A strategy for building neuroanatomy ontologies

David Osumi-Sutherland[1,*], Simon Reeve[1], Christopher J. Mungall[2], Fabian Neuhaus[3], Alan Ruttenberg[4], Gregory S. X. E. Jefferis[5] and J. Douglas Armstrong[6]

[1]Department of Genetics, University of Cambridge, Cambridge, CB2 3EH, UK, [2]Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA, [3]National Center for Ontological Research, Department of Philosophy, University at Buffalo, Buffalo, NY 14260, USA, [4]School of Dental Medicine, University at Buffalo, Buffalo, NY 14214, USA, [5]Division of Neurobiology, MRC Laboratory of Molecular Biology, Cambridge, CB2 0QH, UK and [6]School of Informatics, University of Edinburgh, Edinburgh, EH8 9AB, UK

Associate Editor: Janet Kelso

ABSTRACT

Motivation: Advancing our understanding of how nervous systems work will require the ability to store and annotate 3D anatomical datasets, recording morphology, partonomy and connectivity at multiple levels of granularity from subcellular to gross anatomy. It will also require the ability to integrate this data with other data-types including functional, genetic and electrophysiological data. The web ontology language OWL2 provides the means to solve many of these problems. Using it, one can rigorously define and relate classes of anatomical structure using multiple criteria. The resulting classes can be used to annotate datasets recording, for example, gene expression or electrophysiology. Reasoning software can be used to automate classification and error checking and to construct and answer sophisticated combinatorial queries. But for such queries to give consistent and biologically meaningful results, it is important that both classes and the terms (relations) used to relate them are carefully defined.

Results: We formally define a set of relations for recording the spatial and connectivity relationships of neuron classes and brain regions in a broad range of species, from vertebrates to arthropods. We illustrate the utility of our approach via its application in the ontology that drives the Virtual Fly Brain web resource.

Availability and implementation: The relations we define are available from http://purl.obolibrary.org/obo/ro.owl. They are used in the *Drosophila* anatomy ontology (http://purl.obolibrary.org/obo/fbbt/2011-09-06/), which drives the web resource http://www.virtualflybrain.org

Contact: djs93@gen.cam.ac.uk

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 INTRODUCTION

The neuroanatomical literature is old and diverse. Large amounts of knowledge in this field are trapped in a literature that uses various, sometimes conflicting nomenclatures. Large new datasets are being generated using neuron tracing, image registration and segmentation (Chiang *et al*., 2010; Jefferis *et al*., 2007; Lam *et al*., 2010). To be searchable and query-able in ways that are familiar to neurobiologists, these datasets need to be annotated in a way that reflects commonly used terminology and integrates related genetic data.

Atlas-based approaches (e.g. Dang *et al*., 2007) and the use of image stack registration and image analysis are central to these goals, but are not sufficient. Many important types of neuroanatomical information, such as lineage and function, cannot be extracted from images. It is also difficult to see how image analysis can relate bulk data to the way that biologists typically communicate about neurobiology—as discourse about named classes of cells (neurons, glia) and brain structures.

An ontology-based approach using the W3C standard ontology language OWL2[1], has many advantages as a solution to these problems. An ontology provides a standard, defined vocabulary for annotating data and can therefore act as a hub for integrating disparate datasets. An ontology is also a classification and, potentially at least, a query-able store of knowledge about a particular domain. Biologists classify neuroanatomical structures (neurons in particular) in many different ways and any useful ontology needs to reflect this. But maintaining multiple classification schemes by hand in one ontology is not practical except in small ontologies (Rector, 2003).

OWL2 provides a means to store formal (logical) statements about the relationships between classes using relations (object properties in OWL) and logical quantifiers, as well as the logical characteristics of relations and entailments between them. A variety of reasoning software available for OWL2 can use this information to automate classification, check for errors and allow sophisticated, combinatorial queries. As a result, we can now build and maintain sizeable, high quality ontologies with multiple classification schemes as a way of storing knowledge and classifications in an easily query-able form.

Scaling of reasoning with fully expressive OWL2 DL cannot be made concurrent and is 'worst case intractable': reasoning is sufficiently fast for many applications, but ultimately becomes slow and then impractical as an ontology grows (Baader *et al*., 2005). But for a highly expressive subset of OWL2, OWL2 EL[2], reasoning algorithms can be made concurrent (Kazkov *et al*., 2011) and have reasoning times that scale, at worst, as a polynomial function of

---

*To whom correspondence should be addressed.

[1]http://www.w3.org/TR/owl2-overview/
[2]http://www.w3.org/TR/owl-profiles/#OWL_2_EL

ontology size (Baader *et al.*, 2005; Baader *et al.*, 2008). This means that reasoning does not become intractable as an ontology grows. For example, ELK, an EL reasoner, can classify SNOMED-CT (∼300 000 classes) ∼200 times faster than the DL reasoner FaCT++ (Dentler *et al.*, 2011; Kazakov *et al.*, 2011).

Various neuroanatomy ontologies or ontology-like structures have been built in recent years (reviewed in Larson and Martone, 2009), with varying degrees of formalization. The simplest of these, such as the one developed and used by the Allen brain atlas, are single inheritance partonomies, (Dang *et al.*, 2007). Various semi-formalized ontologies have also been developed in which relations do not, in the main, have any formal, logical definition. These include NIF (Bug *et al.*, 2008), BAMS (Bota and Swanson, 2008) and NeuronBank (Katz *et al.*, 2010), all of which underpin significant neuroinformatics resources. At the other end of the spectrum, Niggemann *et al.* (2008) have developed an impressively well axiomatized model of functional neuroanatomy using a more expressive but less computationally tractable logic than OWL. This work has not, as far as we are aware, been much applied beyond the original paper.

