Published in final edited form as:

Artif Intell Med. 2010 January ; 48(1): 21–28. doi:10.1016/j.artmed.2009.11.003.

Semantic SenseLab: implementing the vision of the Semantic Web in neuroscience

Matthias Samwald^{a,b,c,d}, Huajun Chen^{a,e}, Alan Ruttenberg^f, Ernest Lim^a, Luis Marengo^{a,g}, Perry Miller^{a,g,h}, Gordon Shepherdⁱ, and Kei-Hoi Cheung^{a,g,j,k}

^aCenter for Medical Informatics, Yale University School of Medicine, 300 George Street, New Haven, CT 06520-8009, USA ^bDigital Enterprise Research Institute, National University of Ireland Galway, IDA Business Park, Lower Dangan, Galway, Ireland ^cKonrad Lorenz Institute for Evolution and Cognition Research, Adolf Lorenz Gasse 2 A-3422, Altenberg, Austria ^dSection on Medical Expert and Knowledge-Based Systems, Medical University of Vienna, Spitalgasse 23, A-1090 Vienna, Austria ^eCollege of Computer Science, Zhejiang University, 310027 Hangzhou, China ^fScience Commons, c/o Massachusetts Institute of Technology Computer Science and Artificial Intelligence Laboratory, Building 32-386D, 32 Vassar Street, Cambridge, Massachusetts 02139, USA ^gDepartment of Anesthesiology, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06520-8051, USA ^hDepartment of Molecular, Cellular and Developmental Biology, Yale University, New Haven, CT 06520-8009, USA ⁱDepartment of Neurobiology, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06520-8051, USA ^jDepartment of Genetics, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06520-8005, USA ^kDepartment of Computer Science, Yale University, 51 Prospect Street, New Haven, CT 06511, USA

Summary

Objective—Integrative neuroscience research needs a scalable informatics framework that enables semantic integration of diverse types of neuroscience data. This paper describes the use of the Web Ontology Language (OWL) and other Semantic Web technologies for the representation and integration of molecular-level data provided by several of SenseLab suite of neuroscience databases.

Methods—Based on the original database structure, we semi-automatically translated the databases into OWL ontologies with manual addition of semantic enrichment. The SenseLab ontologies are extensively linked to other biomedical Semantic Web resources, including the Subcellular Anatomy Ontology, Brain Architecture Management System, the Gene Ontology, BIRN Lex and UniProt. The SenseLab ontologies have also been mapped to the Basic Formal Ontology and Relation Ontology, which helps ease interoperability with many other existing and future biomedical ontologies for the Semantic Web. In addition, approaches to representing contradictory research statements are described. The SenseLab ontologies are designed for use on the Semantic Web that enables their integration into a growing collection of biomedical information resources.

© 2009 Elsevier B.V. All rights reserved.

Correspondence address: Matthias Samwald, Konrad Lorenz Institute for Evolution and Cognition Research, Adolf Lorenz Gasse 2, A-3422 Altenberg, Austria, samwald@gmx.at, Phone: +43-2242-32390-19, Fax: +43-2242-323904.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Conclusion—We demonstrate that our approach can yield significant potential benefits and that the Semantic Web is rapidly becoming mature enough to realize its anticipated promises. The ontologies are available online at <http://neuroweb.med.yale.edu/senselab/>

Keywords

Semantic Web; neuroscience; description logic; ontology mapping; Web Ontology Language; integration

1. Introduction

Neuroscience is in need of a new informatics framework that enables semantic integration of diverse data sources [1]. Experimental data is collected across different scales, from cell to tissue to organ, using a wide variety of experimental procedures taken from diverse disciplines. Unfortunately the information systems holding these data do not link related data among them, preventing effective research that could combine the data to achieve new insights. Integrative neuroscience research is key to providing a better understanding of many neurological diseases such as Alzheimer's Disease and Parkinson's Disease, and could potentially lead to a better prevention, diagnosis and treatment of such diseases. The Semantic Web, a maturing set of technologies and standards backed by the World Wide Web consortium [2], offers technical guidance specifically in the area of aggregating and integrating diverse information resources. These Semantic Web technologies can be used to integrate neuroscience knowledge and to make such integrated knowledge more easily accessible to researchers. The foundational technologies of the Semantic Web — Resource Description Framework (RDF [3]), Web Ontology Language (OWL [4]), the SPARQL Protocol and RDF Query Language (SPARQL) — are widely implemented and are backed by a large community of users and developers. The chief advantages of Semantic Web technologies include 1) the widely supported standards backed by the World Wide Web consortium, 2) the ability to make use of the well-established inference mechanisms of description logics, and 3) the availability of a wide range of software tools.

A demonstration of Semantic Web technologies in the neuroscience domain [5-7] has been carried out, in the context of translational research, by the *Semantic Web for Health Care and Life Science Interest Group* of the World Wide Web Consortium. A major goal of translational research is to accelerate the bidirectional communication between basic research and clinical practice, in order to speed up the development of new clinical guidelines, tests, and therapies. The Semantic Web has the potential to facilitate the aggregation and integration of information from different institutions involved in this process.

