially as the user zooms into the scene.

5.4.3 Rotation and Visibility of 3D Brain Regions

Taking advantage of the framework provided by the 3D game engine, the user can rotate the brain model around in 3D dimensions (Fig 5.5) . In addition, each of the visible objects within the visible scene can be made visible or invisible independently of each other, enabling a custom perspective on relationships between datasets.

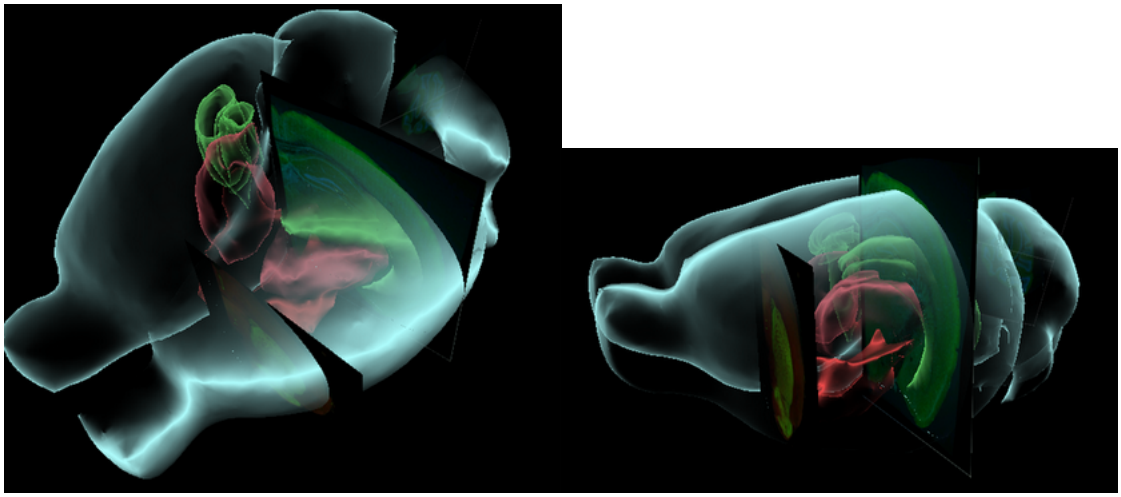

Figure 5.5: Two images from the prototype for the Whole Brain Catalog. At left is a lateral view of the mouse brain with two images cross-cutting the brain taken from mosaic imaging via confocal microscopy. Internal to the outer shell of the mouse brain are some delineated brain structures from the Allen Reference atlas (green: hippocampus, red: Thalamus and Hypothalamus). At right is a dorsal view of the same brain with the olfactory bulb in the bottom left corner.

5.4.4 Import of experimental data into context of atlas

The desktop version of the Whole Brain Catalog allows a user to upload content into the Cell Centered Database after completion of a set of forms. The web version has reversed this practice– data can be uploaded immediately and metadata can be assigned later. When users upload their data, they can see data situated within the brain regions. Most obvious in the default view are the images taken from brain slices that crosscut the brain region models at those places where the tissue is located. The images that are pre-populated in the default view have been manually scaled and warped to fit those brain regions. Many of these images are served from a tile-based image server that enables very large images to be to be displayed by serving fragments on demand.

5.4.5 Dynamic positioning of microanatomy datasets in context of macroanatomy

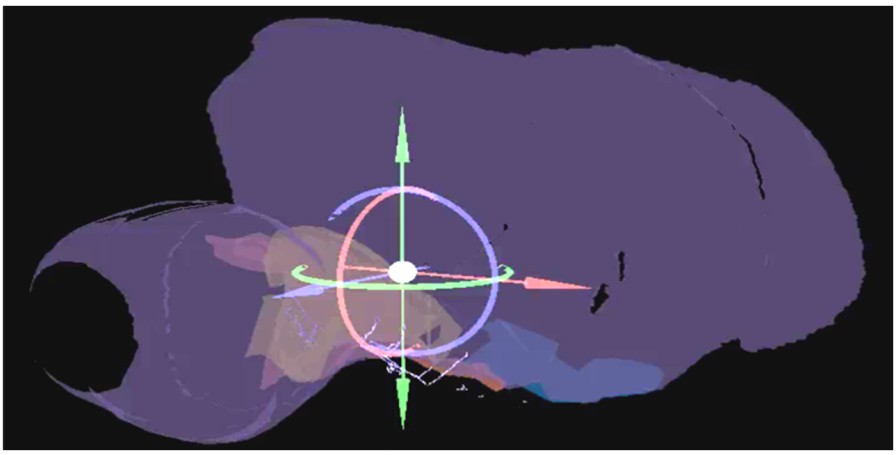

Figure 5.6: Computer assisted drafting-like handles for positioning, rotating, and scaling data within the Whole Brain Catalog

The Whole Brain Catalog includes computer assisted drawing (CAD)-like controls for manipulating data items in the WBC in 3D space much like what can be found in software packages such as OpenWonderland and SolidWorks 3D (Fig 5.6 and Suppl. Movie X). These handles can be used with any entity in the Whole Brain Catalog and can be dragged to cause the object to translate, rotate, or scale in 3D space, thus providing a precise system for object manipulation in 3D space. These controls enlarge or contract as a user zooms in and out of the scene, maintaining a constant size relative to the viewpoint and enabling a user to manipulate entities at any scale. As a user makes modifications to the data objects that are visualized in space, the client requests updates to the positioning, rotation, scale, and other properties of the models from the server processes that manage data, and processes that manage the visualization of those data.

5.4.6 Extending the WBC Framework

The Whole Brain Catalog makes special use of data sets that represent the boundaries of brain regions, using them as part of its spatial framework for

positioning other data sets. Currently, the WBC exposes three reference atlas modes publicly: the Allen Institute mouse atlas, the Waxholm atlas, and a non-authoritative Zebrafish atlas. The mouse atlas was derived from 240 segmented brain regions made available through the Allen Brain Atlas (Lau et al., 2008). In the Whole Brain Catalog, these brain regions have been centered and rotated within the coordinate system inherent in the 3D environment. This enables the geometries of the brain regions to be used as references to the spatial coordinate system created by the Allen Institute with their atlas.

The Waxholm atlas was contributed by the Duke Center for In Vivo Microscopy. We obtained microscopic MRI Zebra Fish surfaces and retinal tract-related optic regions from the UCSD Center for Functional MRI . We have received and incorporated data contributions from Margaret Davis of the NIAAA (a series of two-dimensional hippocampal histological sections). Additionally, we have received data contributions the lab of Steve Fisher at the University of California, Santa Barbara (a mesh derived from retinal ganglion cells, situated in the retina), and Macaque brain surfaces from the University of Washington (not yet exposed through the WBC interface).

5.4.7 Supported data types

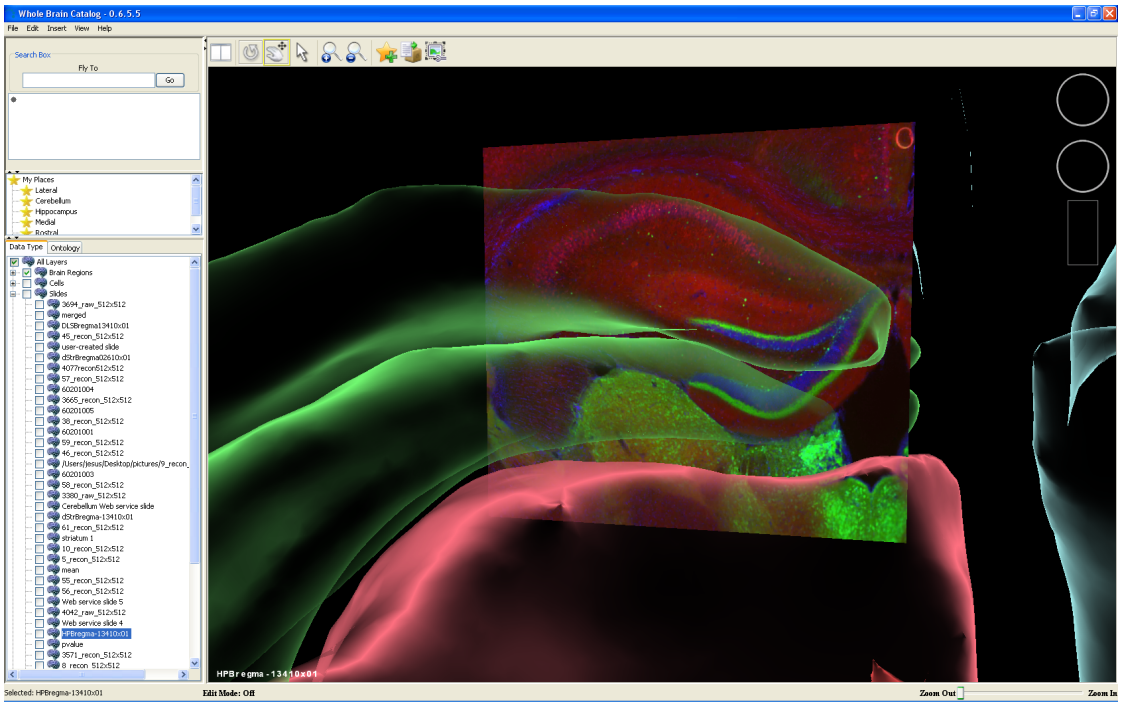

Figure 5.7: A 2D histological section of the hippocampus is displayed situated in the context of the dentate gyrus and hippocampus brain boundaries from the Allen Institutes 3D mouse brain atlas, within the Whole Brain Catalog environment.

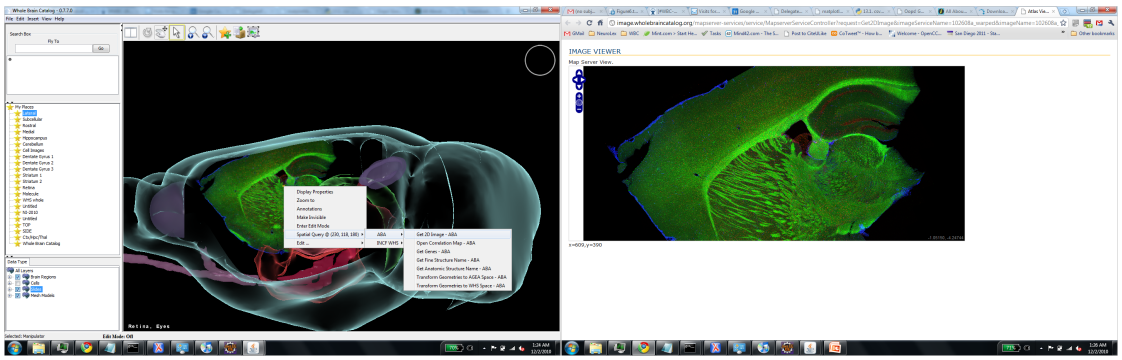

Figure 5.8: Some images within the Whole Brain Catalog are low resolution versions of very large images served on an external image server. These images serve as launch points to a separate web page that allows a user to zoom in and out of the image.

The Whole Brain Catalog can display a two dimensional image in the context of the three-dimensional space of its 3D environment, allowing the 3D boundaries of the brain regions to be brought into alignment with the 2D boundaries depicted in the slice through the brain shown in the image (Fig 5.5).

Two- and three-dimensional images of the morphologies of neurons and glial cells can be visualized in the context of the geometric brain regions and images of tissue slices within the Whole Brain Catalog (Fig 5.4)

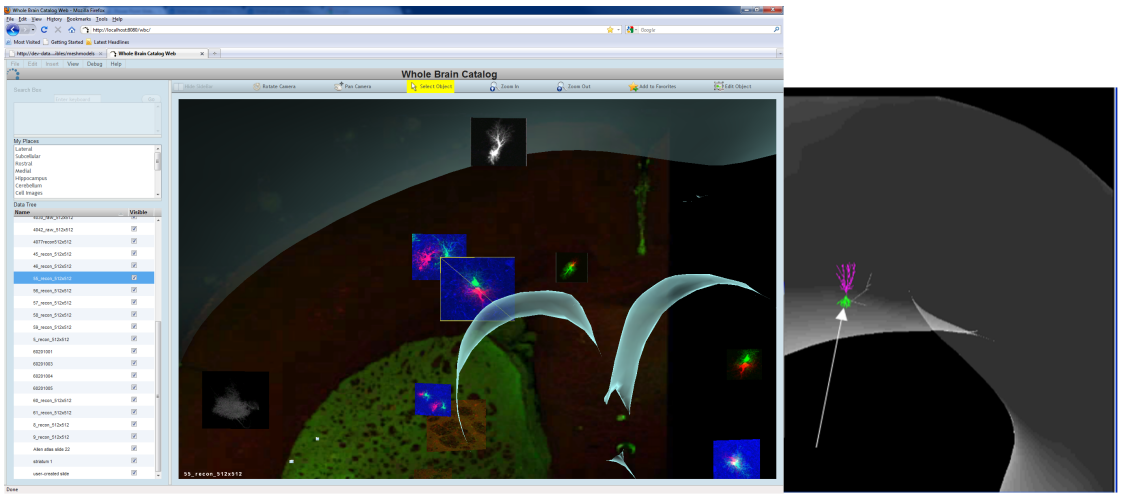

Figure 5.9: Visualization of 2D images and 3D neuronal morphologies in the Whole Brain Catalog. On the left, smaller rectangular images are superimposed into the three dimensional space in a location appropriate to the original position the cell was in when the image was taken, based on manual positioning from the metadata provided for the image. This image takes the perspective of a viewpoint from within the olfactory bulb looking back towards the striatum. Images pictured here originate from the CCDB. On the right, an image of a single 3D morphology of a CA1 pyramidal cell from NeuroMorpho.org positioned within the geometric mesh of the Dentate Gyrus from the Allen Institutes Brain Explorer.