Various relations will be needed for a complete classification of neurons, including relations for capturing neurotransmitter, function, morphology, partonomy and connectivity. Relations for recording the function of anatomical structures have been proposed elsewhere (Meehan *et al.*, 2011). In this article, we propose a standard, integrated set of relations for capturing partonomy and connectivity. These are built using a small number of common building blocks that are sufficiently general to allow our relations to be used for both vertebrate and invertebrate neuroanatomy ontologies. To facilitate integration with bulk 3D image data, the semantics of these relations are designed to cope with basic mereological reasoning. The neural connectivity relations we propose are defined from the synapse level up, allowing logical inference between assertions about connectivity at different levels of granularity.

To achieve fast reasoning now while leaving open the possibility of more sophisticated reasoning in the future, we provide a complete formalization in Common Logic while restricting our OWL formalization, with one exception, to OWL2 EL. We also use a method that hides some details of OWL formalization from users and reasoners unless they choose to expose it.

To ensure consistency with related efforts, we have followed, and participated in the development of, emerging standards for defining relations based on the work of the OBO Foundry (Smith *et al.*, 2007) and the INCF sponsored Project for Ontologies in Neuroscience.[3]

The utility of our solution is clearly demonstrated by its implementation and use in the Virtual Fly Brain project (http://virtualflybrain.org; Milyaev *et al.*, *submitted in parallel*), from which we draw our examples.

2 METHODS

OWL provides numerous constructs for defining classes but limited capabilities for defining relations (object properties) in terms of other relations. One way around this is to use shortcut relations (Mungall *et al.*, 2010). These are OWL object properties annotated with a template that specifies expansion to an OWL class expression. A special-purpose expansion engine[4] can extend the ontology with axioms specified by the expansion.

---

[3]http://www.incf.org/core/programs/pons
[4]http://code.google.com/p/owltools/

Where possible, the semantics of the expanded form of the relation are captured using OWL property characteristics (e.g. transitivity), property hierarchy and property chains. This approach allows more complex relations to be defined in terms of simpler ones. It also allows complexity to be hidden from both users and reasoners unless it is needed, in which case it can be revealed by expansion.

Shortcut relations also have the advantage that they can be used in OBO ontologies and optionally expanded on translation to OWL. In unexpanded form, all of our proposed relations are compatible with OBO format. Code for translation between OBO and OWL with optional expansion of shortcut relations can be downloaded from http://code.google.com/p/oboformat

In some cases we cannot capture the complete meaning of a relation in OWL. For this reason we provided a supplementary document to this article in which all relations in Common Logic Interchange Format (CLIF; ISO/IEC, 2007), providing a translation to an OWL expansion wherever possible. Although CLIF cannot be reasoned over using OWL reasoners, the CLIF formalization serves to document our intentions and may in future be exploited for reasoning in either extended versions of OWL or other reasoning systems.

To keep compatibility with OWL, the relations we use are binary- and instance-level. This is in contrast to the formalism originally proposed for the OBO Relations Ontology, which specifies type-level binary relations defined using instance-level ternary relations taking a time argument (Smith 2005). Please see the supplementary CLIF file for further discussion of how the relations define here relate to those in the OBO relations ontology. Following standard mereology, the mereological relations discussed in the first section of the results (**overlaps**, **part_of**, **has_part**) are locally reflexive. For example, every anatomical structure **overlaps** itself. We reflect this in our CLIF formalization only. There are strategies for expressing this in OWL, but, with current technology, the cost in terms of reasoning speed is very high.

Our relation definitions (shortcut and CLIF) make use of classes from the Gene Ontology (Ashburner *et al.*, 2000), Cell Ontology (Meehan *et al.*, 2011) and CARO (Haendel *et al.*, 2007). All relations defined here have either BFO or RO IDs. An OWL file containing these relations can be downloaded from: http://purl.obolibrary.org/obo/ro.owl

For ease of writing and comprehension, all OWL in this article is written using Manchester syntax (Horridge *et al.*, 2006), which can generally be safely understood as English sentences. All object properties are written in bold. All annotation properties are underlined. All other Manchester syntax is italicized. All examples are taken from the *Drosophila* anatomy ontology, occasionally with some simplification for didactic purposes. With the version of the ontology used in this article can be downloaded here: http://purl.obolibrary.org/obo/fbbt/2011-09-06/ as fbbt.owl, or with relations expanded as fbbt_exp.owl. Both require the imported file FBbt_helper_relations.owl, from the same directory. For browsing and querying this ontology, we recommend the Protégé 4.1 ontology editor in combination with the FaCT++ reasoner.

3 RESULTS

3.1 Outline of approach

We follow a structural approach to defining relations for neuroanatomy, using a single primitive relation—a simple, transitive **part_of** relation. We avoid any granularity constraints regarding what classes this can relate. As a result, this relation provides the glue for an integrated set of relations that work across multiple granularities, from subcellular to gross anatomy.