As part of this community effort, we have created a Semantic Web framework for neuroscience research, based on the *SenseLab* collection of databases [8]. Senselab is a highly accessed information resource for neuroscience research on the Web [9]. Another motivation for converting SenseLab into Semantic Web format was that the “entity-attribute-value with classes and relationships” schema (EAV/CR [10]) on which SenseLab's architecture is based bears considerable resemblance to RDF. As a result, the conversion of SenseLab into the Semantic Web format (e.g., RDF) is facilitated. In fact, we have written a program to automatically convert SenseLab databases in the corresponding RDF structure. Such converted RDF-formatted data can then be loaded into an RDF store (e.g., Oracle RDF Data Model) for RDF-based querying. While we have demonstrated that a straightforward syntactic conversion can be done automatically, the RDF representation has limited expressivity and reusability. For example, RDF is mostly focused on the description of instances and does not allow for the detailed description of class properties, relations

between classes, and automated classification that is central to our integration efforts. It does not offer constructs to describe sameness between entities from different data sources. RDF also lacks important features to enforce consistency checks to identify erroneous and contradictory statements, which is an essential feature when large, complex information repositories need to be merged.

To overcome these limitations, we use a more expressive ontology language, the Web Ontology Language (OWL), for representing richer semantics and logical statements. In addition, we adopt the current ontological standards and best practices in the process of creating the SenseLab ontologies. A goal is to allow the ontologies to have broad interoperability and reusability.

SenseLab databases

SenseLab consists of a number of specialized databases, three of which we have converted to the Semantic Web format: *NeuronDB*, *BrainPharm* and *ModelDB*. *NeuronDB* contains descriptions of anatomic locations, cell architecture and physiologic parameters (membrane properties consisting of transmitters, receptors and ionic channels) of neuronal cells based on compartmental models of neurons (Fig. 1). The pilot *BrainPharm* database is intended to support research on drugs for the treatment of neurological disorders. It enhances the descriptions in a portion of *NeuronDB* with descriptions of the actions of pathological and pharmacological agents. *ModelDB* is a large repository of computational neuroscience models and simulations. The computational models in *ModelDB* are annotated with references to *NeuronDB*. Taken together, these databases allow the researcher to query information and to run simulations pertaining to the function of neurons in healthy and disease states. The *NeuronDB* and *ModelDB* databases contain literature references and excerpts from texts that have been used to curate the database entries. This allows the users of SenseLab to verify the information in the database and can act as a starting point for further literature searches. The highly interconnected and hierarchical nature of these scientifically annotated data makes them suitable candidates for the creation of a Semantic Web resource in neuroscience.

2. Methods

This section describes the process of constructing the ontologies and converting data extracted from the SenseLab databases into the ontological structure. In addition, we discuss how to establish mappings from SenseLab ontologies to other existing ontologies. Finally, we mention the quality control and reasoning capability supported by OWL.

2. 1. Basic ontology development

An ontology ‘scaffold’ made up of basic class hierarchies and relations was manually created, based on the structure of existing SenseLab databases. This scaffold could not be created by an automated process, since some of the structures and entity labels in the database needed to be slightly changed and re-interpreted to create a logically consistent and well-designed ontology.

The design of this scaffold was inspired by the realism described by Smith [11]. The ontologies are primarily organized around direct representations of physical objects and processes (e.g., neuronal cells, ionic currents) in reality, and not around their abstractions (e.g., concepts and database entries). This approach has already been adopted for developing standard biomedical ontologies like those included in the Open Biomedical Ontologies Foundry (OBO Foundry [12]), one of the widely recognized community projects in the area of biomedical ontologies.

The scaffold contains basic classes from the domain of neuroscience, such as ‘brain region’, ‘neuron’, ‘gene’, and ‘serotonin receptor’ (subclass of ‘receptor’). It provides the semantic foundation for data querying, integration and inferencing. For example, based on certain user-defined relationships (e.g., a gene *encodes* a receptor) between different classes, semantic queries can be formulated to answer focused neuroscientific research questions (e.g., *serotonin receptors* are found in specific type(s) of *neurons*). Based on the hierarchical relationship between brain regions, we can infer child/parent regions at any level automatically. Some of the classes (e.g., neurons) can serve as a unit of integration across different data sources. For example, research statements about a particular neuron may be integrated from different databases.

For editing and viewing the SenseLab ontologies, we evaluated several OWL ontology editors including *Protégé 3.2* [13], *Swoop 2.3 alpha* [14, 15] and *TopBraid Composer 2.0* [16]. While the first two are open source, the third is a commercial product. We started with Protégé but experienced some difficulties: i) certain uniform resource identifiers (URIs) that could not be decomposed into XML QNames were not displayed correctly, ii) namespaces and ontology import hierarchies were not handled as expected, and iii) some of the statements automatically created by Protégé did not adhere to the OWL DL standard. While we did not encounter these problems when using Swoop and TopBraid Composer, these ontology editors were not as stable as we had expected. To sum up, more stable, standards-compliant and robust ontology editors are needed for serious ontology design and editing.