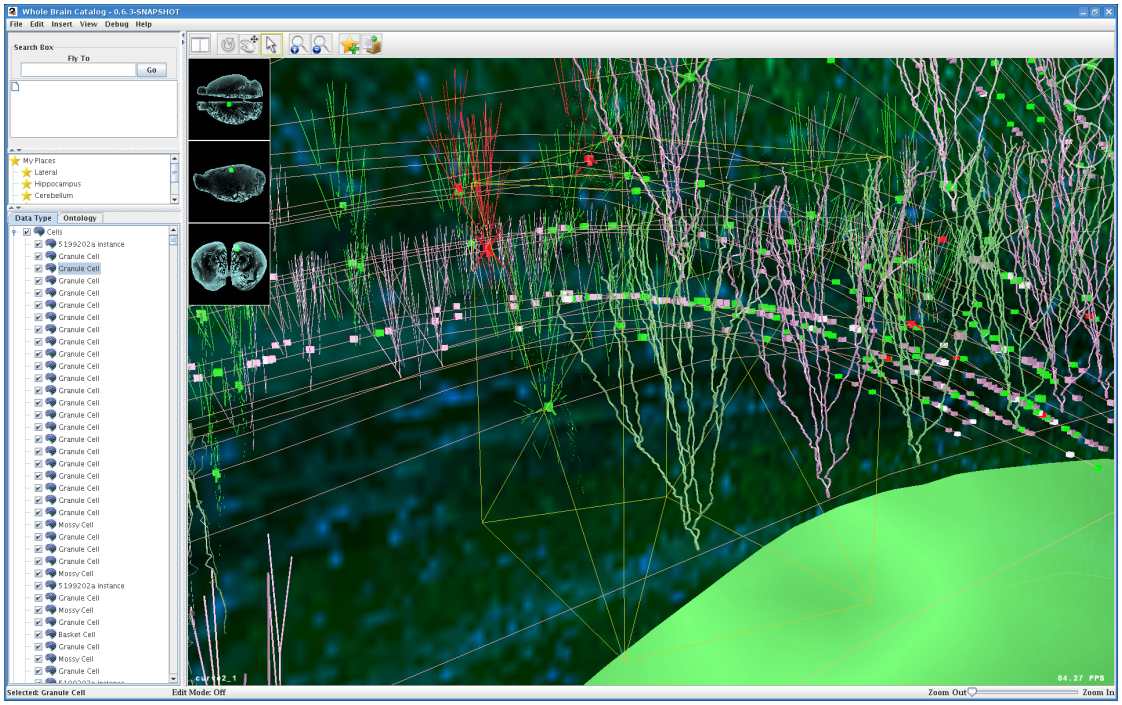

Figure 5.10: Dentate Gyrus granule cells from NeuroMorpho.org embedded within the Dentate Gyrus brain region within the Whole Brain Catalog. In the background is a 2D image of the dentate gyrus. Cells that are green are in the process of growing, based on a time series adapted from Aimone et al., 2009.

The current system has an animation of activity of the neuronal network of the dentate gyrus, composing 1000 neurons of approximately 4000 compartments each, which both spikes and has granule cell neurons that grow into the network and die. This animation is derived from simulation data from the Gage laboratory described in Aimone et al. (2009), contributed to the WBC by the authors. (See supplemental movie 1).

5.4.8 Integration with NEURON simulation engine

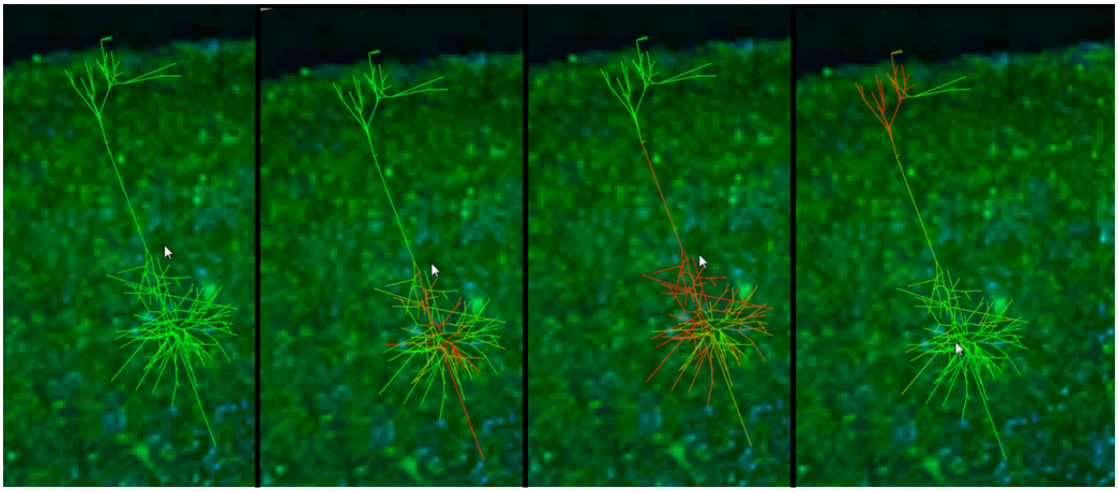

Figure 5.11: Four frames demonstrating the results of a simulation of a multi-compartmental conductance-based model, calculated in NEURON and rendered in real time in the Whole Brain Catalog. This is an animation for a 5000 compartment model of a neocortical pyramidal neuron, composed by Mainen et al. (1995), obtained from Neuroconstruct.org. In the background, a histochemical slice image provides context to the activity pattern as it evolves across layers. The radial extent of the activity pattern across the dendrites of the 3D model can be seen in the context of the rest of the 2D slice image.

The Whole Brain Catalog can display the results of simulations that have been run on 3D cell models described in NeuroML(Gleeson et al., 2010). After a user uploads a model expressed in NeuroML to the WBC, they can elect to have that model processed by the simulation engine resources. A cell model is composed of compartments, and when passed into a simulation engine is able to calculate a time series for the membrane potential of each of its compartments. The interface can read this time series, and for each compartment, convert the value of the membrane potential into a color range from green (-70 mV) to red (+30mV). Therefore a cell model can appear to spike an action potential and the progress of the action potential can be observed as it propagates throughout its dendritic arbor. The current system has an animation for a 5000 compartment model of a

neocortical pyramidal neuron, composed by Mainen et al. (1995), obtained from the Neuroconstruct.org (Gleeson et al., 2007) website (see supplemental movie 2). Additional models that have been imported into the Whole Brain Catalog and run through the simulation engine were derived from the Neuroconstruct.org site, and include a model of the granule cell layer of the cerebellum (Maex and De Schutter, 1998) (see supplemental movie 3) and a thalamocortical model originally created by Traub et al. (2005). These models have been manually situated in space and scaled to the best position possible given the landmark boundaries by the reference brain.

5.4.9 Launch of spatial queries

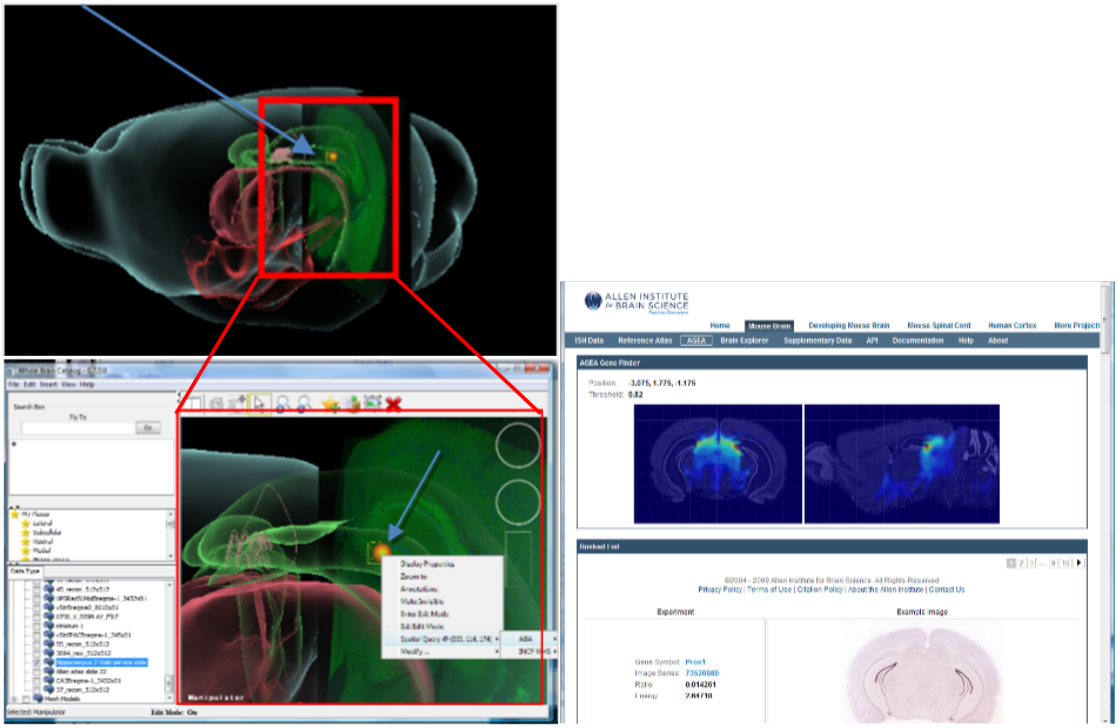

Figure 5.12: A spatial query can be launched from the Whole Brain Catalog from a probe that can be easily and arbitrarily placed that returns data from services provided by the Allen Gene Expression Atlas, pictured above.

Data within the Whole Brain Catalog exist at specific coordinates in 3D

space that are expressed in the units of the Allen Brain atlas voxel space derived from the Allen Brain meshes. We refer to this space as ABA voxel. A spatial query probe can be put down into the space and used as a point of interest (POI) indicator (Fig.X, blue arrow) to launch queries from other information systems that accept POI requests, e.g., the Allen 3D gene expression atlas. Additional spatial queries are available that reuse the INCF Digital Atlasing Infrastructures Waxholm services API (Hawrylycz et al., 2011).

5.4.10 Integration of semantic interoperability

Any data set integrated into the Whole Brain Catalog can have an identifier that refers to the biological entity represented by those data assigned to it from NeuroLex.org, a semantic wiki for neuroscience hosted by the Neuroscience Information Framework. As a result, when a search is run in the Whole Brain Catalog (desktop version only), all synonyms associated with that NeuroLex identifier will be searched. If the keywords being searched match a synonym or part of a synonym, the corresponding dataset will appear in the search results. Additionally, the set of properties that are associated with the NeuroLex identifier such as its definition or its synonyms can be displayed within a native interface inside the Whole Brain Catalog (desktop version only). A button in this interface allows a user to open their web browser directly to the NeuroLex page corresponding to the entity referred to by the identifier. In this way, a user can easily link out to other data sources that contain information about the entity using the NIF Navigator (Akil et al., 2011), or edit the knowledge on NeuroLex.org, such as adding a synonym. Changes made to the NeuroLex are reflected back in the Whole Brain Catalog interface in real time.

