Rule-Based Reasoning With A Multi-Scale Neuroanatomical Ontology

Stephen D. Larson and Maryann E. Martone
{slarson,maryann}@ncmir.ucsd.edu

National Center for Microscopy and Imaging Research,
Center for Research in Biological Structure and Development of Neurosciences,
University of California, San Diego, La Jolla, CA 92093-0608, USA

Abstract. We have demonstrated the use of OWL-based ontologies to infer novel facts from existing knowledge. Using an ontology developed for subcellular anatomy of the nervous system, the SAO, merged with a terminological resource for gross neuroanatomy (Brain Architecture Management System; Bota and Swanson, 2005), we have inferred the existence of axons from the existence of axon terminals and the existence of neurons from the existence of axons. Additionally we have inferred connectivity at a large scale from the existence of synapses at a micro scale. We conclude that OWL is an important technology for merging disparate data and performing multi-scale reasoning.

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 thatneuroscientists 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[1]. 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 specification[2] to construct an ontology of subcellular anatomy [13]. 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 multi-scale 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.

2 OWL Format Enables Ontologies To Be True Data Models

The W3C's OWL specification[2] is an extension of RDF[3], which is itself an extension of XML[4]. As reported in[14], 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[13], we describe the creation of an ontology[1] for subcellular anatomy of the nervous system[2]. Unlike most ontologies developed around gross anatomy, e.g., the Foundational Model of Anatomy[11] 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 1. Many of the relationships between elements in the data are explicitly described, such as in table 1, but many other relationships that could result from description in theSAO 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.

---

[1] The SAO; http://ccdb.ucsd.edu/SAO/1.0/SAO.owl
[2] The version of the SAO used for this work is a predecessor to version 1.0

Table 1: Select Properties of Neuron in SAO

1. has_Neurotransmitter
2. is_Spiny
3. number_Of_Axons
4. number_Primary_Dendrites_Min
5. projects_To
6. has_Compartment
7. has_Component
8. has_Dimension
9. has_Inherent_3D_Shape
10. anatomical_Location_Atlas
11. anatomical_Location_Specific
12. functional_State
13. location_CNS_PNS
14. morphological_Type
15. is_continuous_with
16. observation_Conditions
17. has_parent_cell
18. species_of_origin

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 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, theSAO 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) having its origin in the posterior complex of the thalamus, 2) having glutamate as an associated molecular constituent, 3) having its location in the primary somatosensory area's barrel field, 4) being a compartment of an axon named *Axon_1*, and 5) being 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)[15], 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 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   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)[15]. To collect data on the level of brain anatomy and connectivity between brain regions, we turned to the BAMS website. 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. The BAMS ontology has only two major categories - BrainPart and ConnectionStatement. 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.

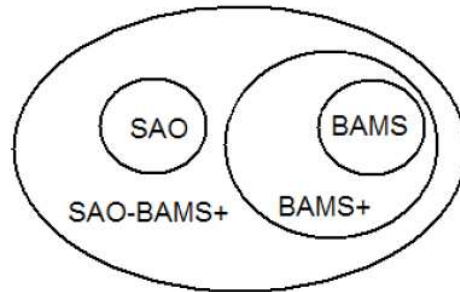Figure 1 illustrates the structure of imports for the SAO merged with BAMS.

Fig. 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.

4 Rule-based Reasoning Allows Inferences To Generate New Knowledge

In the following, we show how we utilized the SAO-BAM+ 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[10]. 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* the cell soma. Almost all electron microscopic data involves the analysis of partial structures,
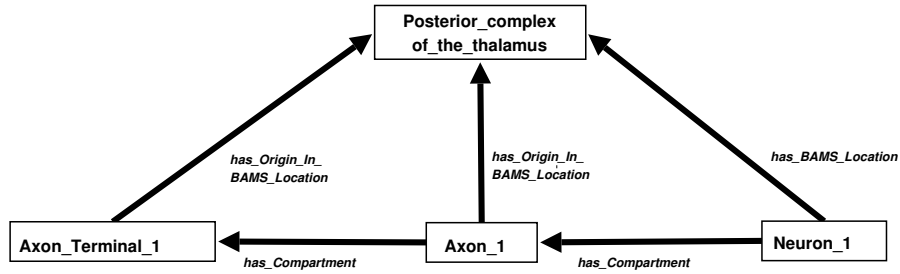
Fig. 2. An illustration of how reasoning enables the creation of new information from old information through logical inference.

i.e. isolated parts of dendrites, axons, and 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 2 illustrates how having some knowledge about where the axon terminal is coming from can be leveraged to infer new information.