The ontologies were mainly developed by a small group of people, and no dedicated software for collaborative ontology editing was used. This worked well for the scope of the current SenseLab ontologies. However, if future SenseLab ontology development involves a wider scope and a greater number of participants, it will make sense to use such software to minimize versioning conflicts.

The ontologies were built upon established foundational ontologies in order to maximize the interoperability with other existing and forthcoming biomedical Semantic Web resources. These ontologies were the Relation Ontology [17, 18] from the Open Biomedical Ontologies repository (OBO [19]), which defines basic relations such as ‘part of’, ‘participant of’ or ‘contained in’; and the Basic Formal Ontology (BFO [20]), which defines basic classes such as ‘process’, ‘object’, ‘quality’ or ‘function’. In [21], the SenseLab ontologies presented here are listed as one of the primary examples of the application of OBO Foundry resources.

2. 2. Data conversion

The data from the SenseLab databases were automatically converted to OWL using programs written in Java and Python. The automated export scripts extended the manually created ontology scaffolds through the creation of subclasses, OWL property restrictions and individuals. In OWL ontologies like the ones created for the SenseLab project, the distinction between ‘ontology’, ‘data schema’ and ‘data’ is blurred. The main practical difference between ontology development and data conversion in our project was that the basic ontological structures needed to be developed manually, while the bulk of ‘data’ could be converted through automated processes. The resulting ontologies show no clearly distinguishable divide between a schema and data.

The OWL export of NeuronDB was based on a transformation from the EAV/CR model of the SenseLab database [22] to RDF serialized as XML (RDF/XML) by a Java program. The transformed information included descriptions of neurons based on research findings (e.g., neuronal receptors, channels and transmitters). The classes and individuals created by these exports were added to the manually created ontology scaffold as subclasses and instances of the classes in the ontology scaffold.

The export from ModelDB and BrainPharm was based on a simple flat text file export of the databases. The text file exports were converted to RDF/XML files with a Python script.

The mapping from neuron receptors to corresponding genes was based upon an automated transformation on the EntrezGene MySQL dump provided by Atlas [23]. In the ontology, the neuron receptors were defined as gene products of genes. The genes in that mapping were identified by their common gene symbols.

Based on this list of gene symbols in the ontology, a mapping between gene symbols and *NCBI Entrez Gene* record identifiers was generated with the *Clone/Gene ID Converter* [24]. This service returned the mapping between gene symbols and identifiers as a tab-delimited text file. The mapping in this text file was used for the generation of an RDF/XML file with a Python script. The RDF/XML was then merged with the main NeuronDB ontology file.

Based on the annotation of receptor proteins with gene symbols, receptor proteins were also annotated with Uniprot records that corresponded to the genes. A mapping between gene symbols and *Uniprot* record identifiers was generated with the *SOURCE* gene annotation service [25]. Again, the resulting tab-delimited text file was used to generate RDF/XML which was merged with the main NeuronDB ontology file. Literature references in the source database were converted to references to *NCBI Pubmed* database entries.

For all of these mappings, we used the URI scheme for database record identifiers established by Science Commons [26]. URIs for database records could simply be generated by concatenating the record identifier to a predefined namespace. For example, the Entrez Gene record with ID ‘3579’ was identified by the URI ‘http://purl.org/commons/record/ncbi_gene/3579’, the Uniprot record ‘P46663’ was identified by ‘<http://purl.org/commons/record/uniprotkb/P46663>’ and the Pubmed record with ID ‘11160518’ was identified by ‘<http://purl.org/commons/record/pmid/11160518>’.

It should be noted that all entries in NCBI Gene and Uniprot are specific to certain animal species. While species specificity is indicated in the annotations and in the ModelDB files, the data entries in the NeuronDB ontology are species-agnostic — they provide general descriptions of mammalian and arthropod physiology, which covers a wide variety of use cases. In cases where species-specific information is required, the textual annotations will be taken into consideration. The NCBI Gene and Uniprot references in the annotations can therefore be seen as species-specific examples, i.e., they do not necessarily cover all homologue proteins from all species.

Research statements in the SenseLab database were interpreted as claims of existence of a certain class of neurons with certain properties, which were added as subclasses to the ontology scaffold. An example of the application of this modelling approach is described in Fig. 2. Information about research statements (e.g., descriptive text, Pubmed references) were attached to these classes. Generalized classes derived from all available research statements for a specific type of neurons were added, which resulted in the three-layer design pattern described in Fig. 2.

The use of OWL reasoning and the creation of manually curated generalizations of research findings make it possible to harness OWL for the formulation of generalized, internally consistent world-views based on changing and often contradictory research findings. The contradictions identified by OWL reasoning in this manner can help in localizing disagreement between different data and hypotheses, and can help in judging the validity of competing hypotheses.