We also restrict ourselves to a small set of very general terms for classes of neural structure. The classes we use are: neuron (CL:0000540), synapse (GO:0045202), pre-synaptic membrane (GO:0042734), post-synaptic membrane (GO:0045211), neuron projection (GO:0043005), neuronal cell body (GO:0043025) and
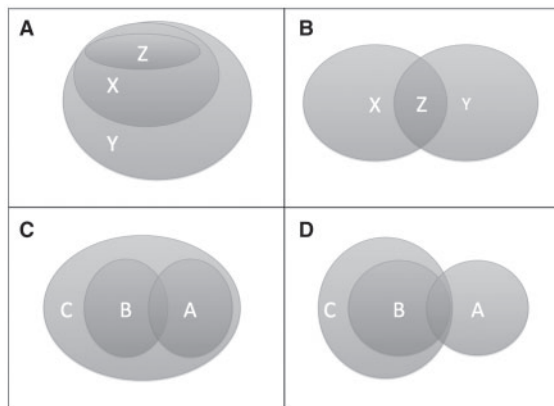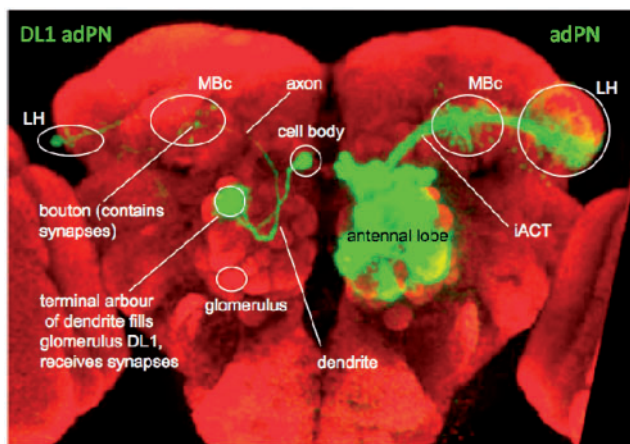
Fig. 1. Examples of overlap.

Fig. 2. Antennal lobe projection neurons. A *Drosophila* brain with labelled antennal lobe projection neurons (green) and synapses (Bruchpilot; red). On the left a single neuron, DL1 adPN has been labelled, on the right a clonally related group of 30 cells derived from a single neuroblast are members of the same superclass adPN, as the neuron on the left. Each of the labelled neurons has a dendrite that innervates a single antennal glomerulus and an axon that passes through the **iACT** to innervate the mushroom body calyx (**MBc**) and the lateral horn (**LH**).

neuron projection bundle (CARO:0001001). These terms are sufficiently general that the resulting relations are applicable to both vertebrate and invertebrate neuroanatomy. We avoid explicit references to types of neuron projection (axon, dendrite), as this distinction does not always hold for invertebrate neurons (for example, see Hasegawa *et al.*, 2011). But, following the patterns we specify here, adding additional relations that reference axons and dendrites explicitly would be straightforward.

3.2 Mereological relations

Biomedical ontologies generally record partonomy using the transitive relation **part_of** or, more rarely, its inverse **has_part** (Fig. 1A). But neither of these relations is particularly useful for recording the mereological relations between neurons and gross anatomy. Consider the neuron DL1 adPN, shown in the left panel of Figure 2. Like many neurons in both vertebrates and invertebrates, its various parts are located in different brain regions. Its soma is located in the cortex of the antennal lobe, it has a dendrite that makes

post-synaptic terminals in antennal lobe glomerulus DL1 and an axon that fasciculates with the inner antenno-cerebral tract (iACT) and has pre-synaptic terminals in the mushroom body calyx and lateral horn (Jefferis *et al.*, 2001; Yu *et al.*, 2010). The relation **overlaps** applies between this neuron and these various regions. Using standard mereology, we can define this using **part_of**:

X **overlaps** Y if and only if, there exists some Z that is **part_of** X and **part_of** Y (Fig. 1A and B).

We can capture this with an OWL property chain:[5]

has_part *o* **part_of** -> **overlaps**

It is important to note here that this relation simply asserts that the two related structures have some part(s) in common. It does not have the other common English implication of 'lying over' or 'covering'.

The **overlaps** relation defined here is a more general relation than **part_of**, it holds between X and Y when X is **part_of** Y (Fig. 1A) as well as when only some **part_of** X is **part_of** Y (see x and y in Fig. 1B.) To infer that if X **part_of** Y then X **overlaps** Y, we can make **part_of** a subproperty of **overlaps**.

We can also define rules that allow overlap to be inferred:

if X **has_part** Z and Z **part_of** Y then X **overlaps** Y (Fig. 1A)
if A **overlaps** B and B **part_of** C then A **overlaps** C (Fig. 1D and E)

As illustrated in Figure 1C and D, **overlaps** is inherited by larger parts: if A **overlaps** B and B **part_of** C, then A **overlaps** C. We can express this using property chains:

overlaps *o* **part_of** -> **overlaps**
has_part *o* **overlaps** -> **overlaps**

These relations are summarized in Supplementary Table S1. Applying this to the example in Figure 2, we *could* use **overlaps** to relate DL1 adPN to various parts of the brain. This is a useful, if rather limited assertion to make. If we know that the region of overlap has more biologically interesting properties, such as fasciculation or synapsing, it would be better to record these and to infer overlap.

3.3 Neural specific overlap relations

We are now in a position to define some neural specific subtypes of overlap. One important differentiating characteristic for many neuronal classes is the location of their cell body. We *could* record this as a nested class expression:

'DL1 adPN' *SubClassOf* **has_part** *some* ('neuronal cell body' *that* **part_of** *some* 'antennal lobe cortex')

In plain English, having a neuronal cell body that is part of some antennal lobe cortex is a necessary condition of belonging the class DL1 adPN.