5.4.11 Use as a research tool

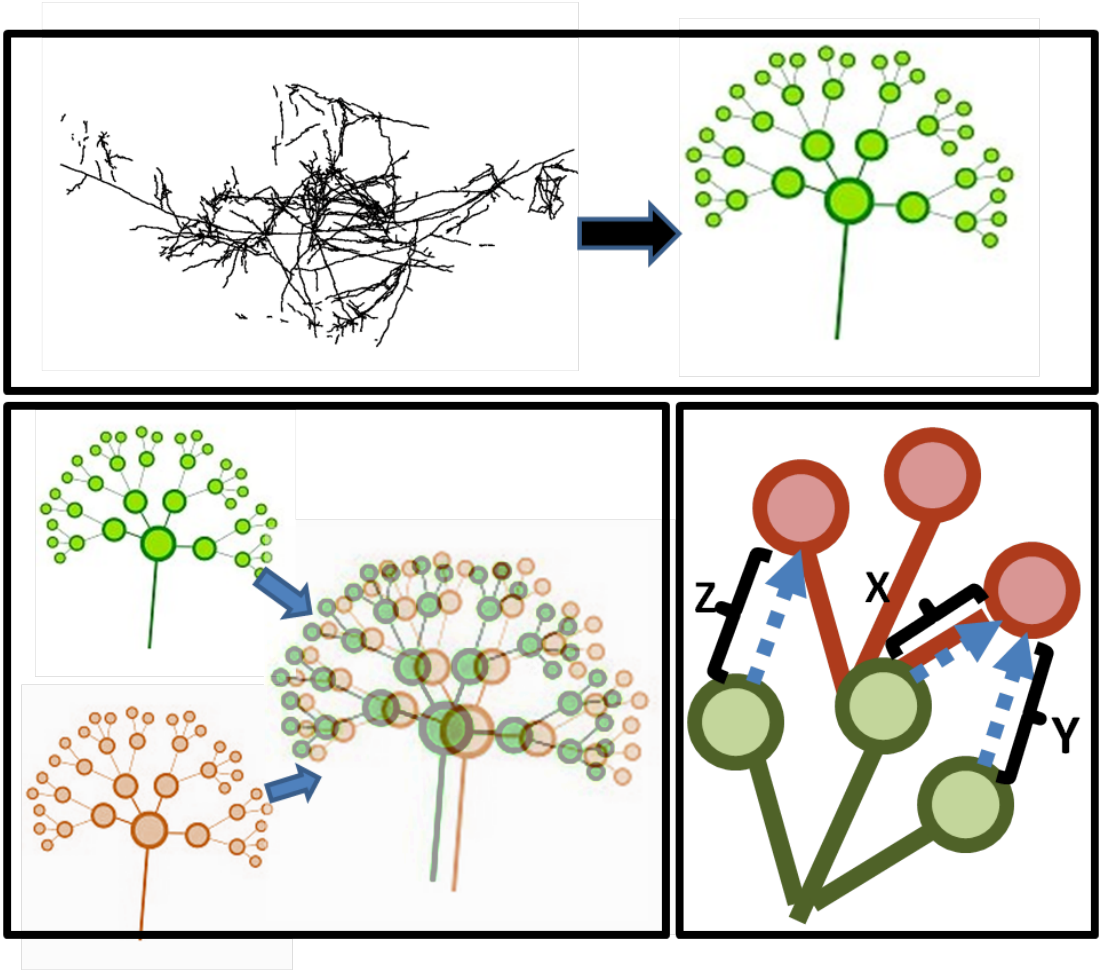

Figure 5.13: Schematic outlining the use of Whole Brain Catalog to perform multi-modal alignment and analysis in a case study involving the olfactory bulb (Ghosh et al., 2011).

Figure 5.13 shows a cartoon representation of multi-modal alignment and analysis process used in a case study involving the olfactory bulb (Ghosh et al., 2011). In the top panel, the black arrow indicates the transformation of the morphology of an olfactory mitral tufted neuron, left, into a cartoon view of a logical tree structure, right, suitable for interrogating parent-child relationships and branch points along its morphology. In the bottom left panel, a cartoon of the near alignment in space of two different neuronal morphologies using the CAD-like

handles described above. (See supplemental movie 4 for a screen capture of this process). In the bottom right panel, a cartoon example of performing an endpoints distance analysis between the two morphologies. For a given endpoint, its nearest (in terms of 3D Euclidean distance) neighbor endpoint in the other morphology is identified. The distribution of distances, X, Y, and Z were analyzed to determine the presence of spatial randomness or regularity with respect to where the endpoints are located in space.

The ability to combine different datatypes and span multiple scales has allowed researchers to ask novel questions about the organization of cells and circuits in the mouse brain. In a recently published study (Ghosh et al., 2011), the Whole Brain Catalog has was used for multi-scale, multi-modal data integration of neuronal morphologies with respect to a reference brain (Figs 5.13 and 5.14). This study used the ability to freely align two-dimensional sections from individual experiments with two-dimensional sections from the reference brain in three-dimensional space. Features that could be seen in the two-dimensional sections, such as the borders between brain areas visible with a Nissl stain, and the outline of the shape of the tissue section, were used for alignment by the use of transparency in the virtual environment, enabling the user to overlay the data set intended for alignment with the reference section (see supplemental movie 4). This, in turn, aligned a set of neuronal morphologies due to their relation to the aligned two-dimensional sections. The aligned morphologies were run through an analysis pipeline built using the Whole Brain Catalog infrastructure (Fig 5.13) that revealed properties of their organization. In addition, this study took advantage of the ability to easily incorporate additional mathematical and graph-analysis libraries from either Java or Python such as NumPy and JUNG with the Whole Brain Catalog (Ghosh et al., 2011, methods).

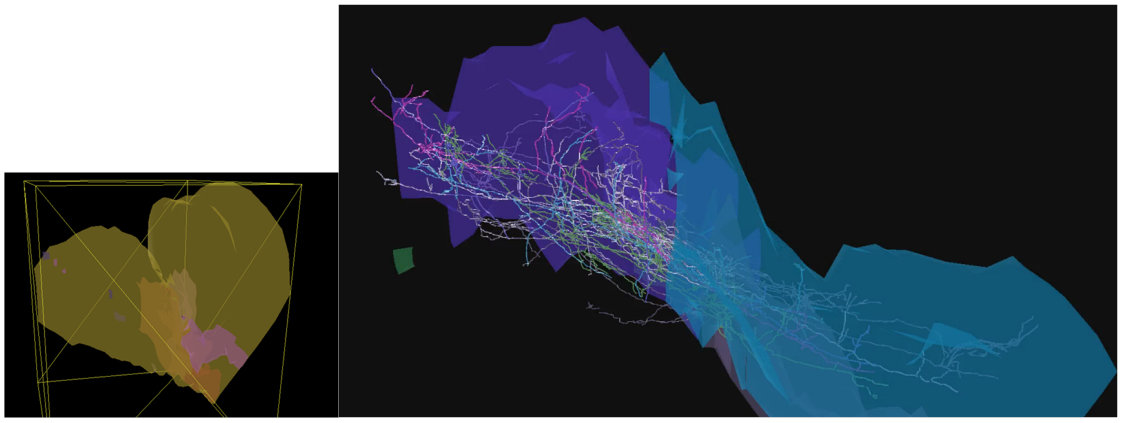

Figure 5.14: Results of the multi-modal alignment in a case study involving the olfactory bulb (Ghosh et al., 2011). Left: the reference brain with olfactory subregions for the olfactory system of a 3 week mouse that was constructed specially for this study. Right: the results of using the dynamic positioning ability of the Whole Brain Catalog to align the olfactory mitral neurons into the olfactory subregions of the reference brain.

5.5 Discussion

We have presented here an approach to address how to gainfully aggregate 2D, 3D and 4D image data sets taken from different spatial scales and by different experimental techniques, to benefit scientific understanding. Current efforts in neuroinformatics focus on addressing a specific need to display, analyze, or integrate one or two types of data together (Fig. 5.1). The approach of these projects is essential to organize and analyze the vast array of data being generated in the neurosciences. However, even the most fully featured of these tools have not engaged in building a general spatial framework for digital neuroanatomy that reaches across all relevant spatial scales and data types necessary to neuroscientists, and that is also designed to be built on and reused in the future. The Whole Brain Catalog, with its 3D engine and its modular architecture demonstrates what is possible with such a framework and point towards the development of even more powerful tools for the organization and understanding of datasets in the neurosciences.

We have applied technologies from computer graphics to the challenge of making complex neuroanatomical structure across spatial scales viewable in the same virtual space. As any student of neuroanatomy knows, images derived from biological tissue reveal irregular and complex shapes. Consider the wrinkled shape of the cortical surface, the roughly banana-shape of the hippocampus or the almond-shape of the amygdala. The use of language in these cases reflects our way of taking a complex shape and analogizing it to a shape that is more intuitive to think about. We recognize, however, that these are intentional simplifications and giving labels to these brain regions leaves out a lot of information we get by examining the raw images. Moreover, with increasing ability to image the shapes and distributions in three dimensions of brain cells e.g. Dodt et al., 2007), researchers are further forced to confront highly irregular and unintuitive spatial patterns and extract understanding from them. By placing three-dimensional signals related to complex brain structure in an environment where a user can position their viewpoint at any arbitrary point, even inside it, we have enabled users of the system to see the spatial relationships of microscale and macroscale neuroanatomy together in the same space, bring them into register, and ask and answer a novel question about neuronal morphology. The use of a dynamic 3D game engine, enabling rotation and zooming through these data sets, enables users to take advantage of the kinetic depth effect (Wallach and O'Connell, 1953), in addition to other monocular cues of depth perception, to translate moving images on a 2D computer screen into a three dimensional perception of neuroanatomical spatial patterns.

The usage of a 3D engine in neuroscience has been shown by other projects (Lau et al., 2008; **?**; Rybak et al., 2010; Gerhard et al., 2011) in neuroinformatics. These projects have focused on visualizing specific types of data, and therefore have created data structures that are specific to their domain. While we have also created domain-specific data structures, we have also dedicated special effort to building a modular architecture that can scale to any arbitrary type of data that would be desirable to render within a virtual environment for neuroscience (Fig 5.2). We have demonstrated the ability to render and spatially integrate two-dimensional images of brain tissue derived from multiple experimen-

tal methodologies, three-dimensional boundaries of brain regions an body parts, three dimensional boundaries of cell surfaces derived from electron microscopy, three-dimensional morphologies of neurons derived from confocal microscopy, and simulation-derived morphologies of neurons constructed to demonstrate function, and display a time series altering the visual appearance of those shapes. Several other kinds of data could be imported into the system following the modular pattern that we have laid out for the existing ones.

In addition to building a system that shows how multiple types of data across multiple spatial scales can be assembled into a single space, we have taken special effort to make it freely available to others on the internet. A significant inspiration for the Whole Brain Catalog was Google Earth, the 3D mapping software produced by Google for performing multi-scale integration of geospatial data sets from a variety of data sources (Satellite imagery, arial photography, street maps, land-based photographs of street views, digital elevation models, underwater seabed mapping imagery, 3D models of structures such as buildings, individual photographs). Because Google Earth has shown that it is possible to build an information system accessible from any network-linked computer that can grapple with spatial data on a comparable order of magnitude to what we would want for the nervous system, it appeared to us approachable to explore using similar methods within neuroscience. Furthermore, the uses of Google Earth by scientists has shown that making data sets freely available in this manner can drive scientific inquiry (Butler, 2006; Burda et al., 2009; Begall et al., 2008) and we believe that this can also happen within neuroscience.

There have already been efforts between the geospatial mapping community and the neuroinformatics community to reuse infrastructure, such as the SMART Atlas (Zaslavsky et al., 2004; Martone et al., 2008a). Some previous efforts have used a volumetric representation of the brain as a starting point Bertrand and Nissanov (2008); Martone et al. (2008a). The lack of ability for raster-based graphics such as volumetric representations to scale to arbitrary resolutions can be problematic. For example, information systems are unable to embed visualizations of neuronal morphologies within an MRI image because the MRI images voxel size

will typically be much larger than the width of a neuron, leaving no resolution to display the neuron. In order to address this problem, we have taken advantage of the vector-based approach of a 3D game engine, which allows arbitrarily small shapes to be embedded inside larger ones.

During the development of the Whole Brain Catalog, we have invested extra effort in following the recommended guidelines for building open source software (Fogel, 2005). The publication of our source code from the beginning has been done with the hope that ideas, snippets of source code, architectural methods, or whole modules built into the Whole Brain Catalog will be used in the future, perhaps in ways that we cannot currently predict. We believe that the value of computational models or tools in these areas goes far beyond what can be captured in a scientific publication, and we encourage others to follow in these footsteps.

We have shown that using a three-dimensional environment for spatial data integration in the neurosciences enables the embedding of neuronal morphologies in the context of the brain regions that contain them. Our first example of this, the dentate gyrus model (Fig 5.10, supplemental movie 2), has a precedent in neuroinformatics (Ascoli, 2002). Our version represents a demonstration of what is possible in building a 3D model of the Dentate Gyrus using neuronal morphologies. The process of constructing this model provided insight and a first step towards the construction of more sophisticated models. We were forced to confront the spatial realities of the complex shape of the dentate gyrus as we placed these neurons together, such as how the dendrites should be oriented with respect to the hippocampal crescent. The morphologies, while derived from actual data downloaded from NeuroMorpho.org, do not reside in the precise locations from which they were derived because this information had not been preserved in the original data, but rather have been manually placed based on the relationship between their cell soma and the granule cell layer. Individual cells have been duplicated in the scene as appropriate for the cell type and region region they are in. The growth patterns that were reproduced were applied as linear scaling of neuronal morphologies based on time series patterns adapted from Aimone et al. (2009). In addition our initial difficulties manipulating morphologies to create the model in

an early version prompted the construction of CAD-like handles (Fig 5.6). Our need to visualize the growth patterns of the neurons prompted us to address the challenge of a data representation capable of representing both spike times and voltage values changing over time. All in all, the exercise proved an important first effort to show what is possible with multi-scale data integration in a virtual environment.