It was found that the design pattern for the representation of research findings and evidence used in the SenseLab ontology was easily expressed consistently in OWL and was well integrated with other ontologies in our collection. Other approaches (e.g., the definition of named RDF subgraphs for each set of research statements) were also considered, but they did not meet these criteria.

2. 3. Connecting to other biomedical Semantic Web ontologies

The three ontologies representing the SenseLab data were mapped to several related Semantic Web ontologies from the domains of neuroscience and biomedicine: 1) the BAMS ontology (created by John Barkley, National Institute of Standards and Technology, USA) which was derived from the Brain Architecture Management System (BAMS [27, 28]); 2) the Subcellular Anatomy Ontology (SAO [29]) created by the Cell Centered Database project [30]; 3) the BirnLex ontology [31] developed by members of the Biomedical Informatics Research Network [32]; 4) the Common Anatomy Reference Ontology (CARO [33]); 5) the Gene Ontology [34]; 6) the Ontology of Biomedical Investigation (OBI) [35] (a mapping still quite rudimentary at the time of this writing). URIs from SenseLab ontologies are also referenced in the OWL version of the Psychoactive Drug Screening Program (PDS) Ki database of receptor-ligand interactions [36].

The mappings were created by a person with expertise in both ontology engineering and neuroscience, which was indispensable for carrying out this task. They were created with standard ontology editing software. No automated algorithms for ontology mapping were used.

2. 4. Quality control and automated reasoning

The W3C RDF validator [37], a web-based tool hosted by the World Wide Web consortium, was used for checking well-formedness of RDF/XML and basic RDF syntax validation. The Java-based reasoner Pellet 1.4 [38, 39] was used for consistency checking and classification. It turned out to be essential to check syntax and ontological consistency after each major step of ontology development, as both syntactic and semantic errors were often introduced through human error or malfunction of software tools.

OWL inference was used to test which neurons in the database were in accordance with one of the ‘canonical neuronal forms’ described by SenseLab, for example the canonical form “neuron having an axon and apical dendrite”. While such a classification could also be done manually, the use of automated reasoning has the potential to speed up the process and allows flexible re-classification of all neurons when the definitions of canonical forms should be changed.

However, the greatest utility of OWL reasoning did not lie in the inference of new relationships based on complex logical deductions, but rather on consistency checking and the avoidance of errors in the knowledge base. During the development of the knowledge base, some errors were identified through simple reasoning processes. For example, based on class disjoints in the ontology scaffold, the OWL reasoner pointed us to an error: some classes (e.g. ‘GABA’, which is a common acronym of ‘gamma-aminobutyric acid’) were subclasses of both ‘neurotransmitter’ and ‘receptor’, which was wrong. This was an error caused by the automated conversion — both the GABA transmitters and the GABA receptors were simply labeled with ‘GABA’ in the source database. The conversion algorithm generated URIs based on these labels, so they were represented with identical URIs (http://neuroweb.med.yale.edu/senselab/neuron_ontology.owl #GAB A). Since ‘neurotransmitter’ and ‘receptor’ were declared as disjoint in the ontology scaffold, we could identify this problem early on and revise our conversion scripts accordingly. This

error would have been noticed much later without the use of OWL reasoning, and would certainly have led to unexpected bugs in software that makes use of the ontology.

2. 5. Dissemination

The Web addresses for downloading or importing all OWL files of the SenseLab Semantic Web infrastructure are listed in [40]. OWL makes it possible to import these ontologies into future ontologies by a simple reference to the URL of the ontologies. The ontologies can also be queried via Hypertext Transfer Protocol (HTTP) with the SPARQL RDF query language. The SPARQL server is based on the open source version of Virtuoso [41], a web server with an integrated, highly scalable RDF database. Instructions for accessing the SPARQL server are available at [42].

3. Results

The resulting SenseLab Semantic Web ontology collection is made up of seven ontology modules. Each ontology module is available as a separate OWL file with a specific Web address. The ontologies conform to the “OWL DL” specifications so that they can be classified by standard description logic reasoners. The separate ontology files give users the flexibility to selectively import or query those ontologies with a particular focus. The dependencies between ontologies are encoded in the ontology files through OWL ‘import’ statements. OWL-aware software can use these statements to load recursively all required ontology modules from the Web.

The basic statistics for each ontology module are summarized in Table 1. The NeuronDB ontology [42], ModelDB ontology [43] and BrainPharm ontology [44] contain the bulk of data from the respective SenseLab databases, together with some additional references to the *NCBI Gene* sequence database and the *Uniprot* sequence database. The other ontologies are mainly comprised of links/mappings between the SenseLab ontologies and ontologies created by other groups.