But this cannot be expressed in OBO format and would make queries of the ontology quite verbose. For these reasons, we define a short-cut relation, **has_soma_location** in terms of an expansion to a class expression:

label: **has_soma_location**

---

[5]http://semwebprogramming.org/?p=175

expands_to: **has_part** *some* ('neuronal cell body' *that* **part_of** *some* Y?)

We give the short-cut relation as much of the semantics of the expanded expression as is possible in OWL. If **has_soma_location** is expanded then **overlaps** is inferred by the property chain '**has_part** *o* **part_of** -> **overlaps**'. We capture this for the unexpanded relation by making **has_soma_location** a subproperty of **overlaps**.

has_soma_location is also inherited by larger parts:

- X **has_part** *some* 'neuronal cell body' *that* **part_of** *some* Y;

- Y **part_of** some Z, therefore X **part_of** some Z;

- Therefore X **has_part** *some* 'neuronal cell body' *that* **part_of** *some* Z.

We capture this for the unexpanded relation with another property chain: '**has_soma_location** *o* **part_of** -> **has_soma_location**'.

For error checking purposes, we restrict the domain of this relation to 'neuron' (CL:0000540). The range is kept broad (anatomical structure; CARO:0000003), as neuronal cell bodies may be part of many different types of anatomical structure.

A more difficult relation to define formally is that between a neuron projection and a tract or nerve that it is **fasciculated_with**. For example, the left panel of Figure 2 shows the path of a single neuron projection (an axon) of a neuron of the class DL1 adPN. Another neuron of this class is among the many neurons visible in the right panel, all of which have axons that follow a similar path for some of their length, collectively forming the iACT. This is more than just overlap. A projection that crossed through iACT perpendicularly would overlap that tract but would not be fasciculated with it. For this to be the case, the projection must overlap with nerve or tract for some distance along its long axis. A complete formalization of this in CLIF is presented in the Supplementary Material. In OWL, **fasciculates_with** is a *SubProperyOf* **overlaps**, but is not inherited by larger parts. It has domain 'neuron *or* **part_of** *some* neuron' and range 'neuron projection bundle' (CARO).

3.4 Connectivity relations

One of the major aims of neurobiology is to map synaptic connections between neurons and between neurons and regions. To record chemical synapsing between two neurons, we again follow a simple, partonomy-based approach using GO cell component terms.

We use an expansion to define **synapsed_to** as applying between neurons N1 and N2 where N1 has a pre-synaptic membrane that is part of a synapse that has a post-synaptic membrane that is part of N2:

label: **synapsed_to**
expands_to: **has_part** *some* ('pre-synaptic membrane' *that* **part_of** *some* (synapse *that* **has_part** *some* ('post-synaptic membrane' *that* **part_of** *some* ?Y)))

We use a similar pattern to specify its inverse, **synapsed_by** (see Supplementary Table S1 for details).

In this case, we have an additional reason to prefer the shortcut approach: the class expression is sufficiently complicated that always recording it explicitly would be laborious and error prone.

With these in place, we define relations for querying neural paths: **upstream_in_neural_path_with** and **downstream_in_neural_path_with** are transitive super-properties of **synapsed_by** and **synapsed_to**, respectively. These can be used to find neurons that are upstream or downstream of some specified neuron. Further, we define a general relation, **in_neural_path_with**, as a transitive super-property of both of these relations, in order to query for neurons upstream or downstream in the same circuit (for CLIF formalizaion, see Supplementary Material).

These relations are useful when neuron to neuron connectivity data are quite limited. But they become less useful as representations of connectivity become more complete as ultimately all neurons in a nervous system are indirectly connected. A more scale-able query system would allow the number of intervening neurons between two neurons in the same circuit to be specified. This is compatible with the CLIF formalization (which uses a sequence variable). In OWL2, this can be achieved less elegantly by using property chains to define separate relations for connections with different numbers of intervening neurons.

Much available data about connectivity of neurons is neuron to region, rather than neuron to neuron. For this, we define a broad relation between a neuron and a region in which it has synaptic terminals:

label: **has_synaptic_terminal_in**
expands_to: **overlaps** *some* (synapse *that* **part_of** *some* Y?)

This is a *SubPropertyOf* **overlaps** and is inherited by larger parts.

In addition, we define two subproperties of **has_synaptic_terminal_in** that distinguish the direction of synapsing:

label: **has_pre-synaptic_terminal_in**
expands_to: **has_part** *some* ('pre-synaptic membrane' *that* **part_of** *some* (synapse *that* **part_of** *some* ?Y))
label: **has_post-synaptic_terminal_in**
expands_to: **has_part** *some* ('post-synaptic membrane' *that* **part_of** *some* (synapse *that* **part_of** *some* ?Y))

The domain for all these relations is defined as the class expression: 'neuron *or* **part_of** *some* neuron'. This allows these relations to be applied in nested class-level expressions that capture the neuron part, e.g. axon or dendrite, where synaptic terminals are located.

Applying these relations to the neuron class DL1 adPN gives:

SubClassOf: **has_post-synaptic_terminal_in** *some* 'antennal lobe glomerulus DL1'
SubClassOf: **has_pre-synaptic_terminal_in** *some* 'lateral horn'
SubClassOf: **has_pre-synaptic_terminal_in** *some* 'mushroom body calyx'

We also need a relation between a tract or nerve and a region in which neuron projections that are bundled in it have synaptic terminals. In line with common usage, we call this relation **innervates**. First we define a relation, **has_fasciculating_neuron_projection**, between a tract and a neuron projection that fasciculates with it. Using the same pattern as for **fasculates_with**: X **has_fasciculating_neuron_projection** Y if and only if X is a 'neuron projection bundle', Y is a neuron projection and some fiber segment of Y is both **part_of** and follows the path of some fiber segment of X. We then used this relation to define **innervates** as an expansion with the domain 'neuron projection bundle':

label: **innervates**
expands_to: **has_fasciculating_neuron_projection**   *some*
('neuron projection' *and* (**overlaps** *some* (synapse *that* **part_of**
some Y?)))