In addition to the construction of models with data-derived morphology, we have demonstrated we can embed existing models built by researchers from computational neuroscience, and visualize their effects (Fig 5.11). This demonstrates the alignment of computational models into the spatial context of gross neuroanatomy, an area that has been called for (Ascoli, 2002). We have manually aligned the two network models we have imported into the space of the Allen Institute reference brain, by scaling and positioning the models based on criteria of brain region landmark boundaries and by comparing their positions to histological sections that corresponded to the cell types in the models. Because the models were not originally designed to reside at precise brain coordinates they can never be perfectly positioned. The value of this demonstration is not in the precision of the positioning of the model, however, but in the potential to computational neuroscientists to develop future models that follow the spatial boundaries of a brain atlas. Using methods we have developed here, a computational model builder can base a model on a set of images in 2D or 3D of a given brain feature. We believe that environments like the Whole Brain Catalog are pre-requisites for the construction of a collaborative computational neuroscience modeling environment that is deeply connected with brain anatomy.

We have shown how a data-integration environment for neuroscience can draw together images and data sets that reside on different data resources on the internet. The Whole Brain Catalog contains data connectors to the Allen Brain Institutes gene expression atlas through the use of a common spatial reference system (Fig 5.12), and the Cell Centered Database through the usage of database identifiers. Rather than focus on being a single repository that must hold onto a copy of all data visible through it, the Whole Brain Catalog has also been designed

to visualize data from other sources (e.g. CCDB) and link out to data held in other sources (e.g. AGEA). Because of the effort to integrate with projects from other groups, The International Neuroinformatics Coordinating Facility (INCF: `http://incf.org`) Digital Atlasing Task force leverages the Whole Brain Catalog as a first example of how distributed and federated brain data can be collected into views using the standards it has created (Hawrylycz et al., 2011).

Where the data to import into the Whole Brain Catalog are large, special care has been taken to only bring in those parts of a data set that are needed for display. This has been done with methods such as the tile-rendering approach to large images drawn from image servers and the streaming of time-series data sets for simulation rendering. Because of the growing file sizes coming from a big data approach to data collection across the sciences, pulling an entire data set into memory is rapidly becoming one of the major limitations to data analysis. We believe that the approach taken here, partially-loading subsets of data sets, is a more scalable one.

Additionally we have shown the value of integrating with the semantic framework of the Neuroscience Information Framework through the usage of ontological identifiers. Users of the Whole Brain Catalog are able to benefit from definitions, synonyms, and other properties that are being aggregated at NeuroLex.org while they are using the Whole Brain Catalog to investigate the data sets there. They are able to follow a link from within the Whole Brain Catalog to the NIF Navigator, floating on the side of each NeuroLex.org page, which rapidly brings them to the results of a focused, ontologically indexed query of data sets relevant to the object they are looking at. This simultaneously provides a simplified way to access resources that are otherwise hard to find individually, and by providing a useful service to an end user, encourages others to do the same.

We have also shown that the integration of different modalities of data into the same space can provide unique advantages to research in neuroscience. As shown in Figs 5.13 and 5.14, the ability to easily transform the morphology of those neurons into computable graphs to be able to perform the analysis, reusing open source libraries for the mathematics, led to insight in the structure of neuronal

morphologies in the olfactory bulb (Ghosh et al., 2011). We believe the ability to bring to bear a diverse set of computational tools in this example demonstrates the flexibility and genericity of the architecture. Moreover the application of the CAD-like controls and the use of 2D features in the data to align 3D features shows a novel means of using relationships in one modality of data to inform otherwise implicit relationships in anothera validation of what is possible within multi-scale multi-modal virtual environments.

In summary, we have designed, implemented, and validated a computational system for multi-scale, multi-modal neuroanatomy that is built to be generic and reused. We have offered examples of the value of this framework for visualizing data sets in the neuroanatomical context from which they were taken, comparing the system to the intention of virtual globes such as Google Earth. We have incorporated computational simulations into the environment to demonstrate how this approach to neuroanatomy could be useful to computational neuroscientists. We have shown examples of how data integration with other resources in neuroinformatics can be helpful to an end user. Finally, we have provided a case study of a usage for the Whole Brain Catalog in improving scientific understanding.

5.6 Future directions

The ideal outcome of digital neuroanatomy would be a what we would consider the ultimate microscope. We would want to be able to digitially reconstruct, visualize, and analyze the entire multi-scale organization of the nervous system. Every cell morphology would be visible, and every spine on that morphology, if it had them, would also be visible. The molecular constituents of all cells contained would be viewable, and comparable against the patterns of gene expressions. While this vision may seem like a fantastic one because of the modern limitations of imaging technology, from the perspective of what is achievable by information systems, it may not be so far off. As with our comparison to Google Earth, the limits of the scales that must be spanned are comparable megameters (10^6) to centimeters (10^{-2}) for the earth (8 orders of magnitude) and centimeters to nanometers (10^{-9})

for the nervous system (7 orders of magnitude). We believe that with the proper investment of effort, the equivalent of Google Earth for the brain is possible.

In addition, as we have pointed out, the object oriented system that we have designed here is intended for extension, which means that as new imaging methodologies produce novel spatial patterns that should be embedded into the Whole Brain Catalog, we would like that future software developers can write the appropriate code needed to bring that data in. We ourselves have made progress on the display of data such as DTI fibers and volumes, but have not yet finalized this. Ideally, we would like to see a model building environment like NeuroConstruct (Gleeson et al., 2007) be incorporated into the Whole Brain Catalog. This is much harder to do if the source code is kept closed, and also disincentivizes would-be contributors who prefer to have the ability to investigate how a method is performing its task. Therefore, in addition to being important for the full value of the work to be recognized, we also believe that a strategy of open source development in neuroinformatics is essential for allowing for effective aggregation of software development effort in the future.

Finally, we would like to see the Whole Brain Catalog become the standard 3D repository of neuroanatomical structure for neuroscience. In order to facilitate this, the environment must be extended to become a service that provides neuroscientists end-to-end management of their datasets and continually simplifies and reduces the barrier of entry to their usage. We believe this is possible and are working towards this goal.

Chapter 6

Contributions

On Feynman's last day, his chalkboard had a box around the words: "What I cannot create, I cannot understand." In this dissertation, I have demonstrated computational systems that enable the formalization of knowledge in the neurosciences into forms that can be used to ask and answer novel questions. The purpose of these systems has been to create frameworks that allow the building of representations of the nervous system that augment the ability of the neuroscience researcher to ask and answer novel questions. Through this constructive effort, of assembling what we know about the nervous system into computational forms, neuroscience researchers can ask questions and get back answers about the nervous system that have not been possible in the past.

In chapters 2 and 3, I demonstrated the utility of formal structures for capturing knowledge in the form of OWL ontologies. Here the formation of an ontology for subcellular anatomy, and related reasoning methodologies, enabled the inference of knowledge that was not explictly called out in an electron micrograph, but could be derived from the combination of those data with additional knowledge about the structure of the nervous system. In chapter 4, I introduced an online system for the collaborative construction of a parts list for the nervous system that was graph-based and which lowers barriers of entry to discovery and editing of the concepts contained within it. Using the knowledge structured within it, I was able to pose a novel question about cerebellar neuroanatomy that allowed the investigation of existing facts from a novel perspective. Through semantic

reasoning, I was able to answer this question programmatically. In chapter 5, I introduced an online system for the assembly of multi-scale image data of the nervous system in a multi-scale 3D coordinate system. By assembling knowledge into this form, I assisted an investigation to both ask and answer a question about the regularity of the structure of neurons in the olfactory bulb.

Bibliography

Putting the brain back together. *Nature neuroscience*, 9(4):457, April 2006. ISSN 1097-6256. doi: 10.1038/nn0406-457. URL `http://www.ncbi.nlm.nih.gov/pubmed/16568097`.

James B Aimone, Janet Wiles, and Fred H Gage. Computational influence of adult neurogenesis on memory encoding. *Neuron*, 61(2): 187–202, January 2009. ISSN 1097-4199. doi: 10.1016/j.neuron.2008.11.026. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2670434\&tool=pmcentrez\&rendertype=abstract`.

H. Akil, M. E. Martone, and D. C. Van Essen. Challenges and Opportunities in Mining Neuroscience Data. *Science*, 331(6018):708–712, February 2011. ISSN 0036-8075. doi: 10.1126/science.1199305. URL `http://www.sciencemag.org/cgi/doi/10.1126/science.1199305`.

Laurent Alquier, T. Schultz, and Susie Stephens. Exploration of a Data Landscape using a Collaborative Linked Data Framework. In *Proceedings of the Workshop on the Future of the Web for Collaborative Science (WWW2010)*, 2010. URL `http://imageweb.zoo.ox.ac.uk/pub/2010/Proceedings/FWCS2010/07/Paper7.pdf`.

Joseph Altman and Shirley A. Bayer. *Development of the Cerebellar System: In relation to its evolution, structure, and functions*. CRC Press, 1997. ISBN 0-8493-9490-2.

Grigoris Antoniou and Frank Van Harmelen. Web Ontology Language: OWL. In Steffen Staab and Rudi Studer, editors, *Handbook on Ontologies*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009. ISBN 978-3-540-70999-2. doi: 10.1007/978-3-540-92673-3. URL `http://www.springerlink.com/index/10.1007/978-3-540-92673-3`.

Giorgio a Ascoli. Computing the Brain and the Computing Brain. In Giorgio a Ascoli, editor, *Computational Neuroanatomy: Principles and Methods*, pages 3–23. Humana Press, 2002. ISBN 1-58829-000-X.

Giorgio a Ascoli. The coming of age of the hippocampome. *Neuroinformatics*, 8 (1):1–3, March 2010. ISSN 1559-0089. doi: 10.1007/s12021-010-9063-0. URL `http://www.ncbi.nlm.nih.gov/pubmed/20127205`.

Giorgio a Ascoli, Lidia Alonso-Nanclares, Stewart a Anderson, German Barrionuevo, Ruth Benavides-Piccione, Andreas Burkhalter, György Buzsáki,

Bruno Cauli, Javier Defelipe, Alfonso Fairén, Dirk Feldmeyer, Gord Fishell, Yves Fregnac, Tamas F Freund, Daniel Gardner, Esther P Gardner, Jesse H Goldberg, Moritz Helmstaedter, Shaul Hestrin, Fuyuki Karube, Zoltán F Kisvárday, Bertrand Lambolez, David a Lewis, Oscar Marin, Henry Markram, Alberto Muñoz, Adam Packer, Carl C H Petersen, Kathleen S Rockland, Jean Rossier, Bernardo Rudy, Peter Somogyi, Jochen F Staiger, Gabor Tamas, Alex M Thomson, Maria Toledo-Rodriguez, Yun Wang, David C West, and Rafael Yuste. Petilla terminology: nomenclature of features of GABAergic interneurons of the cerebral cortex. *Nature reviews. Neuroscience*, 9(7):557–68, July 2008. ISSN 1471-0048. doi: 10.1038/nrn2402. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2868386\&tool=pmcentrez\&rendertype=abstract`.

M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, and Others. Gene ontology: tool for the unification of biology. *Nature genetics*, 25(1):25–29, 2000. URL `http://www.nature.com/ng/journal/v25/n1/abs/ng0500\_25.html`.

Rembrandt Bakker, Stephen D Larson, Strobelt Sandra, Andreas Hess, Daniel Wojcik, Piotr Majka, and Rolf Kotter. Scalable Brain Atlas: From Stereotaxic Coordinate to Delinated Brain Region. In *INCF 3rd Annual Meeting*. Frontiers In Neuroscience, 2010. doi: 10.3389/conf.fnins.2010.13.00028. URL `http://www.frontiersin.org/Journal/Abstract.aspx?f=55\&name=neuroscience\&ART\_DOI=10.3389/conf.fnins.2010.13.00028`.

Jonathan Bard, Seung Y Rhee, and Michael Ashburner. An ontology for cell types. *Genome biology*, 6(2):R21, January 2005. ISSN 1465-6914. doi: 10.1186/gb-2005-6-2-r21. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=551541\&tool=pmcentrez\&rendertype=abstract`.

Susan M Baxter, Steven W Day, Jacquelyn S Fetrow, and Stephanie J Reisinger. Scientific Software Development Is Not an Oxymoron. *PLoS Computational Biology*, 2(9):4, 2006. URL `http://www.ncbi.nlm.nih.gov/pubmed/16965174`.

Sabine Begall, Jaroslav Cerveny, Julia Neef, Oldrich Vojtech, and Hynek Burda. Magnetic alignment in grazing and resting cattle and deer. *Proceedings of the National Academy of Sciences of the United States of America*, 105(36):13451–5, September 2008. ISSN 1091-6490. doi: 10.1073/pnas.0803650105. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2533210\&tool=pmcentrez\&rendertype=abstract`.

Louise Bertrand and Jonathan Nissanov. The Neuroterrain 3D Mouse Brain Atlas. *Frontiers in neuroinformatics*, 2(July):3, January 2008. ISSN 1662-5196. doi: 10.3389/neuro.11.003.2008. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2525976\&tool=pmcentrez\&rendertype=abstract`.

F E Bloom. The multidimensional database and neuroinformatics requirements for molecular and cellular neuroscience. *NeuroImage*, 4(3 Pt 2):S12–3, December 1996. ISSN 1053-8119. doi: 10.1006/nimg.1996.0042. URL `http://www.ncbi.nlm.nih.gov/pubmed/9345517`.

F E Bloom and W G Young. Database Needs of Neuroscience: Schema and Design.