The biological function of receptors and transmitters was represented through subclasses of the ‘Function’ class from BFO (Fig. 3). Where applicable, classes from the ‘molecular function’ branch of the Gene Ontology were used (e.g. ‘dopamine receptor activity’). When no corresponding classes could be found in the Gene Ontology, new classes were created as part of the NeuronDB ontology and placed in the existing hierarchy of classes from the Gene Ontology. For example, the class ‘Dopamine D1 receptor activity function’ was created as a subclass of ‘dopamine receptor activity’ from the Gene Ontology. Molecules were linked to their molecular functions through the ‘has function’ property. For example, the dopamine receptor class has the defining property

has_function some ‘dopamine receptor activity’

The motivation for this exercise was to enable interoperability with the Gene Ontology, and other domain ontologies that make use of the Gene Ontology. In this way, the widely accepted Gene Ontology can be used as a bridge between ontologies about neuroreceptors, a knowledge domain where a widely accepted standard ontology is still lacking. For example, if another group would develop their own ontology of neuroreceptors and would reference the Gene Ontology in a similar fashion, it would be possible to infer class equivalence between the independently developed ontologies based on the references to the Gene Ontology.

The ‘has part’ relation from the OBO Relation Ontology found extensive use in the ontology. For example, the anatomic structure of the Archicortex was described with the following restrictions

has_part some Dentate

has_part some Hippocampus

The Hippocampus was described with

has_part some 'CA1 oriens alveus interneuron'

has_part some 'CA1 pyramidal neuron'

has_part some 'CA3 pyramidal neuron'

The finding that some CA1 pyramidal neurons have receptors for the neurotransmitter GABA in the Soma region was captured by the creation of a class with the following properties

has_part some ('Soma' that 'has receptors' some 'GABA receptor')

Some basic examples of queries that are possible based on the SenseLab ontologies are listed in Table 2.

In RDF/OWL, relations between entities defined in different ontologies do not differ from relations defined inside a single ontology, i.e., querying and inferencing can be done over several ontologies as if they were one. Classes in the SenseLab ontologies were connected to classes in other ontologies through class equivalence relations, class-subclass relations and whole-part relations. Examples for such relations spanning different ontologies are given in Fig 4. The import dependencies between ontology modules are depicted in Fig. 5.

The SenseLab ontologies presented in this paper are part of the Health Care and Life Science demo [45] of the W3C Semantic Web for Health Care and Life Science Interest Group and Science Commons. The demo consists of a large collection of ontologies and RDF data from the biomedical domain. It has been further extended and maintained by Science Commons, forming the 'Neurocommons Knowledge Base' [45]

4. Discussion

We reaped several benefits from the use of Semantic Web standards and tools. The integration of the SenseLab ontology with several other neuroscientific Semantic Web resources was easily accomplished based on the foundational ontologies. The use of established ontologies like BFO and the Gene Ontology has led to a clear, consistent and transparent representation of biological reality that would not have been readily achieved with relational databases or XML documents. This facilitates shared understanding between developers as well as between users of the ontology. Furthermore, the semantics associated with ontology constructs are described in human-readable form directly in the ontologies, which makes most ontologies self-documenting. The use of OWL ontologies helped us focus our work on the description of biological reality, and less on unnecessary artefacts such as database tables, columns or documents. OWL reasoning and consistency checking allowed the automatic identification of logical errors introduced during data entry and conversion, as well as true contradictions in the research information. Many of these errors and contradictions would not have been identified without the use of reasoners and would have caused complications or incomplete results when querying and mapping the ontologies.

The use of foundational ontologies such as BFO [20] or the Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) [47] is beneficial and in certain cases indispensable for the integration of independent ontologies. Foundational ontologies allow the creators of domain ontologies to reuse basic ontological constructs instead of re-inventing them again and again.

Turning an existing database into a useful and semantically consistent ontology is in most cases not a purely mechanical endeavour. A useful ontology cannot simply be generated through a generic syntactic conversion. A semantic and ontological *re-interpretation* is necessary. Syntactic conversion alone is not enough for realizing complex integration of different databases, since the associated semantics often do not match or are highly ambiguous. The conversion has to be informed by biomedical domain knowledge, as well as knowledge of basic ontological principles. For example, the ontology creator should invest some time in answering questions such as “Is an electrical current across a membrane an object, a process or a property of the membrane?”, “Is the relation between the ‘hippocampus’ and the ‘hippocampus proper’ an is-a relation or a part-of relation?”, or “Is ‘neurotransmitter’ a class of molecules, or a role that certain molecules can play in a certain scenario?”. On the other hand, an overly precise ontology may hamper its effective use. A major factor in the success of any ontology is the balance between solid, logically consistent and unambiguous description of entities on one hand, and pragmatic features such as intuitiveness, ease of queries, openness to change and overall simplicity on the other hand.

One outstanding issue that needs to be addressed is the agreement on stable, preferably resolvable URIs for bioinformatics resources such as protein and publication records. Unfortunately, most primary data providers have not started producing usable URIs for their resources. The URI system that is being developed by the Science Commons based on persistent uniform resource locators (PURLs [48]) may be a possible solution to this problem.