As shown in figure 2, neurons of the class adPN have projections
that fasciculate with the iACT and make synaptic terminals in
the antennal lobe, lateral horn and mushroom body calyx. We
can therefore record the innervation pattern of this tract with the
following restrictions:

label: iACT
SubClassOf : **innervates** *some* 'antennal lobe'
SubClassOf : **innervates** *some* 'lateral horn'
SubClassOf : **innervates** *some* 'mushroom body calyx'

It is straightforward to extend the relations defined so far to include
relations between connected regions. These are likely to be useful in
cases where neuronal tracing indicates region to region connectivity
without defining neuron classes (Bohland *et al*., 2009). Defining
some inverse relations for the relations defined so far allows us to
add property chains for reasoning these relations without expansion.
For example:

label: **directly_connected_by_neuron_to**
description: A relation that holds between two gross brain
structures that are connected by a neuron that has at least one
synaptic terminal each structure.
expands_to: **has_part** *some* (synapse *that* **overlaps** *some* (*neuron*
that **has_part** some (*synapse* that **part_of** some ?Y)))
property_chain: **has_synaptic_terminal_of** *o* **has_synaptic_
terminal_in** -> **directly_connected_by_neuron_to**

where:

has_synaptic_terminal_of   *InverseOf*   **has_synaptic_
terminal_in**

In OWL, this relation has the undesirable property that a neuron
with a synapse in a region of overlap between X and Y would satisfy
both the expansion definition and the property chain. This cannot be
fixed by making the relation irreflexive as the property chain would
then have to be removed to ensure decidability. A complete CLIF
formalization that avoids this issue is presented in the Supplementary
Material.

Relations for less direct connections could be formalized using
this pattern with the neural circuit relations defined above.

3.5   Non-structural relations

We also need to be able to capture non-structural properties
of neurons. For example, much recent work in *Drosophila* has
focused on classifying neurons by lineage (Pereanu *et al*., 2011;
Yu *et al*., 2010). To record lineage, we use the transitive relation
develops_from. In the case of our example neuron, DL1 adPN, we
can add a further restriction:

SubClassOf : **develops_from** *some* 'antero-dorsal antennal lobe
neuroblast'

3.6   Applying the relations

The relations defined in this article have been used extensively to
construct the *Drosophila* anatomy ontology and are used in queries

underlying the Virtual Fly Brain website (http://www.virtualflybrain
.org). The subset of this ontology visible on the site includes over
350 brain regions, 59 tracts and ∼500 neuron classes, over 80% of
which have relationships recording the location of their synaptic
terminals in a total of 1147 statements. It also records detailed
patterns of synapsing for a test set of 23 neurons in the optic
lobes (Meinertzhagen and O'Neil, 1991; Takemura *et al*., 2008).
All these ontology terms have textual definitions and references
attached.

3.6.1   Patterns of generalization and autoclassification   So far we
have defined our example neuron, DL1 adPN in terms of necessary
conditions for class membership using *SubClassOf*. Some of these
properties can be used to automate classification. For example, a
reasoner can conclude that DL1 adPN is a *SubClassOf* of adPN (the
class shown in the right half of Fig. 2), given:

label: DL1 adPN
SubClassOf : 'antennal lobe projection neuron'
SubClassOf : **develops_from** *some* 'antero-dorsal antennal lobe
neuroblast'
label: adPN
EquivalentTo: 'antennal   lobe   projection   neuron'   *that*
develops_from *some* 'antero-dorsal antennal lobe neuroblast'

In line with Alan Rector's normalization pattern for ontology design
(Rector, 2003), we assert as little classification as possible and
mostly state the properties of classes at leaf nodes. We then use a
reasoner to automate classification, making it practical to maintain
multiple classification schemes in a single ontology.

But the literature also contains assertions about the synapsing
and fasciculation patterns of general classes of neuron, such as
adPN. Furthermore, leaf node classes to which assertions are
attached can become general classes as detail is added to the
ontology. For example, the lineage of antennal lobe projection
neurons shown in the right panel of Figure 2, defined by
development from the 'antero-dorsal antennal lobe neuroblast', has
recently been mapped completely (Yu *et al*., 2010). This work
shows that all uniglomerular projection neurons in this lineage
(u adPN) have pre-synaptic terminals in the mushroom body
calyx.

We could capture this as a general class inclusion (GCI) axiom
in OWL, but for compatibility with OBO format, we record it as
assertions about a named class:

label: antennal lobe projection neuron u adPN
EquivalentTo: 'uniglomerular antennal lobe projection neuron'
that **part_of** *some* 'adult brain' *and* **develops_from** *some* 'antero-
dorsal antennal lobe neuroblast'
SubClassOf : **has_pre-synaptic_terminal_in** *some* 'mushroom
body calyx'

A reasoner can now classify the following term as an 'antennal
lobe projection neuron u adPN' and infer that this neuron **has_
pre-synaptic_terminal** in some mushroom body calyx:

label: DL1 adPN
EquivalentTo: 'uniglomerular antennal lobe projection neuron'
that **dendrite_innervates** *some* 'antennal lobe glomerulus DL1' *and*
develops_from *some* 'antero-dorsal AL neuroblast' *and* **part_of**
some 'adult brain'.