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)))
```

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)))
```

Common sense 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 (Figure 2). Here we demonstrate schematically the result of applying Algernon[7] rules 1.1 and 1.2 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.
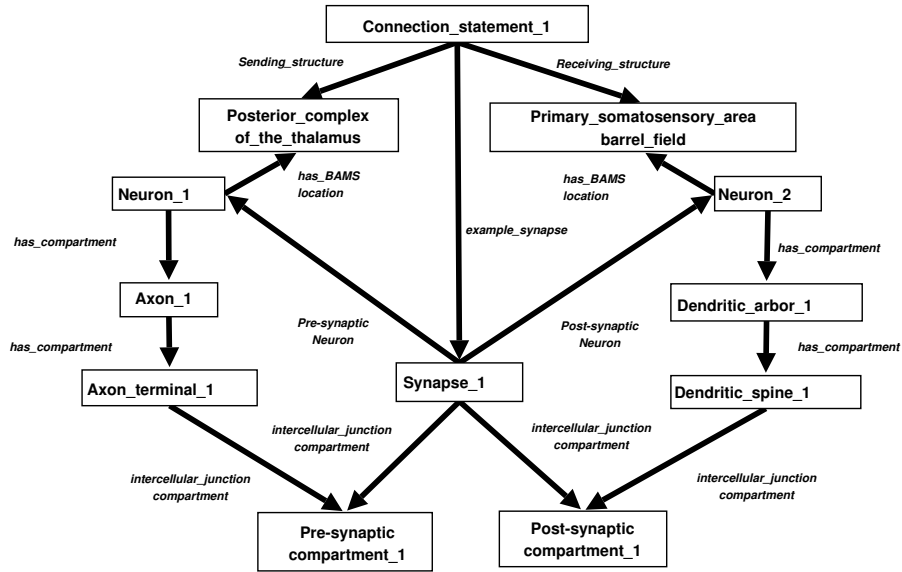

Fig. 3. Reasoning across three levels of anatomical scale, merging the cellular, supracellular and gross anatomy.

Figure 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 related to *Axon_Terminal_1* and our post-synaptic compartment is related to *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

has_BAMS_location relations, Algernon rule 2.3 can make an explicit connection statement about the brain areas that the neurons are found within. Additionally the connection statement to *Synapse_1* can be flagged with *Synapse_1* as an *example_synapse*, providing a powerful means of organizing the data in the brain across scales. 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.

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))
```

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))
```

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)))
```

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. 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 [13]), 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.

6 Contributions

In this paper, we have demonstrated that in order to perform inference across physical scales in neuroscience, one requires data organized in a representation that enables integration. OWL-based ontologies are ideal common grounds for many neuroscientific data because they allow information from different sources and on different topics to be analyzed together 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 commonsense 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.

7 Appendix A: 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[7], 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 rules files are available at `http://ccdb.ucsd.edu/SAO`.

References

1. "Putting the brain back together", Editorial, Nature Neuroscience, http://www.nature.com/doifinder/10.1038/nn0406-457
2. OWL Web Ontology Language Overview, W3C Recommendation 10 February 2004, http://www.w3.org/TR/owl-features/
3. Resource Description Framework (RDF), W3C Recommendation, May 2001, http://www.w3.org/RDF/
4. Extensible Markup Language (XML), W3C Specification, 2003, http://www.w3.org/XML/
5. Protege, Stanford Medical Informatics at the Stanford University School of Medicine http://protege.stanford.edu/
6. http://protege.stanford.edu/community/conferences.html
7. Micheal Hewett, Algernon - Rule-Based Programming, http://algernon-j.sourceforge.net/
8. Martin OConnor, Holger Knublauch, Samson Tu, Mark Musen, Writing Rules for the Semantic Web using SWRL and JESS, Stanford Medical Informatics, Stanford University School of Medicine, http://www.med.univ-rennes1.fr/ cgolb/Protege2005/SWRLJessOConnor.pdf
9. Jess is a rule engine and scripting environment written entirely in Sun's JavaTM language by Ernest Friedman-Hill at Sandia National Laboratories in Livermore, CA.
10. C. Golbreich, O. Bierlaire, O. Dameron, B. Gibaud. What reasoning support for ontology and rules? the brain anatomy case study: workshop on Protg and rules . 8th International Protg Conference, 2005.
11. Martin, RF, Mejino, JL Jr, Bowden, DM, Brinkley, JF 3rd, Rosse, C. (2001) Foundational model of neuroanatomy: implications for the Human Brain Project. Proc AMIA Symp 438-42.
12. W3C Member Submission, SWRL: A Semantic Web Rule Language Combining OWL and RuleML, http://www.w3.org/Submission/SWRL/, 21 May 2004
13. Fong et al., An Ontology-Driven Knowledge Environment For Subcellular Neuroanatomy, submitted to OWLED 2007, March 2007
14. Martone ME, Gupta A, Ellisman MH. (2004) E-neuroscience: challenges and triumphs in integrating distributed data from molecules to brains. Nat Neurosci. 7: 467-72.
15. Bota M, Dong HW, Swanson LW. Brain architecture management system. Neuroinformatics. 2005;3(1):15-48.