Another pressing problem that caused difficulties during the development and use of our ontologies is the lack of scalable querying and reasoning support for OWL by triplestores. This makes it much difficult to write queries and applications for OWL ontologies. The solution is the creation and standardization of new, OWL-aware triplestores and query languages. Such a solution may take a considerable amount time. Therefore, our approach to to apply simple algorithms and best-practices to make complex OWL ontologies amenable to existing, standard RDF tools and query languages. In addition, we have been collaborating with Oracle in exploring the use of Oracle 11g [49] as a proprietary OWL triplestore for storing, querying and reasoning about OWL ontologies. This academic-industrial collaboration may help contribute to the future standardization of OWL-based triplestore technologies.

Lastly, more work needs to be done on the representation of uncertainty, evidence and data provenance in OWL ontologies. These are currently addressed by several working groups, including the *W3C Semantic Web in Health Care and Life Science Interest Group* (HCLSIG [50]).

5. Conclusion and Future Work

We have demonstrated how Semantic Web technologies can be used in the context of neuroscience data integration. While other projects have adopted Semantic Web standards like RDF and OWL for local information representation, our project is among the first that actually use Semantic Web technologies to create a neuroscience semantic web that spans over different information sources hosted on different web servers and developed by independent groups. We also showed that the use of more advanced logical formalism like OWL, as well as the use of foundational ontologies, has real practical advantages. The Semantic Web has the potential to become a standard platform for semantic integration of neuroscience data.

Two future threads of development are based on the current work. First is the development of an easily accessible and intuitive web user interface to query the ontologies without needing to write verbose SPARQL queries. The development of Entrez Neuron [51], a web portal based on the ontologies presented in this paper, is one step in this direction. The second future thread of development is the exploration of strategies to make syntactically complex OWL ontologies such as NeuronDB better accessible to standard RDF tools and query languages. Furthermore, we are expanding the SenseLab ontology collection by: 1) adding mappings to other ontologies (e.g., the OBO Chemical Entities ontology) and 2) converting new databases to OWL.

The Semantic Web development in SenseLab is integral to the activities within the *Semantic Web in Health Care and Life Science Interest Group* (HCLS IG). The activities of this group span many different disciplines and are driven by participants from different sectors and countries. The existence of such a group with a strong backing in the communities of biology, medicine, computer science and philosophy is essential for the kind of large-scale information integration that is so often demanded — e.g., to realize a working infrastructure for translational medicine. The HCLS IG will continue to explore how to build a Semantic Web infrastructure for integrating biomedical data and disciplines, and to raise the awareness for Semantic Web technologies in the scientific community. The Semantic Web development in SenseLab will continue to contribute to this community activity.

Acknowledgments

This work is supported in part by: NIH grant P01 DC04732 and Fidelity Foundation, a postdoctoral fellowship from the Konrad Lorenz Institute for Evolution and Cognition Research, Austria and by the Science Foundation Ireland under Grant No. SFI/08/CE/I1380 (Lion-2). We thank the members of the W3C Health Care and Life Science Interest Group, the Science Commons/Neurocommons project and the developers of the Basic Formal Ontology for their feedback and cooperation.

References

1. Martone ME, Gupta A, Ellisman MH. E-neuroscience: challenges and triumphs in integrating distributed data from molecules to brains. *Nature Neurosci.* 2004; 7(5):467–472. [PubMed: 15114360]
2. <http://www.w3.org/>
3. <http://www.w3.org/RDF/>
4. <http://www.w3.org/TR/owl-features/>
5. Ruttenberg A, Clark T, Bug W, Samwald M, Bodenreider O, Chen H, Doherty D, Forsberg K, Gao Y, Kashyap V, Kinoshita J, Luciano J, Marshall MS, Ogbuji C, Rees J, Stephens S, Wong GT, Wu E, Zaccagnini D, Hongsermeier T, Neumann E, Herman I, Cheung KH. Advancing translational research with the Semantic Web. *BMC Bioinformatics.* 2006; 8(Suppl 3):S2. [PubMed: 17493285]
6. <http://esw.w3.org/topic/HCLS/Banff2007Demo>
7. Samwald, M.; Bug, W.; Rees, J.; Mungall, C.; Barkley, J.; Hookway, R.; Chen, H.; Stephens, S.; Bodenreider, O.; Cheung, K.; Ciccarese, P.; Clark, T.; Doherty, D.; Forsberg, K.; Kashyap, V.; Kinoshita, J.; Luciano, J.; Marshall, MS.; Neumann, E.; Prud'hommeaux, E.; Travers, M.; Wong, G.; Wu, E.; Ruttenberg, A. The Semantic Web Health Care and Life Sciences Interest Group work in progress: A large scale, OBO inspired, repository of biological knowledge based on Semantic Web technologies. Poster, Bio-Ontologies Special Interest Group Workshop at the International Conference on Intelligent Systems for Molecular Biology 2007; 2007.
8. Crasto CJ, Marenco LN, Liu N, Morse TM, Cheung KH, Lai PC, Bahl G, Masiar P, Lam HY, Lim E, Chen H, Nadkarni P, Migliore M, Miller PL, Shepherd GM. SenseLab: new developments in disseminating neuroscience information. *Brief Bioinform.* 2007; 8(3):150–62. [PubMed: 17510162]
9. Liu N, Marenco L, Miller PL. ResourceLog: an embeddable tool for dynamically monitoring the usage of web-based bioscience resources. *J Am Med Inform Assoc.* 2006; 13(4):432–7. [PubMed: 16622167]