In S H Koslow and S Subramaniam, editors, *Databasing the Brain: From Data to Knowledge*, pages 3–25. Wiley & Sons, 2005. ISBN 0-471-30921-4.

Jyl Boline, Allan MacKenzie-Graham, David Shattuck, H Yuan, S Anderson, DM Sforza, J Wang, Robert W Williams, Willy Waiho Wong, Maryann E Martone, Ilya Zaslavsky, and Arthur W Toga. A Digital Atlas and Neuroinformatics Framework for Query and Display of Disparate Data. In *Society for Neuroscience*, page 100.12, 2006.

Mihail Bota and Larry W Swanson. BAMS Neuroanatomical Ontology: Design and Implementation. *Frontiers in neuroinformatics*, 2 (May):2, January 2008. ISSN 1662-5196. doi: 10.3389/neuro.11.002. 2008. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2525975\&tool=pmcentrez\&rendertype=abstract`.

Mihail Bota, Hong-Wei Dong, and Larry W Swanson. Brain architecture management system. *Neuroinformatics*, 3(1):15–48, January 2005. ISSN 1539-2791. doi: 10.1385/NI:3:1:015. URL `http://www.ncbi.nlm.nih.gov/pubmed/15897615`.

D.M. Bowden and R.F. Martin. NeuroNames brain hierarchy. *Neuroimage*, 2(1):63–83, 1995. URL `http://linkinghub.elsevier.com/retrieve/pii/S1053811985710099`.

Douglas M Bowden and Mark F Dubach. NeuroNames 2002. *Neuroinformatics*, 1(1):43–59, January 2003. ISSN 1539-2791. doi: 10.1385/NI:1:1:043. URL `http://www.ncbi.nlm.nih.gov/pubmed/15055392`.

Douglas M Bowden, Mark Dubach, and Jack Park. Creating neuroscience ontologies. *Methods in molecular biology (Clifton, N.J.)*, 401(1):67–87, January 2007. ISSN 1064-3745. doi: 10.1007/978-1-59745-520-6_5. URL `http://www.ncbi.nlm.nih.gov/pubmed/18368361`.

a Brevik, T B Leergaard, M Svanevik, and J G Bjaalie. Three-dimensional computerised atlas of the rat brain stem precerebellar system: approaches for mapping, visualization, and comparison of spatial distribution data. *Anatomy and embryology*, 204(4):319–32, October 2001. ISSN 0340-2061. URL `http://www.ncbi.nlm.nih.gov/pubmed/11720236`.

James F Brinkley, Jeffrey S Prothero, John W Prothero, and Cornelius Rosse. A Framework for the Design of Knowledge-Based Systems in Structural Biology. *Current*, pages 61–65, 1989.

James F Brinkley, Dan Suciu, Landon T Detwiler, H John, and Biomedical Informatics. A framework for using reference ontologies as a foundation for the semantic web Departments of Biological Structure , 2 Computer Science and Engineering and 3 Medical AMIA 2006 Symposium AMIA 2006 Symposium. *Medical Education*, pages 96–100, 2006.

William J Bug, Giorgio a Ascoli, Jeffrey S Grethe, Amarnath Gupta, Christine Fennema-Notestine, Angela R Laird, Stephen D Larson, Daniel Rubin, Gordon M Shepherd, Jessica a Turner, and Maryann E Martone. The NIFSTD and BIRNLex vocabularies: building comprehensive ontologies for neuroscience. *Neuroinformatics*, 6(3):175–94,

September 2008. ISSN 1559-0089. doi: 10.1007/s12021-008-9032-z. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2743139\&tool=pmcentrez\&rendertype=abstract.

Hynek Burda, Sabine Begall, Jaroslav Cerveny, Julia Neef, and Pavel Nemec. Extremely low-frequency electromagnetic fields disrupt magnetic alignment of ruminants. *Proceedings of the National Academy of Sciences*, June 2009. ISSN 1536-8386.

Declan Butler. The web-wide world. *Nature*, 439(7078):776–778, 2006. doi: 10.1038/439776a. URL http://cat.inist.fr/?aModele=afficheN\&cpsidt=17483706.

Chao Cai, Mingyue Ding, Hao Lei, Jie Cao, Ailing Liu, Intelligent Control, and M Physics. A Novel 3D Correspondence-Less Method for MRI and Paxinos-Watson Atlas of Rat Brain Registration. *Image Rochester NY*, pages 269 – 276, 2006.

Stuart G Campbell and Andrew D McCulloch. Multi-scale computational models of familial hypertrophic cardiomyopathy: genotype to phenotype. *Journal of the Royal Society, Interface / the Royal Society*, (August), August 2011. ISSN 1742-5662. doi: 10.1098/rsif.2011.0184. URL http://www.ncbi.nlm.nih.gov/pubmed/21831889.

Jie Cao. Localization and labeling of rat brain in MR image based on Paxinos-Watson atlas. *Proceedings of SPIE*, 6141:614129–614129–10, 2006. ISSN 0277786X. doi: 10.1117/12.652658. URL http://link.aip.org/link/PSISDG/v6141/i1/p614129/s1\&Agg=doi.

W Ceusters, B Smith, and L Goldberg. A terminological and ontological analysis of the NCI Thesaurus. *Methods of information in medicine*, 44(4):498–507, January 2005. ISSN 0026-1270. URL http://www.ncbi.nlm.nih.gov/pubmed/16342916.

E Chan, N Kovacevíc, S K Y Ho, R M Henkelman, and J T Henderson. Development of a high resolution three-dimensional surgical atlas of the murine head for strains 129S1/SvImJ and C57Bl/6J using magnetic resonance imaging and micro-computed tomography. *Neuroscience*, 144(2):604–15, January 2007. ISSN 0306-4522. doi: 10.1016/j.neuroscience.2006.08.080. URL http://www.ncbi.nlm.nih.gov/pubmed/17101233.

Li Chen, Maryann Martone, Lisa Fong, and Mona Wong-barnum. OntoQuest : Exploring Ontological Data Made Easy. *Context*, pages 1183–1186, 2006.

Kei-Hoi Cheung, Ernest Lim, Matthias Samwald, Huajun Chen, Luis Marenco, Matthew E Holford, Thomas M Morse, Pradeep Mutalik, Gordon M Shepherd, and Perry L Miller. Approaches to neuroscience data integration. *Briefings in bioinformatics*, 10(4):345–53, July 2009. ISSN 1477-4054. doi: 10.1093/bib/bbp029. URL http://www.ncbi.nlm.nih.gov/pubmed/19505888.

Tim Clark and June Kinoshita. Alzforum and SWAN: the present and future of scientific web communities. *Briefings in bioinformatics*, 8(3):163–71, May 2007. ISSN 1467-5463. doi: 10.1093/bib/bbm012. URL http://www.ncbi.nlm.nih.

gov/pubmed/17510163.

Jay S Coggan, Thomas M Bartol, Eduardo Esquenazi, Joel R Stiles, Stephan Lamont, Maryann E Martone, Darwin K Berg, Mark H Ellisman, and Terrence J Sejnowski. Evidence for ectopic neurotransmission at a neuronal synapse. *Science (New York, N.Y.)*, 309(5733):446–51, July 2005. ISSN 1095-9203. doi: 10.1126/science.1108239. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2915764\&tool=pmcentrez\&rendertype=abstract.

Neil G Connelly and Ture Damhus. Nomenclature of inorganic chemistry: IUPAC recommendations 2005. In *Nomenclature Of Inorganic Chemistry IUPAC Recommendations 2005 Red Book*, chapter 5.1, page 98. Royal Society of Chemistry, 2005. ISBN 0854044388. URL http://books.google.com/books?hl=en\&lr=\&id=w1Kf1CakyZIC\&oi=fnd\&pg=PR2\&dq=Nomenclature+of+inorganic+chemistry:+IUPAC+recommendations+2005\&ots=rMGMXAYY3l\&sig=XtV9DKTt5x35HgnfqI8L5JiHF6s.

J D Cooke, B Larson, O Oscarsson, and B Sjölund. Origin and termination of cuneocerebellar tract. *Experimental brain research. Experimentelle Hirnforschung. Expérimentation cérébrale*, 13(4):339–58, October 1971. ISSN 0014-4819. URL http://www.ncbi.nlm.nih.gov/pubmed/5123642.

Ronald Cornet and Nicolette de Keizer. Forty years of SNOMED: a literature review. *BMC medical informatics and decision making*, 8 Suppl 1:S2, January 2008. ISSN 1472-6947. doi: 10.1186/1472-6947-8-S1-S2. URL http://www.ncbi.nlm.nih.gov/pubmed/19007439.

Chiquito J Crasto, Luis N Marenco, Nian Liu, Thomas M Morse, Kei-Hoi Cheung, Peter C Lai, Gautam Bahl, Peter Masiar, Hugo Y K Lam, Ernest Lim, Huajin Chen, Prakash Nadkarni, Michele Migliore, Perry L Miller, and Gordon M Shepherd. SenseLab: new developments in disseminating neuroscience information. *Briefings in bioinformatics*, 8(3):150–62, May 2007. ISSN 1467-5463. doi: 10.1093/bib/bbm018. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2756159\&tool=pmcentrez\&rendertype=abstract.

Mike D R Croning, Michael C Marshall, Peter McLaren, J Douglas Armstrong, and Seth G N Grant. G2Cdb: the Genes to Cognition database. *Nucleic acids research*, 37(Database issue):D846–51, January 2009. ISSN 1362-4962. doi: 10.1093/nar/gkn700. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2686544\&tool=pmcentrez\&rendertype=abstract.

Jacob Czech, Markus Dittrich, and Joel R Stiles. Systems Biology. *Methods*, 500, 2009. doi: 10.1007/978-1-59745-525-1. URL http://www.springerlink.com/index/10.1007/978-1-59745-525-1.

Randall Davis, Howard Shrobe, and Peter Szolovits. What is a knowledge representation? *AI magazine*, 14(1):17, 1993. URL http://www.aaai.org/ojs/index.php/aimagazine/article/viewArticle/1029.

John Day-Richter, Midori a Harris, Melissa Haendel, and Suzanna Lewis. OBO-Edit–an ontology editor for biologists. *Bioinformatics (Oxford, England)*, 23(16):

2198–200, August 2007. ISSN 1367-4811. doi: 10.1093/bioinformatics/btm112. URL http://www.ncbi.nlm.nih.gov/pubmed/17545183.

Juan J de Pablo. Coarse-grained simulations of macromolecules: from DNA to nanocomposites. *Annual review of physical chemistry*, 62:555–74, May 2011. ISSN 0066-426X. doi: 10.1146/annurev-physchem-032210-103458. URL http://www.ncbi.nlm.nih.gov/pubmed/21219152.

Dirk Derom, R.A. Schmidt, I. McLeod, and B.A. Hewitt. Using Metaneva for Structuring, Managing and Retrieving Animal Data in the Cognitive Neurosciences. *Victoria*, 2010. URL http://precedings.nature.com/documents/4146/version/1.

Ivo D Dinov, Daniel Valentino, Bae Cheol Shin, Fotios Konstantinidis, Guogang Hu, Allan MacKenzie-Graham, Erh-Fang Lee, David Shattuck, Jeff Ma, Craig Schwartz, and Arthur W Toga. LONI visualization environment. *Journal of digital imaging : the official journal of the Society for Computer Applications in Radiology*, 19(2):148–58, June 2006. ISSN 0897-1889. doi: 10.1007/s10278-006-0266-8. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3045182\&tool=pmcentrez\&rendertype=abstract.

H.U. Dodt, Ulrich Leischner, Anja Schierloh, N. J\\"ahrling, C.P. Mauch, Katrin Deininger, J.M. Deussing, Matthias Eder, W. Zieglg\\"ansberger, and Klaus Becker. Ultramicroscopy: three-dimensional visualization of neuronal networks in the whole mouse brain. *Nature methods*, 4(4):331–336, 2007. doi: 10.1038/NMETH1036. URL http://www.nature.com/nmeth/journal/vaop/ncurrent/full/nmeth1036.html.

David Eberly. *3D Game Engine Design: A Practical Approach to Real-Time Computer Graphics*. Morgan Kaufmann, 2000. ISBN 1558605932.

Roy T. Fielding and Richard N. Taylor. Principled design of the modern Web architecture. *ACM Transactions on Internet Technology*, 2(2):115–150, May 2002. ISSN 15335399. doi: 10.1145/514183.514185. URL http://portal.acm.org/citation.cfm?doid=514183.514185.

Karl Fogel. *Producing Open Source Software How to Run a Successful Free Software Project*. O'Reilly Media, Inc., 2005. ISBN 0596007590. URL http://portal.acm.org/citation.cfm?id=1121560.

LL Fong, SD Larson, A. Gupta, C. Condit, W.J. Bug, L. Chen, R. West, S. Lamont, M. Terada, and ME Martone. An ontology-driven knowledge environment for subcellular neuroanatomy. In *CEUR Workshop Proceedings*, volume 258, pages 1613–0073, 2007. URL http://owl1-1.googlecode.com/svn-history/r546/trunk/www.webont.org/owled/2007/PapersPDF/submission_41.pdf.