In all of this, judgments are required about whether evidence is sufficient to include an assertion in the ontology. It is essential to record reasons for believing assertions about properties of neuron classes and the associated references so that users can judge for themselves what to trust and so that future editors of the ontology can understand the reasons for assertions made. We record these as references attached to all terms, frequently adding comments about evidence.

3.6.2 Queries The *Drosophila* anatomy ontology can be queried using the same types of DL class expressions we have used to define relations. These may be simple queries, such as one for neurons with synaptic terminals in the antennal lobe ('**has_synaptic_terminal_in** *some* "antennal lobe"' currently finds 127 neuron classes), or compound queries, such as one for one for neurons that fasciculate with the iACT, have post-synaptic terminals in the 'antennal lobe' and pre-synaptic terminals in the 'lateral horn' (**has_post-synaptic_terminal_in** *some* 'antennal lobe' *and* **fasciculates_with** *some* iACT *and* **has_pre-synaptic_terminal_in** *some* lateral horn.) These queries require reasoning over the property hierarchy to infer relationships from, more specific subproperties and over property chains to infer, for example, that a relationship applies to the antennal lobe based on a relationship to one of its parts.

There is great potential for sophisticated DL queries generated by combining the various relations defined here into more complicated DL queries. For example, the following query, which finds eight neuron classes, combines information about circuits and the location of synapses: 'neuron *that* **has_synaptic_terminals_in** *some* "medulla layer M5" *and* **downstream_in_circuit_with** *some* "photoreceptor cell R8"'.

Because all recorded statements are backed up with references, users can easily check the veracity of the query results. A query that previously would have taken some extensive searching of the literature now takes a few seconds and the result can be quickly checked against the papers found.

3.7 Computability

The full *Drosophila* anatomy ontology provides a good test of the computability of ontologies built using our system. This ontology has 7344 classes, 12 623 SubClass axioms, 1679 Equivalent Class axioms and 1966 uses of relations defined in this article. Classification with the DL reasoner, FaCT++ (Tsarkov and Horrocks, 2006), takes ∼74 s,[6] but after classification queries are returned in about a second. With ELK (Kazakov *et al*., 2011) a parallel EL reasoner, classification time is slashed to 780 ms. With shortcut relations expanded to GCI axioms, ELK still classifies in under 2 s, whereas FaCT++ fails to complete classification within 10 min.

Another approach to improving reasoning performance is to denormalize by instantiating inferences from a pre-reasoning step and then relaxing equivalence axioms to subclassing axioms. We currently use this strategy on Virtual Fly Brain, giving us a classification time of 5513 ms (at build time) and a query response time of under a second with FaCT++.

---

[6]All reasoning tests with Mac Pro 2 × 2.8 Ghz Quad-Core Intel Xeon.

4 DISCUSSION

4.1 Advantages over a conventional database-based approach

It is possible to use a conventional relational database approach to record detailed and high quality information about neurons and their relation to gross brain anatomy and to use this to drive a web resource (for example, see Shinomiya *et al*., 2011; http://flybrain-ndb.iam.u-tokyo.ac.jp/). But an ontology-based approach has a number of advantages. First, it provides a well-defined standard for annotation, and so provides a basis for integrating multiple external datasets. Second, unlike the OWL-based approach described here, a conventional relational database approach provides no means to automate classification. Without this, maintaining the multiple inheritance classification schemes biologists typically use quickly becomes impractical (Rector, 2003). Finally, implementing basic models of mereology and neuroanatomy in OWL gives us a clear, well documented and easily extensible query system.

There is great potential for building powerful web-based query systems using the combinatorial DL queries as described in the previous section. The formalized nature of the relations used means that we can accompany such queries with precise sentences describing what the query does, making the nature of queries transparent to users. This is difficult to achieve with compound query systems developed around conventional databases. We have begun to implement such queries on the Virtual Fly Brain website.

4.2 Application to vertebrate neuroanatomy

The practicality of our approach for building ontologies of arthropod nervous systems is demonstrated by its successful implementation in the *Drosophila* anatomy ontology. But how applicable is our system to the much larger and more complicated nervous systems of vertebrates? For some systems in vertebrate neuroanatomy, classes are sufficiently well characterized that application of our system should be no more problematic than it is for invertebrate nervous systems. For example, the neuronal populations of the mammalian retina consist of ∼72 cell types, with each type having a regular spacing and pattern of stratification within the highly regular, layered structure of the retina (Masland, 2004). But this is certainly not the case for much of vertebrate neuroanatomy.

The most formalized existing system aimed at capturing connectivity in vertebrate nervous systems is that proposed by Niggemann *et al*. (2008). Their proposed system models the connectivity of groups of co-localized[7] neurons sharing common structure, function and connectivity patterns. Individual neurons and neuron parts stand in a **member-of** relation to the group. They successfully apply this to modeling of a number of major functional pathways.