10. Marenco L, Tosches N, Crasto C, Shepherd G, Millera PL, Nadkarni PM. Achieving evolvable Web-database bioscience applications using the EAV/CR framework: recent advances. *Journal of the American Medical Informatics Association*. 2003; 10(5):444–53. [PubMed: 12807806]
11. Smith, B. Beyond Concepts: Ontology as Reality Representation. In: Varzi, A.; Vieu, L., editors. *Proceedings of the International Conference on Formal Ontology in Information Systems*; Amsterdam. 2004. p. 319-330.
12. <http://obofoundry.org/>
13. <http://protege.stanford.edu>
14. Sirin E, Parsia B, Grau BC, Kalyanpur A, Katz Y. Pellet: A practical OWL-DL reasoner. *Web Semantics: Science, Services and Agents on the World Wide Web*. 2007; 5(2):51–53.
15. <http://www.mindswap.org/2004/SWOOP/>
16. <http://www.topbraidcomposer.com>
17. Smith B, Ceusters W, Klagges B, Köhler J, Kumar A, Lomax J, Mungall C, Neuhaus F, Rector AL, Rosse C. Relations in biomedical ontologies. *Genome biology*. 2005; 6(5):R46. [PubMed: 15892874]
18. <http://www.obofoundry.org/ro/>
19. <http://www.obofoundry.org/>
20. <http://www.ifomis.uni-saarland.de/bfo>
21. Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, Goldberg LJ, Eilbeck K, Ireland A, Mungall CJ, Leontis N, Rocca-Serra P, Ruttenberg A, Sansone S, Scheuermann RH, Shah N, Whetzel PL, Lewis S. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology*. 2007; 25:1251–1255.
22. Marenco L, Tosches N, Crasto C, Shepherd G, Miller PL, Nadkarni PM. Achieving evolvable Web-database bioscience applications using the EAV/CR framework: recent advances. *J Am Med Inform Assoc*. 2003; 10(5):444–53. [PubMed: 12807806]
23. <http://bioinformatics.ubc.ca/atlas/downloads/>
24. <http://idconverter.bioinfo.cnio.es/>
25. <http://smd.stanford.edu/cgi-bin/source/sourceBatchSearch>
26. <http://sw.neurocommons.org/2007/uri-explanation.html>
27. Bota M, Dong H, Swanson LW. Brain architecture management system. *Neuroinformatics*. 2005; 3(1):15–48. [PubMed: 15897615]
28. <http://brancusi.usc.edu/bkms/>
29. <http://ccdb.ucsd.edu/sao.html>
30. <http://ccdb.ucsd.edu/>
31. <http://fireball.drexelmed.edu/birnlex/OWLDocs/>
32. <http://www.nbirn.net/>
33. http://www.bioontology.org/wiki/index.php/CARO:Main_Page
34. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics*. 2000; 25(1):25–9. [PubMed: 10802651]
35. <http://obi.sourceforge.net/>
36. <http://pdsp.med.unc.edu/>
37. <http://www.w3.org/RDF/Validator/>
38. Sirin E, Parsia B, Grau BC, Kalyanpur A, Katz Y. Pellet: A practical OWL-DL reasoner. *Web Semantics: Science, Services and Agents on the World Wide Web*. 2007; 5(2):51–53.
39. <http://pellet.owldl.com/>
40. <http://neuroweb.med.yale.edu/senselab/>
41. <http://virtuoso.openlinksw.com/>
42. http://neuroweb.med.yale.edu/senselab/neuron_ontology.owl
43. <http://neuroweb.med.yale.edu/senselab/model-db.owl>

44. <http://neuroweb.med.yale.edu/senselab/brainpharm.owl>
45. Marshall, MS.; Prud'hommeaux, E. W3C Interest Group Note; A Prototype Knowledge Base for the Life Sciences. web publication, <http://www.w3.org/TR/hcls-kb/>
46. <http://neurocommons.org/>
47. Gangemi, A.; Guarino, N.; Masolo, C.; Oltramari, A.; Schneider, L. Sweetening ontologies with DOLCE. In: Gómez-Pérez, A.; Benjamins, VR., editors. Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management; London, UK. 2002. p. 166-181.
48. <http://purl.org>
49. <http://www.oracle.com/database/>
50. <http://www.w3.org/2001/sw/hcls/>
51. Cheung KH, Lim E, Samwald M, Chen H, Marengo L, Holford ME, Morse TM, Mutalik P, Shepherd GM, Miller PL. Approaches to neuroscience data integration. Briefings in Bioinformatics. 2009; 10(4):345–353. [PubMed: 19505888]

Figure 1.