Peter T Fox, Angela R Laird, Sarabeth P Fox, P Mickle Fox, Angela M Uecker, Michelle Crank, Sandra F Koenig, and Jack L Lancaster. BrainMap taxonomy of experimental design: description and evaluation. *Human brain mapping*, 25 (1):185–98, May 2005. ISSN 1065-9471. doi: 10.1002/hbm.20141. URL http://www.ncbi.nlm.nih.gov/pubmed/15846810.

Leon French, Suzanne Lane, Lydia Xu, and Paul Pavlidis. Automated recognition of brain region mentions in neuroscience literature. *Frontiers in neuroinformatics*, 3(September):29, January 2009. ISSN 1662-5196. doi: 10.3389/neuro.11.029.2009. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2741206\&tool=pmcentrez\&rendertype=abstract.

G.A. Frishkoff, P. LePendu, R.M. Frank, Haishan Liu, and Dejing Dou. Development of Neural Electromagnetic Ontologies (NEMO): Ontology-based Tools for Representation and Integration of Event-related Brain Potentials. *Development*, 2009. URL http://precedings.nature.com/documents/3458/version/1.

Daniel Gardner, H. Akil, G.A. Ascoli, D.M. Bowden, W. Bug, D.E. Donohue, D.H. Goldberg, B. Grafstein, J.S. Grethe, A. Gupta, and Others. The Neuroscience Information Framework: a data and knowledge environment for neuroscience. *Neuroinformatics*, 6(3):149–160, 2008. ISSN 1539-2791. doi: 10.1007/s12021-008-9024-z.The. URL http://www.springerlink.com/index/Q204000137166608.pdf.

M Garwicz, H Jorntell, and C F Ekerot. Cutaneous receptive fields and topography of mossy fibres and climbing fibres projecting to cat cerebellar C3 zone. *The Journal of physiology*, 512 (Pt 1):277–93, October 1998. ISSN 0022-3751. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2231180\&tool=pmcentrez\&rendertype=abstract.

The Gene and Ontology Consortium. Creating the gene ontology resource: design and implementation. *Genome research*, 11(8):1425–33, August 2001. ISSN 1088-9051. doi: 10.1101/gr.180801. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=311077\&tool=pmcentrez\&rendertype=abstract.

Stephan Gerhard, Alessandro Daducci, Alia Lemkaddem, Reto Meuli, Jean-Philippe Thiran, and Patric Hagmann. The connectome viewer toolkit: an open source framework to manage, analyze, and visualize connectomes. *Frontiers in neuroinformatics*, 5(June):3, January 2011. ISSN 1662-5196. doi: 10.3389/fninf.2011.00003. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3112315\&tool=pmcentrez\&rendertype=abstract.

Sulagna Ghosh, Stephen D Larson, Hooman Hefzi, Zachary Marnoy, Tyler Cutforth, Kartheek Dokka, and Kristin K Baldwin. Sensory maps in the olfactory cortex defined by long-range viral tracing of single neurons. *Nature*, 472 (7342):217–20, April 2011. ISSN 1476-4687. doi: 10.1038/nature09945. URL http://www.ncbi.nlm.nih.gov/pubmed/21451523.

Jim Giles. Internet encyclopaedias go head to head. *Nature*, 438(7070):900–1, December 2005. ISSN 1476-4687. doi: 10.1038/438900a. URL http://www.ncbi.nlm.nih.gov/pubmed/16355180.

Georgios V Gkoutos, Eain C J Green, Ann-Marie Mallon, John M Hancock, and Duncan Davidson. Using ontologies to describe mouse phenotypes. *Genome biology*, 6(1):R8, January 2005. ISSN 1465-6914. doi: 10.1186/gb-2004-6-1-r8. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2447424\&tool=pmcentrez\&rendertype=abstract.

Padraig Gleeson, Volker Steuber, and R Angus Silver. neuroConstruct: a tool for modeling networks of neurons in 3D space. *Neuron*, 54(2):219–35, April 2007. ISSN 0896-6273. doi: 10.1016/j.neuron.2007.03.025. URL `http://www.ncbi.nlm.nih.gov/pubmed/17442244`.

Padraig Gleeson, Sharon Crook, Robert C. Cannon, Michael L. Hines, Guy O. Billings, Matteo Farinella, Thomas M. Morse, Andrew P. Davison, Subhasis Ray, Upinder S. Bhalla, Simon R. Barnes, Yoana D. Dimitrova, and R. Angus Silver. NeuroML: A Language for Describing Data Driven Models of Neurons and Networks with a High Degree of Biological Detail. *PLoS Computational Biology*, 6(6):e1000815, June 2010. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1000815. URL `http://dx.plos.org/10.1371/journal.pcbi.1000815`.

Christine Golbreich, Olivier Dameron, Olivier Bierlaire, and Bernard Gibaud. What reasoning support for Ontology and Rules? the brain anatomy case study. In *Proceedings of the workshop" OWL Experiences and Directions", Nov*, pages 11–12. Citeseer, 2005. URL `http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.96.5802\&rep=rep1\&type=pdf`.

P Grenon. BFO in a Nutshell: A bi-categorical axiomization of BFO and Comparison with DOLCE. Technical Report December, IFOMIS Report, 2003.

Pierre Grenon, Barry Smith, and Louis Goldberg. Biodynamic ontology: applying BFO in the biomedical domain. *Studies in health technology and informatics*, 102(ii):20–38, January 2004. ISSN 0926-9630. URL `http://www.ncbi.nlm.nih.gov/pubmed/15853262`.

Jeffrey S Grethe. NeuroLex and the Neuroscience Information Framework: Building comprehensive neuroscience ontologies with and for the community. In *Frontiers in neuroscience conference abstract: Neuroinformatics 2009*, 2009. doi: 10.3389/conf.neuro.11.2009.08.140.

T Gruber. Toward principles for the design of ontologies used for knowledge sharing? *International Journal of Human-Computer Studies*, 43(5-6):907–928, November 1993. ISSN 10715819. doi: 10.1006/ijhc.1995.1081. URL `http://linkinghub.elsevier.com/retrieve/doi/10.1006/ijhc.1995.1081`.

A. Gupta, S. Larson, Christopher Condit, Sandeep Gupta, Lisa Fong, Li Chen, and M. Martone. Toward an ontological database for subcellular neuroanatomy. *Advances in Conceptual ModelingFoundations and Applications*, (8):64–73, 2007. URL `http://www.springerlink.com/index/w7p010307q411251.pdf`.

Carl Gustafson, William J Bug, and Jonathan Nissanov. NeuroTerrain–a client-server system for browsing 3D biomedical image data sets. *BMC bioinformatics*, 8:40, January 2007. ISSN 1471-2105. doi: 10.1186/1471-2105-8-40. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1802997\&tool=pmcentrez\&rendertype=abstract`.

David J Hamilton and Giorgio A Ascoli. The Neuron Registry Curator Interface: an Infrastructure for the Collaborative Definition of Neuronal Properties. In *3rd INCF Congress of Neuroinformatics*, page 59, 2010.

Stephan Hartmann. The World as a Process : Simulations in the Natural and

Social Sciences. In R Hegselmann, editor, *Simulation and Modelling in the Social Sciences from the Philosophy of Science Point of View*, pages 77–100. Theory and Decision Library. Kluwer, Dordrect, 1996.

Michael Hawrylycz, Richard a. Baldock, Albert Burger, Tsutomu Hashikawa, G. Allan Johnson, Maryann Martone, Lydia Ng, Chris Lau, Stephen D. Larsen, Jonathan Nissanov, Luis Puelles, Seth Ruffins, Fons Verbeek, Ilya Zaslavsky, and Jyl Boline. Digital Atlasing and Standardization in the Mouse Brain. *PLoS Computational Biology*, 7(2):e1001065, February 2011. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1001065. URL `http://dx.plos.org/10.1371/journal.pcbi.1001065`.

Shan He, Senthil K Nachimuthu, Shaun C Shakib, and Lee Min Lau. Collaborative authoring of biomedical terminologies using a semantic Wiki. *AMIA ... Annual Symposium proceedings / AMIA Symposium. AMIA Symposium*, 2009:234–8, January 2009. ISSN 1942-597X. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2815405\&tool=pmcentrez\&rendertype=abstract`.

Michael Hewett. Algernon - Rule-based Programming. URL `http://algernon-j.sourceforge.net`.

M L Hines and N T Carnevale. The NEURON simulation environment. *Neural computation*, 9(6):1179–209, August 1997. ISSN 0899-7667. URL `http://www.ncbi.nlm.nih.gov/pubmed/9248061`.

Trine Hjornevik, Trygve B Leergaard, Dmitri Darine, Olve Moldestad, Anders M Dale, Frode Willoch, and Jan G Bjaalie. Three-dimensional atlas system for mouse and rat brain imaging data. *Frontiers in neuroinformatics*, 1(November):4, January 2007. ISSN 1662-5196. doi: 10.3389/neuro.11.004.2007. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2525992\&tool=pmcentrez\&rendertype=abstract`.

M F Horstemeyer. Multiscale Modeling: A Review. In Jerzy Leszczynski and Manoj K Shukla, editors, *Practical Aspects of Computational Chemistry: Methods, Concepts and Applications*, pages 87–136. Springer, Heidelberg, 2009. ISBN 978-90-481-2686-6. URL `http://j.mp/o5RQ2n`.

Jeff Howe. The Rise of Crowdsourcing. *Wired Magazine*, 14(14):1–5, 2006. URL `http://www.clickadvisor.com/downloads/Howe\_The\_Rise\_of\_Crowdsourcing.pdf`.

M Hucka, K Shankar, D Beeman, and James M Bower. The Modeler's Workspace. In Giorgio a Ascoli, editor, *Computational Neuroanatomy: Principles and Methods*, pages 83–103. Humana Press, 2002. ISBN 1-58829-000-X.

M F Huerta, S H Koslow, and A I Leshner. The Human Brain Project: an international resource. *Trends in Neurosciences*, 16(11):436–438, 1993.

Jon W Huss, Camilo Orozco, James Goodale, Chunlei Wu, Serge Batalov, Tim J Vickers, Faramarz Valafar, and Andrew I Su. A gene wiki for community annotation of gene function. *PLoS biology*, 6 (7):e175, July 2008. ISSN 1545-7885. doi: 10.1371/journal.pbio.

0060175. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2443188\&tool=pmcentrez\&rendertype=abstract.

Shantanu H Joshi, John Darrell Van Horn, and Arthur W Toga. Interactive exploration of neuroanatomical meta-spaces. *Frontiers in neuroinformatics*, 3(November):38, January 2009. ISSN 1662-5196. doi: 10.3389/neuro.11.038.2009. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2776489\&tool=pmcentrez\&rendertype=abstract.

a Kalyanpur, B Parsia, E Sirin, B Grau, and J Hendler. Swoop: A Web Ontology Editing Browser. *Web Semantics: Science, Services and Agents on the World Wide Web*, 4(2):144–153, June 2006. ISSN 15708268. doi: 10.1016/j.websem.2005.10.001. URL http://linkinghub.elsevier.com/retrieve/pii/S1570826805000326.

Paul S Katz, Robert Calin-Jageman, Akshaye Dhawan, Chad Frederick, Shuman Guo, Rasanjalee Dissanayaka, Naveen Hiremath, Wenjun Ma, Xiuyn Shen, Hsui C Wang, Hong Yang, Sushil Prasad, Rajshekhar Sunderraman, and Ying Zhu. NeuronBank: A Tool for Cataloging Neuronal Circuitry. *Frontiers in systems neuroscience*, 4 (April):9, January 2010. ISSN 1662-5137. doi: 10.3389/fnsys.2010.00009. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2859812\&tool=pmcentrez\&rendertype=abstract.

S.H Koslow and M F Huerta. *Neuroinformatics: an overview of the human brain project*. Lawrence Erlbaum Associates, Mahwah, New Jersey, 1997.

Rolf Kötter. Online retrieval, processing, and visualization of primate connectivity data from the CoCoMac database. *Neuroinformatics*, 2(2):127–44, January 2004. ISSN 1539-2791. doi: 10.1385/NI:2:2:127. URL http://www.springerlink.com/content/c6503007275881n7/.

Markus Krotzsch, D. Vrandecic, and Max Volkel. *Semantic MediaWiki*, volume 4273 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006. ISBN 978-3-540-49029-6. doi: 10.1007/11926078. URL http://www.springerlink.com/content/j671440610746246/.

Tobias Kuhn. Acewiki: Collaborative ontology management in controlled natural language. *Arxiv preprint arXiv:0807.4623*, 2008. URL http://arxiv.org/abs/0807.4623.

A Kuß, S Prohaska, B Meyer, J Rybak, and HC Hege. Ontology-Based Visualization of Hierarchical Neuroanatomical Structures. *Proc Visual Computing for Biomedicine*, pages 177–184, 2008. URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.150.766\&rep=rep1\&type=pdf.

S. Larson and M. Martone. Rule-based reasoning with a multi-scale neuroanatomical ontology. In *CEUR Workshop Proceedings*, volume 258, pages 1613–0073, 2007. URL http://scholar.google.com/scholar?hl=en\&btnG=Search\&q=intitle:Rule-based+reasoning+with+a+multi-scale+neuroanatomical+ontology\#0.