Our system, based around neuron classes, does not preclude any criteria for class membership that prove useful for recording generalizations about connectivity. This includes co-localization with some group of other class members. This flexibility combined with the ease of automating classification and error checking in OWL gives ontology editors the means to build and manage large ontologies with multiple overlapping classification schemes. It is not clear how this could work with the Niggemann system without

---

[7]Co-localization here means that all members of a group have some part (soma, projection) that is localized to the same region.

extending it to include a specification of underlying neuron classes. But if it proves useful, our system could be extended to encompass some aspects of the Niggemann system. In particular, for any neuron class with a restriction specifying location, we could define a corresponding neuron group class. With appropriate formalization in OWL2, such groups could be automatically populated and related to each other based on classification of the underlying neuron classes. One possible use for such terms is to record the number of neurons in a co-localized group (perhaps as a range or average). This could be useful in recording connectivity strength between regions.

Ultimately, assessing the usefulness of our system for ontological modeling of vertebrate nervous systems will be an empirical exercise. We fully expect that further extension of our system will be needed for this task. Necessary extensions are likely to include relations specific to the grossest levels of vertebrate neuroanatomy.

4.3 Querying annotation of genetic resources

Anatomy ontologies have been used extensively to annotate gene expression and phenotype. Typically this work is done by model organism databases that use **part_of** and classification hierarchies to group related annotations. But, as we have outlined, **part_of** relations are not particularly useful for relating neurons to gross neuroanatomy. Overlap, which we can now infer with ease, is much more useful, but comes with dangers for grouping annotations. A gene product or phenotype annotated with a term for a neuron may be localized to some part of that neuron that does not overlap the region queried. Despite this danger, overlap can still be very useful for grouping annotations. For example, many commonly used transgenes in *Drosophila* (e.g. GAL4; Brand and Perrimon, 1993), make gene products that are not localized within cells. Even where grouping annotations in this way is not completely safe, it may still be worth using if the enrichment of positive results is significant compared with the number of false hits.

4.4 Image annotation

Ontology terms refer to classes—of neuron, brain region etc. In contrast, when annotating individual images or image stacks, we are referring to individual members of some class. These can be represented in OWL and related to each other using the relations we define here and typed using class expressions. As for classes, a reasoner can be used to auto-classify individuals, check their consistency with the ontology and query for them.

As part of the VFB project we are annotating 16 000 neuron images from the FlyCircuit (Chiang *et al.*, 2010) project via a mixture of hand annotation and auto-annotation using information extracted from registered image stacks. In reasoning tests with a set of 800 annotated individuals and the JCEL EL reasoner (http://jcel.sourceforge.net/) classification takes several minutes. But a denormalized ontology produced from this classification step classifies in ∼7 s with FaCT++.

By using expanded forms of the relations, there is also the potential to reason across detailed annotation of individual neuron parts, such as the electron micrographs being used to map individual synapses in the *Drosophila* brain (e.g. Sprecher *et al.*, 2011). The one limitation on this is that inverse property assertions are likely to be useful in such reasoning, but are outside of EL and so reasoning may be slow with current systems.

4.5 Limitations and possible extensions

The shortcut relations defined here have a major limitation; they can only be used with existential quantifiers. This places some limits on the types of queries that are possible. But as the alternatives, universal quantification and cardinality constraints, are outside OWL2 EL, this limitation will not affect attempts to build large ontologies with good scaling properties.

Our system has no way to formally record connectivity strength. In many cases this is plastic, but it may still be useful to record, for example, a range or average for: the number of synaptic connections between individual members of neuron classes; the number of cells of some particular class that connect to a region; the numbers of neurons connecting two regions. This can, still be usefully recorded in free text definitions of terms. Cardinality constraints, if they were allowed in our system, could be used to record at least some information on connectivity strength. But their use is known to scale badly with DL reasoners, especially when cardinality numbers are higher than single figures (Boeker *et al.*, 2011). A more promising if less rigorous approach would be to use OWL data properties to record the number of neurons in a co-localized group. These might best be attached to terms that, as in Niggeman, refer to classes of co-localized groups of neurons rather than to neuron classes themselves.

The limited mereological reasoning available in our system means that it is not suitable for reconciling multiple parcellation schemes for a single type of brain. Cross-registration and mapping to common co-ordinate systems are much better suited to this task.

5 CONCLUSIONS

As illustrated by the Virtual Fly Brain project, the relations we define here provide a practical basis for building query-able stores of neuroanatomical knowledge in the form of ontologies and collections of annotated neuroanatomical images. By using the resulting ontologies to annotate external datasets including genetic data neurophysiological data, the resulting ontologies can act as integration hubs. Restriction to OWL2 EL and the ongoing development of fast, concurrent EL reasoners means our approach will be fast for many uses and should not become intractable even with very large datasets.

Further formalized relations are needed for more complete neuroanatomy ontologies, including relations for recording electrical synapsing and neurotransmitter. We intend to publish on these in the near future. Application of our relations to modeling vertebrate neuroanatomy is likely reveal the need for various extensions to our proposed system. The relations defined here can provide a solid base to build these on.

REFERENCES

Ashburner,M. *et al*. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.

Baader,F. *et al*. (2005) Pushing the EL Envelope. In *Proceedings of 19th Int. Joint Conf. on Artificial Intelligence (IJCAI'05)*, Professional Book Center, Denver CO, pp. 364–369.

Baader,F. *et al*. (2008) Pushing the EL Envelope Further. In Clark,K. and Patel-Schneider,P.F. (eds), *OWLED 2008DC. CEUR conference Proceedings*, 496, Aachen.

Boeker,M. *et al*. (2011) A T-Box Generator for testing scalability of OWL mereotopological patterns. In Dumontier,M. and Courtot,M. (eds), *OWLED 2011. CEUR conference Proceedings*, 796, Aachen.