An example of the simplified representation of neuronal structure in NeuronDB (right side) as compared to the actual morphology (left side, textbook illustration of a Purkinje neuron). In accordance with common practice in neuroscience, the neuron is seen as divided into sections such as soma, axon and dendrite. The electrical and molecular properties of each section can be described separately.

Figure 2.

Formulating generalized views of biological reality based on partially contradictory research statements in the SenseLab ontologies. The statements shown here are simplified. The SenseLab ontologies use a three-layer design pattern to organize these classes and to use OWL reasoning to create generalized views of biological reality based on research statements.

(a) Example based on the description of 'Purkinje cells' (a certain type of neuron): The first layer (shown on the left) is made up of a class that is defined in the 'ontology scaffold', e.g., the class 'Purkinje cell'. The second layer (in the middle) is made up of subclasses of the class in the first layer which have been derived from research findings. These classes were automatically exported from the SenseLab database. Research statements (e.g., "we observed a sodium current in a Purkinje cell") are interpreted as claims of the existence of a certain class with certain properties (e.g., the class "Purkinje cell with sodium current"). Of course, it is highly likely that at least some of these classes of neurons are identical in reality and that their definitions differ only from each other because the researchers investigated different properties. However, because of the open world assumption of OWL the mere definition of two classes does not imply that these two classes are distinct in reality -- they can still be declared equivalent through additional statements. This is done in the third layer (shown on the right). The third layer is only made up of a single class, which is a subclass of all the classes in the second layer. It bundles all of the properties of these classes in a single, generalized class, e.g., a generalized class of a Purkinje cell based on all available research findings. When research findings contradict each other, this generalized class can become unsatisfiable, which can be detected by a reasoner.

(b) When a contradictory research finding such as the one in (a) is added to the ontology and identified by the reasoner, the curator of the ontology may be alerted to formulate new generalized classes to re-establish consistency in the third layer. This process can be used to approach incrementally new, generalized findings about biological reality. For example, further investigation could confirm that there indeed are two distinct populations of Purkinje cells with different properties, although they were considered to be the same at first. OWL and OWL reasoning can assist in identifying these distinctions and in formulating new hypotheses.

Figure 3.

The class hierarchy of the biological functions of receptors and transmitters was represented through subclasses of the 'Function' class from BFO.

Figure 4.

Ontology import dependencies. The arrows point from the imported ontology to the importing ontology, e.g., the NeuronDB Ontology imports the Relation Ontology. Import statements are transitive, e.g., the ModelDB Ontology imports both the NeuronDB ontology and the Relation ontology. Ontologies created by SenseLab are printed in bold, all other ontologies have been created by second parties. Some imported ontologies of minor importance have been omitted. This graph demonstrated how tightly interrelated ontologies on the Semantic Web can be, even when they have been developed by independent groups and are housed on different servers.

Figure 5.

Relations ('mappings') between classes from the NeuronDB ontology (in the middle) and classes from external ontologies. The Uniform Resource Identifier (URI) for each class is shown. This example demonstrates the use of 'subclass of', 'equivalent class' and 'has part' relations between ontologies. In OWL, the relations between entities in different ontology files are formulated with the exact same syntax as relations within a single ontology. All of the URIs in this example can be resolved via HTTP to yield the ontology structures encoded in RDF/XML.

Table 1

Statistics of SenseLab ontologies. The last digit of each number has been rounded. The statistics for the 'NeuronDB (including generalisations)' ontology are omitted.

Ontology	Subject-predicate-object triples ('RDF triples')	Named classes (including imports)	Individuals (including imports)	Properties (including imports)
NeuronDB	21010	1400	1510	60
ModelDB	3720	1410	1800	60
BrainPharm	810	1710	1830	80
NeuronDB – BFO mapping	380	1500	1510	70
NeuronDB – SAO mapping	40	2160	1550	330
NeuronDB – Birnlex – BAMS – OBI mapping	130	3640	1650	190

Table 2

Examples of possible queries and the ontologies that are needed for each query.

Example query	Ontologies needed for query
Return all neuron types that are located in the Neocortex or some part of the Neocortex, and show research notes and Pubmed references for each.	NeuronDB
Return all neurons that use GABA as a neurotransmitter and that have receptors for Glutamate located on their dendrites.	NeuronDB, ModelDB
Return all neurons that might be affected in the early phase of Alzheimer's disease.	NeuronDB, BrainPharm
Return available computational models for all neurons that exhibit A-type potassium ion currents on their membranes.	NeuronDB, ModelDB
Return all ligands that bind with high affinity to neurons located in the hippocampus.	NeuronDB, PDSP Ki database
Return all neurons and their properties in regions that receive neuronal projections from a cortical brain region.	NeuronDB, BAMS