Stephen D Larson and Maryann E Martone. Ontologies for Neuroscience:

What are they and What are they Good for? *Frontiers in neuroscience*, 3(1):60–7, May 2009. ISSN 1662-453X. doi: 10.3389/neuro.01.007.2009. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2695392\&tool=pmcentrez\&rendertype=abstract.

Christopher Lau, Lydia Ng, Carol Thompson, Sayan Pathak, Leonard Kuan, Allan Jones, and Mike Hawrylycz. Exploration and visualization of gene expression with neuroanatomy in the adult mouse brain. *BMC bioinformatics*, 9:153, January 2008. ISSN 1471-2105. doi: 10.1186/1471-2105-9-153. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2375125\&tool=pmcentrez\&rendertype=abstract.

Ed S Lein, Michael J Hawrylycz, Nancy Ao, Mikael Ayres, Amy Bensinger, Amy Bernard, Andrew F Boe, Mark S Boguski, Kevin S Brockway, Emi J Byrnes, Lin Chen, Li Chen, Tsuey-Ming Chen, Mei Chi Chin, Jimmy Chong, Brian E Crook, Aneta Czaplinska, Chinh N Dang, Suvro Datta, Nick R Dee, Aimee L Desaki, Tsega Desta, Ellen Diep, Tim a Dolbeare, Matthew J Donelan, Hong-Wei Dong, Jennifer G Dougherty, Ben J Duncan, Amanda J Ebbert, Gregor Eichele, Lili K Estin, Casey Faber, Benjamin a Facer, Rick Fields, Shanna R Fischer, Tim P Fliss, Cliff Frensley, Sabrina N Gates, Katie J Glattfelder, Kevin R Halverson, Matthew R Hart, John G Hohmann, Maureen P Howell, Darren P Jeung, Rebecca a Johnson, Patrick T Karr, Reena Kawal, Jolene M Kidney, Rachel H Knapik, Chihchau L Kuan, James H Lake, Annabel R Laramee, Kirk D Larsen, Christopher Lau, Tracy a Lemon, Agnes J Liang, Ying Liu, Lon T Luong, Jesse Michaels, Judith J Morgan, Rebecca J Morgan, Marty T Mortrud, Nerick F Mosqueda, Lydia L Ng, Randy Ng, Geralyn J Orta, Caroline C Overly, Tu H Pak, Sheana E Parry, Sayan D Pathak, Owen C Pearson, Ralph B Puchalski, Zackery L Riley, Hannah R Rockett, Stephen a Rowland, Joshua J Royall, Marcos J Ruiz, Nadia R Sarno, Katherine Schaffnit, Nadiya V Shapovalova, Taz Sivisay, Clifford R Slaughterbeck, Simon C Smith, Kimberly a Smith, Bryan I Smith, Andy J Sodt, Nick N Stewart, Kenda-Ruth Stumpf, Susan M Sunkin, Madhavi Sutram, Angelene Tam, Carey D Teemer, Christina Thaller, Carol L Thompson, Lee R Varnam, Axel Visel, Ray M Whitlock, Paul E Wohnoutka, Crissa K Wolkey, Victoria Y Wong, Matthew Wood, Murat B Yaylaoglu, Rob C Young, Brian L Youngstrom, Xu Feng Yuan, Bin Zhang, Theresa a Zwingman, and Allan R Jones. Genome-wide atlas of gene expression in the adult mouse brain. *Nature*, 445(7124): 168–76, January 2007. ISSN 1476-4687. doi: 10.1038/nature05453. URL http://www.ncbi.nlm.nih.gov/pubmed/17151600.

D a Lindberg, B L Humphreys, and a T McCray. The Unified Medical Language System. *Methods of information in medicine*, 32(4):281–91, August 1993. ISSN 0026-1270. URL http://www.ncbi.nlm.nih.gov/pubmed/8412823.

Jane Lomax and Alexa T McCray. Mapping the gene ontology into the unified medical language system. *Comparative and functional genomics*, 5(4):354–61, January 2004. ISSN 1531-6912. doi: 10.1002/cfg.407. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2447454\&tool=pmcentrez\&rendertype=abstract.

Courtney L Lopreore, Thomas M Bartol, Jay S Coggan, Daniel X Keller, Gina E Sosinsky, Mark H Ellisman, and Terrence J Sejnowski. Computational modeling of three-dimensional electrodiffusion in biological sys-

tems: application to the node of Ranvier. *Biophysical journal*, 95(6): 2624–35, September 2008. ISSN 1542-0086. doi: 10.1529/biophysj.108.132167. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2527256\&tool=pmcentrez\&rendertype=abstract.

Y Ma, P R Hof, S C Grant, S J Blackband, R Bennett, L Slatest, M D McGuigan, and H Benveniste. A three-dimensional digital atlas database of the adult C57BL/6J mouse brain by magnetic resonance microscopy. *Neuroscience*, 135 (4):1203–15, January 2005. ISSN 0306-4522. doi: 10.1016/j.neuroscience.2005.07.014. URL http://www.ncbi.nlm.nih.gov/pubmed/16165303.

Allan MacKenzie-Graham, Erh-Fang Lee, Ivo D Dinov, Mihail Bota, David W Shattuck, Seth Ruffins, Heng Yuan, Fotios Konstantinidis, Alain Pitiot, Yi Ding, Guogang Hu, Russell E Jacobs, and Arthur W Toga. A multimodal, multidimensional atlas of the C57BL/6J mouse brain. *Journal of anatomy*, 204(2): 93–102, February 2004. ISSN 0021-8782. doi: 10.1111/j.1469-7580.2004.00264.x. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1571243\&tool=pmcentrez\&rendertype=abstract.

R Maex and E De Schutter. Synchronization of golgi and granule cell firing in a detailed network model of the cerebellar granule cell layer. *Journal of Neurophysiology*, 80(5):2521–2537, 1998. URL http://www.ncbi.nlm.nih.gov/pubmed/9819260.

Z F Mainen, J Joerges, J R Huguenard, and T J Sejnowski. A model of spike initiation in neocortical pyramidal neurons. *Neuron*, 15(6):1427–39, December 1995. ISSN 0896-6273. URL http://www.ncbi.nlm.nih.gov/pubmed/8845165.

Henry Markram. The blue brain project. *Nature reviews. Neuroscience*, 7(2): 153–60, February 2006. ISSN 1471-003X. doi: 10.1038/nrn1848. URL http://www.ncbi.nlm.nih.gov/pubmed/16429124.

R F Martin, J L Mejino, D M Bowden, J F Brinkley, and C Rosse. Foundational model of neuroanatomy: implications for the Human Brain Project. *Proceedings / AMIA ... Annual Symposium. AMIA Symposium*, pages 438–42, January 2001. ISSN 1531-605X. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2243655\&tool=pmcentrez\&rendertype=abstract.

Richard F Martin, Joev Dubach, and Douglas M Bowden. NeuroNames: Human/Macaque Neuroanatomical Nomenclature. *Proceedings of the Annual Symposium on Computer Application in Medical Care*, pages 1018–1019, 1990.

Maryann E Martone, Amarnath Gupta, Mona Wong, Xufei Qian, Gina Sosinsky, Bertram Ludäscher, and Mark H Ellisman. A cell-centered database for electron tomographic data. *Journal of structural biology*, 138(1-2):145–55, 2002. ISSN 1047-8477. URL http://www.ncbi.nlm.nih.gov/pubmed/12160711.

Maryann E Martone, Shenglan Zhang, Amarnath Gupta, Xufei Qian, Haiyun He, Diana L Price, Mona Wong, Simone Santini, and Mark H Ellisman. The cell-centered database: a database for multiscale structural and protein localization data from light and electron microscopy. *Neuroinformatics*, 1(4):379–95, January 2003. ISSN 1539-2791. doi: 10.1385/NI:1:4:379. URL http://www.ncbi.nlm.

nih.gov/pubmed/15043222.

Maryann E Martone, Amarnath Gupta, and Mark H Ellisman. E-neuroscience: challenges and triumphs in integrating distributed data from molecules to brains. *Nature neuroscience*, 7(5):467–72, May 2004. ISSN 1097-6256. doi: 10.1038/nn1229. URL http://www.ncbi.nlm.nih.gov/pubmed/15114360.

Maryann E Martone, Joy Sargis, Joshua Tran, Willy Waiho Wong, Heather Jiles, and Cem Mangir. Database resources for cellular electron microscopy. *Methods in cell biology*, 79(06):799–822, January 2007. ISSN 0091-679X. doi: 10.1016/S0091-679X(06)79031-8. URL http://www.ncbi.nlm.nih.gov/pubmed/17327184.

Maryann E. Martone, Giorgio a Ascoli, Alan Ruttenberg, and David C. Van Essen. The INCF Program on Ontologies for Neural Structures. In *3rd INCF Congress of Neuroinformatics*, page 61, 2010.

M.E. Martone, Joshua Tran, W.W. Wong, Joy Sargis, Lisa Fong, Stephen Larson, S.P. Lamont, Amarnath Gupta, and M.H. Ellisman. The cell centered database project: an update on building community resources for managing and sharing 3D imaging data. *Journal of structural biology*, 161(3):220–231, 2008a. URL http://www.sciencedirect.com/science/article/pii/S1047847707002547.

M.E. Martone, Ilya Zaslavsky, Amarnath Gupta, Asif Memon, Joshua Tran, Willy Wong, Lisa Fong, S.D. Larson, and M.H. Ellisman. The smart atlas: spatial and semantic strategies for multiscale integration of brain data. *Anatomy Ontologies for Bioinformatics*, pages 267–286, 2008b. URL http://www.springerlink.com/index/V002843600Q13V0L.pdf.

a Maye, T H Wenckebach, and H-C Hege. Visualization, reconstruction, and integration of neuronal structures in digital brain atlases. *The International journal of neuroscience*, 116(4):431–59, April 2006. ISSN 0020-7454. doi: 10.1080/00207450500505860. URL http://www.ncbi.nlm.nih.gov/pubmed/16574581.

Ammar Mechouche, Christine Golbreich, and Bernard Gibaud. Towards an hybrid system for annotating brain MRI images. In *Workshop on OWL Experiences and Directions*, volume 216, pages 1–11. Citeseer, 2006. URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.142.9369\&rep=rep1\&type=pdf.

J.L.V. Mejino, L.T. Detwiler, J.A. Turner, M.E. Martone, D.L. Rubin, and J.F. Brinkley. Enabling RadLex with the Foundational Model of Anatomy Ontology to Organize and Integrate Neuro-imaging Data. *Symposium A Quarterly Journal In Modern Foreign Literatures*, pages 2010–2010, 2010. URL http://sigpubs.biostr.washington.edu/archive/00000246/01/AMIA-1310-A2009.PDF.

Michele Migliore and G.M. Shepherd. An integrated approach to classifying neuronal phenotypes. *Nature Reviews Neuroscience*, 6(10):810–818, 2005. URL http://www.nature.com/nrn/journal/v6/n10/abs/nrn1769.html.

Eric Miller, Christian Seppa, Aniket Kittur, F.W. Sabb, and R.A. Poldrack. The Cognitive Atlas: Employing Interaction Design Processes to Facilitate Collabo-

rative Ontology Creation. *Nature Precedings*, June 2010. ISSN 1756-0357. doi: 10.1038/npre.2010.4532. URL http://precedings.nature.com/documents/4532/version/1.

Perry L Miller, Luis N Marenco, Gordon M Shepherd, and Prakash Nadkarni. Entity-Attribute-Value database approaches for heterogeneous, evolving neuroscience. In S H Koslow and S Subramaniam, editors, *Databasing the Brain: From Data to Knowledge*, pages 83–97. Wiley & Sons, 2005. ISBN 0-471-30921-4.

Asif Jan Muhammad and Henry Markram. NEOBASE: Databasing the Neocortical Microcircuit. In Tony Solomonides, Richard McClatchey, Vincent Breton, Yannick Legre, and Sofie Norager, editors, *From Grid to Healthgrid - Proceedings of Healthgrid 2005*, pages 167–177, 2005. doi: 978-1-58603-510-5.

Hans-Michael Müller, Eimear E Kenny, and Paul W Sternberg. Textpresso: an ontology-based information retrieval and extraction system for biological literature. *PLoS biology*, 2(11):e309, November 2004. ISSN 1545-7885. doi: 10.1371/journal.pbio.0020309. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=517822\&tool=pmcentrez\&rendertype=abstract.

Robert F Murphy. Cytomics and location proteomics: automated interpretation of subcellular patterns in fluorescence microscope images. *Cytometry. Part A : the journal of the International Society for Analytical Cytology*, 67 (1):1–3, September 2005. ISSN 1552-4922. doi: 10.1002/cyto.a.20179. URL http://www.ncbi.nlm.nih.gov/pubmed/16082712.

Eric Neumann and Larry Prusak. Knowledge networks in the age of the Semantic Web. *Briefings in bioinformatics*, 8(3):141–9, May 2007. ISSN 1467-5463. doi: 10.1093/bib/bbm013. URL http://www.ncbi.nlm.nih.gov/pubmed/17502336.