Bohland,J.W. *et al*. (2009) A proposal for a coordinated effort for the determination of brainwide neuroanatomical connectivity in model organisms at a mesoscopic scale. *PLoS Comput. Biol.*, **5**, e1000334.

Bota,M. and Swanson,L.W. (2008) BAMS Neuroanatomical Ontology: Design and Implementation, *Front. Neuroinform.*, **2**, 2.

Brand,A.H. and Perrimon,N. (1993) Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development*, **118**, 401–415.

Bug,W.J. *et al*. (2008) The NIFSTD and BIRNLex vocabularies: building comprehensive ontologies for neuroscience. *Neuroinformatics*, **6**, 175–194.

Chiang,A.S. *et al*. (2010) Three-dimensional reconstruction of brain-wide wiring networks in Drosophila at single-cell resolution. *Curr. Biol.*, **21**, 1–11.

Dang,C. *et al*. (2007) *The Allen Brain Atlas: Delivering Neuroscience to the Web on a Genome Wide Scale Data Integration in the Life Sciences.* Vol. 4544 of *Lecture Notes in Computer Science*, Springer, pp. 17–26.

Dentler,K. *et al*. (2011) Comparison of reasoners for large ontologies in the OWL 2 EL profile. *Semantic Web*, **2**, 71–87.

Haendel,M.A. *et al*. (2007) CARO – the common anatomy reference ontology: principles and practice. In Burger,A. *et al*. (eds) *Anatomy Ontologies for Bioinformatics*. Springer.

Hasegawa,E. *et al*. (2011) Concentric zones, cell migration and neuronal circuits in the Drosophila visual center. *Development*, **138**, 983–993.

Horridge,M. *et al*. (2006) The Manchester OWL syntax. In Cuenca Grau,B. *et al*. (eds), *OWLED 2006. CEUR Workshop Proceedings*, 216, Aachen.

ISO/IEC. (2007) ISO/IEC 24707:2007: common logic (CL)–a framework for a family of logic-based languages. ISO Standards – JTC1 Information Technology. International Organisation for Standardisation.

Jefferis,G.S. *et al*. (2001) Target neuron prespecification in the olfactory map of Drosophila. *Nature*, **414**, 204–208.

Jefferis,G.S. *et al*. (2007) Comprehensive maps of Drosophila higher olfactory centers: spatially segregated fruit and pheromone representation. *Cell*, **128**, 1187–1203.

Katz,P.S. *et al*. (2010) NeuronBank: a tool for cataloging neuronal circuitry. *Front. Syst. Neurosci.*, **4**, 9.

Kazakov,Y. *et al*. (2011) *Concurrent Classification of EL Ontologies*. In *Proceedings of the 10th International Semantic Web Conference (ISWC-11)*, *Lecture Notes in Computer Science 7032*, Springer.

Lam,S.C. *et al*. (2010) Segmentation of center brains and optic lobes in 3D confocal images of adult fruit fly brains. *Methods*, **50**, 63–69.

Larson,S.D. and Martone,M.E. (2009) Ontologies for neuroscience: what are they and what are they good for? *Front. Neurosci.*, **3**, 60–67.

Masland,R.H. (2004) Neuronal cell types. *Curr. Biol.*, **14**, R497–R500.

Meehan,T.F. *et al*. (2011) Logical development of the cell ontology. *BMC Bioinformatics*, **12**, 6.

Meinertzhagen,I.A. and O'Neil,S.D. (1991) Synaptic organization of columnar elements in the lamina of the wild type in Drosophila melanogaster. *J. Comp. Neurol.*, **305**, 232–263.

Mungall,C.J. *et al*. (2010) Taking shortcuts with OWL using safe macros. *Nature Precedings*.

Niggemann,J.M. *et al*. (2008) Modeling functional neuroanatomy for an anatomy information system. *J. Am. Med. Inform. Assoc.*, **15**, 671–678.

Pereanu,W. *et al*. (2011) Lineage-based analysis of the development of the central complex of the Drosophila brain. *J. Comp. Neurol.*, **519**, 661–689.

Rector,A. (2003) Modularisation of domain ontologies implemented in description logics and related formalisms including OWL. In *Proceedings of the 2nd international conference on Knowledge capture.*, ACM, Sanibel Island, FL, USA, pp. 121–129.

Shinomiya,K. *et al*. (2011) Flybrain neuron database: a comprehensive database system of the Drosophila brain neurons. *J. Comp. Neurol.*, **519**, 807–833.

Smith,B. *et al*. (2005) Relations in biomedical ontologies. *Genome Biol.*, **6**, R46.

Smith,B. *et al*. (2007) The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat. Biotechnol.*, **25**, 1251–1255.

Sprecher,S.G. *et al*. (2011) The Drosophila larval visual system: high-resolution analysis of a simple visual neuropil. *Dev. Biol*, **358**, 33–43.

Takemura,S.Y. *et al*. (2008) Synaptic circuits of the Drosophila optic lobe: the input terminals to the medulla. *J. Comp. Neurol.*, **509**, 493–513.

Tsarkov,D. and Horrocks,I. (2006) *FaCT++ Description Logic Reasoner: System Description. Proceedings of the Int. Joint Conf. on Automated Reasoning (IJCAR 2006). Lecture Notes in Artificial Intelligence 4130*, Springer, pp 292-297.

Yu,H.H. *et al*. (2010) A complete developmental sequence of a Drosophila neuronal lineage as revealed by twin-spot MARCM. *PLoS Biol.*, **8**, e1000461.