Finn Å rup Nielsen. Brede Wiki: A neuroinformatics Web service with structured information. In *2nd INCF Congress of Neuroinformatics*. Informatics and Mathematical Modelling, Technical University of Denmark, 2009a. doi: 10.3389/conf.neuro.11.2009.08.072. URL http://frontiersin.org/Community/AbstractDetails.aspx?ABS_DOI=10.3389/conf.neuro.11.2009.08.072.

Finn Arup Nielsen. Lost in localization: a solution with neuroinformatics 2.0? *NeuroImage*, 48(1):11–3, October 2009b. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2009.05.073. URL http://www.ncbi.nlm.nih.gov/pubmed/19497377.

Natalya F Noy, Nigam H Shah, Patricia L Whetzel, Benjamin Dai, Michael Dorf, Nicholas Griffith, Clement Jonquet, Daniel L Rubin, Margaret-Anne Storey, Christopher G Chute, and Mark a Musen. BioPortal: ontologies and integrated data resources at the click of a mouse. *Nucleic acids research*, 37(Web Server issue):W170–3, July 2009. ISSN 1362-4962. doi: 10.1093/nar/gkp440. URL http://www.ncbi.nlm.nih.gov/pubmed/19483092.

N.F. Noy, Michael Sintek, Stefan Decker, Monica Crubézy, R.W. Fergerson, and M.A. Musen. Creating semantic web contents with protege-2000. *Intelligent*

Systems, IEEE, 16(2):60–71, 2001. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=920601.

A.L Peters, S Palay, and H D F Webster. *The fine structure of the nervous system*. Harper and Row, New York, 1991.

A Reichenbach and H Wolberg. Astrocytes and Ependymal glia. chapter 2. Oxford University Press, New York, 2005.

Lorna Richardson, Shanmugasundaram Venkataraman, Peter Stevenson, Yiya Yang, Nicholas Burton, Jianguo Rao, Malcolm Fisher, Richard a Baldock, Duncan R Davidson, and Jeffrey H Christiansen. EMAGE mouse embryo spatial gene expression database: 2010 update. *Nucleic acids research*, 38 (Database issue):D703–9, January 2010. ISSN 1362-4962. doi: 10.1093/nar/gkp763. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2808994\&tool=pmcentrez\&rendertype=abstract.

C Rosse, J L Mejino, B R Modayur, R Jakobovits, K P Hinshaw, and J F Brinkley. Motivation and organizational principles for anatomical knowledge representation: the digital anatomist symbolic knowledge base. *Journal of the American Medical Informatics Association : JAMIA*, 5(1):17–40, 1998a. ISSN 1067-5027. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=61273\&tool=pmcentrez\&rendertype=abstract.

C Rosse, L G Shapiro, and J F Brinkley. The digital anatomist foundational model: principles for defining and structuring its concept domain. *Proceedings / AMIA ... Annual Symposium. AMIA Symposium*, pages 820–4, January 1998b. ISSN 1531-605X. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2232229\&tool=pmcentrez\&rendertype=abstract.

Cornelius Rosse and José L V Mejino. A reference ontology for biomedical informatics: the Foundational Model of Anatomy. *Journal of biomedical informatics*, 36(6):478–500, December 2003. ISSN 1532-0464. doi: 10.1016/j.jbi.2003.11.007. URL http://www.ncbi.nlm.nih.gov/pubmed/14759820.

Badrinath Roysam, William Shain, and Giorgio a Ascoli. The central role of neuroinformatics in the National Academy of Engineering's grandest challenge: reverse engineer the brain. *Neuroinformatics*, 7(1):1–5, January 2009. ISSN 1559-0089. doi: 10.1007/s12021-008-9043-9. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2726926\&tool=pmcentrez\&rendertype=abstract.

Daniel L Rubin, Natalya F Noy, and Mark a Musen. Protégé: a tool for managing and using terminology in radiology applications. *Journal of digital imaging : the official journal of the Society for Computer Applications in Radiology*, 20 Suppl 1(0):34–46, November 2007. ISSN 0897-1889. doi: 10.1007/s10278-007-9065-0. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2039856\&tool=pmcentrez\&rendertype=abstract.

S J Russell and Peter Norvig. *Artificial Intelligence, A Modern Approach - 2nd Edition.pdf*, volume 82 of *Prentice Hall series in artificial intelligence*. Prentice hall Englewood Cliffs, NJ, 2003. ISBN 0136042597. doi: 10.1016/0004-3702(96)00007-0. URL http://www.cis.uab.edu/courses/cs760/Spring-660-2007/

7A-660-PLUS-SYLLABUS-DRAFT.pdf.

Alan Ruttenberg, Jonathan A Rees, Matthias Samwald, and M Scott Marshall. Life sciences on the Semantic Web: the Neurocommons and beyond. *Briefings in Bioinformatics*, 10(2):193–204, 2009. URL http://www.ncbi.nlm.nih.gov/pubmed/19282504.

Jürgen Rybak, Anja Kuß, Hans Lamecker, Stefan Zachow, Hans-Christian Hege, Matthias Lienhard, Jochen Singer, Kerstin Neubert, and Randolf Menzel. The Digital Bee Brain: Integrating and Managing Neurons in a Common 3D Reference System. *Frontiers in systems neuroscience*, 4 (July):1–15, January 2010. ISSN 1662-5137. doi: 10.3389/fnsys.2010.00030. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2935790\&tool=pmcentrez\&rendertype=abstract.

Clay Shirky. Shirky: Ontology is Overrated – Categories, Links, and Tags, 2005. URL http://www.shirky.com/writings/ontology_overrated.html.

Nicholas Sioutos, Sherri de Coronado, Margaret W Haber, Frank W Hartel, Wen-Ling Shaiu, and Lawrence W Wright. NCI Thesaurus: a semantic model integrating cancer-related clinical and molecular information. *Journal of biomedical informatics*, 40(1):30–43, February 2007. ISSN 1532-0480. doi: 10.1016/j.jbi.2006.02.013. URL http://www.ncbi.nlm.nih.gov/pubmed/16697710.

E Sirin, B Parsia, B Grau, a Kalyanpur, and Y Katz. Pellet: A practical OWL-DL reasoner. *Web Semantics: Science, Services and Agents on the World Wide Web*, 5(2):51–53, June 2007. ISSN 15708268. doi: 10.1016/j.websem.2007.03.004. URL http://linkinghub.elsevier.com/retrieve/pii/S1570826807000169.

B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L.J. Goldberg, K. Eilbeck, A. Ireland, C.J. Mungall, OBI Consortium., Neocles. Leontis, Philippe. Rocca-Serra, Alan. Ruttenberg, Susanna-Assunta. Sansone, R. Scheuermann, Nigam. Shah, Patricia L. Whetzel, and Suzanna Lewis. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature biotechnology*, 25(11):1251–1255, 2007. doi: 10.1038/nbt1346.The. URL http://www.nature.com/nbt/journal/vaop/ncurrent/full/nbt1346.html.

Barry Smith and Cornelius Rosse. The role of foundational relations in the alignment of biomedical ontologies. *Studies in health technology and informatics*, 107 (Pt 1):444–8, January 2004. ISSN 0926-9630. URL http://www.ncbi.nlm.nih.gov/pubmed/15360852.

Barry Smith, Werner Ceusters, Bert Klagges, Jacob Köhler, Anand Kumar, Jane Lomax, Chris Mungall, Fabian Neuhaus, Alan L Rector, and Cornelius Rosse. Relations in biomedical ontologies. *Genome Biology*, 6(5):R46, 2005. URL http://www.ncbi.nlm.nih.gov/pubmed/15892874.

Harold Solbrig and Guoqian Jiang. The BiomedGT Wiki, 2009. URL http://j.mp/p1w00G.

Gina E Sosinsky, Thomas J Deerinck, Rocco Greco, and Casey H Buitenhuys. Development of a Model for Microphysiological Small Nodes of Ranvier From

Peripheral Nerves of Mice Reconstructed by Electron Tomography. *Neuroinformatics*, pages 133–162, 2005. doi: 10.1385/NI.

Diomidis Spinellis and Panagiotis Louridas. The collaborative organization of knowledge. *Communications of the ACM*, 51(8):68, August 2008. ISSN 00010782. doi: 10.1145/1378704.1378720. URL `http://portal.acm.org/citation.cfm?doid=1378704.1378720`.

Peter Spyns, Robert Meersman, and Mustafa Jarrar. Data modelling versus ontology engineering. *ACM SIGMOD Record*, 31(4):12, December 2002. ISSN 01635808. doi: 10.1145/637411.637413. URL `http://portal.acm.org/citation.cfm?doid=637411.637413`.

K E Stephan, L Kamper, a Bozkurt, G a Burns, M P Young, and R Kötter. Advanced database methodology for the Collation of Connectivity data on the Macaque brain (CoCoMac). *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 356(1412):1159–86, August 2001. ISSN 0962-8436. doi: 10.1098/rstb.2001.0908. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1088509\&tool=pmcentrez\&rendertype=abstract`.

Roger D Traub, Diego Contreras, Mark O Cunningham, Hilary Murray, Fiona E N LeBeau, Anita Roopun, Andrea Bibbig, W Bryan Wilent, Michael J Higley, and Miles A Whittington. Single-column thalamocortical network model exhibiting gamma oscillations, sleep spindles, and epileptogenic bursts. *Journal of Neurophysiology*, 93(4):2194–2232, 2005. URL `http://www.ncbi.nlm.nih.gov/pubmed/15525801`.

Tania Tudorache, N. Noy, Samson Tu, and M. Musen. Supporting collaborative ontology development in Protégé. *The Semantic Web-ISWC 2008*, pages 17–32, 2010. URL `http://www.springerlink.com/index/uu1453660t047775.pdf`.

Tania Tudorache, N.F. Noy, S.M. Falconer, and M.A. Musen. A knowledge base driven user interface for collaborative ontology development. In *Proceedings of the 15th international conference on Intelligent user interfaces*, pages 411–414. ACM, 2011. URL `http://portal.acm.org/citation.cfm?id=1943478`.

Jessica a Turner, Jose L V Mejino, James F Brinkley, Landon T Detwiler, Hyo Jong Lee, Maryann E Martone, and Daniel L Rubin. Application of neuroanatomical ontologies for neuroimaging data annotation. *Frontiers in neuroinformatics*, 4(June):1–12, January 2010. ISSN 1662-5196. doi: 10.3389/fninf.2010.00010. URL `http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2912099\&tool=pmcentrez\&rendertype=abstract`.

Pedro A Valdés-Hernández. An in vivo MRI rat template set in the Paxinos and Watson space: Applications in fMRI, morphometry and anatomical connectivity, 2010.

D. Vrandecic and M. Krötzsch. Reusing ontological background knowledge in semantic wikis. In *Proceedings of the 1st Workshop on Semantic Wikis*. Citeseer, 2006. URL `http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.87.9182\&rep=rep1\&type=pdf\#page=24`.

H Wallach and D N O'Connell. The kinetic depth effect. *Journal of Experi-*

mental Psychology, 45(4):205–217, 1953. URL http://www.ncbi.nlm.nih.gov/pubmed/13052853.

Jintao Wang, Robert W Williams, and Kenneth F Manly. WebQTL: Web-Based Complex Trait Analysis. *Neuroinformatics*, 1:299–308, 2003.

G Paxinos C Watson. *The Rat Brain in Stereotacxic Coordinates*. Elsevier, 2005.

Tal Yarkoni, Russell a. Poldrack, David C. Van Essen, and Tor D. Wager. Cognitive neuroscience 2.0: building a cumulative science of human brain function. *Trends in Cognitive Sciences*, 14(11):489–496, September 2010. ISSN 13646613. doi: 10.1016/j.tics.2010.08.004. URL http://linkinghub.elsevier.com/retrieve/pii/S1364661310002019.

Lynn Young, Samson W. Tu, Lakshika Tennakoon, David Vismer, Vadim Astakhov, Amarnath Gupta, Jeffrey S. Grethe, Maryann E. Martone, Amar K. Das, and Matthew J. McAuliffe. Ontology Driven Data Integration for Autism Research. *2009 22nd IEEE International Symposium on Computer-Based Medical Systems*, pages 1–7, August 2009. doi: 10.1109/CBMS.2009.5255362. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5255362.

Ilya Zaslavsky, Haiyun He, Joshua Tran, Maryann E. Martone, and Amarnath Gupta. Integrating Brain Data Spatially: Spatial Data Infrastructure and Atlas Environment for Online Federation and Analysis of Brain Images. In *2nd International Workshop on Biological Data Management*, 2004.

Wuxue Zhang, Yong Zhang, Hui Zheng, Chen Zhang, Wei Xiong, John G Olyarchuk, Michael Walker, Weifeng Xu, Min Zhao, Shuqi Zhao, Zhuan Zhou, and Liping Wei. SynDB: a Synapse protein DataBase based on synapse ontology. *Nucleic acids research*, 35(Database issue):D737–41, January 2007. ISSN 1362-4962. doi: 10.1093/nar/gkl876. URL http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1669723\&tool=pmcentrez\&rendertype=abstract.