ICBO: International Conference on Biomedical Ontology
July 28-30, 2011 · Buffalo, NY, USA

Ontology-Based Analysis of Event-Related Potentials

Gwen Frishkoff [1,2], Robert Frank[2], Paea LePendu[3]

[1]Georgia State University, Atlanta, GA, USA
[2]University of Oregon, Eugene, OR, USA
[3]Stanford University, Stanford, CA, USA

Abstract. We describe recent progress in the development and application of NEMO (Neural ElectroMagnetic Ontology), a formal ontology for the event-related potentials (ERP) domain. The ontology encodes knowledge about patterns that are commonly seen in ERP studies. The patterns are defined using equivalent class descriptions, which specify the spatial, temporal, and functional constraints that must be satisfied for an ERP instance, or datum, to belong to a particular pattern class. The data themselves are represented in RDF, using N-triples that link the data to the ontology. Our analysis pipeline automatically generates these RDF data. We then apply a reasoner, such as Hermit, to classify the data. By creating this pipeline, we have enabled our consortium partners to compare results across experiment paradigms using a common knowledge base and to refine that base (i.e., to add or adjust pattern descriptions) based on cross-lab study results. We discuss implications for ERP meta-anlaysis, discovery of new knowledge, and resolution of current controversies in the ERP literature.

1 Introduction

This paper describes recent progress in the development and application of NEMO (Neural ElectroMagnetic Ontology), a biomedical ontology for the event-related potentials (ERP) domain. The driving motivation for NEMO is the need to make valid comparisons across ERP datasets. Although ERPs have been used in human neuroscience for over 50 years, there have been remarkably few meta-analyses, and the few that exist are of questionable validity [1]. By contrast, although functional magnetic resonance imaging (fMRI) is a newer technique, meta-analyses are now routine in the fMRI literature [2].

One problem that hinders meta-analysis in the ERP domain is the lack of a standard vocabulary and the absence of explicit, formal definitions for patterns that are commonly seen in a particular experimental context (e.g., visual word recognition). A second challenge is the complexity of ERP data, which has led to a variety of approaches to ERP pattern extraction. These cross-lab differences may result in incommensurable data, which cannot serve as inputs to a valid meta-analysis. The goal of the NEMO project is to address these problems by developing a seamless pipeline for ERP analysis, statistical measure generation, and classification of data, which can be used across studies and across labs.

In previous work [3-5], we described the structure of the NEMO ontology, which represents knowledge about ERPs and foundational concepts from various domains. The present paper describes how the ontology can function as a tool for classification and labeling of ERP data. Two recent developments have been central to this effort. First, we have added equivalent class descriptions (aka “rules”) for ERP pattern classes within the ontology.

Second, the instance-level data themselves are now automatically created as part of our ERP analysis pipeline. The pipeline takes raw ERP data as input, extracts pattern instances, and produces summary metrics for each pattern. The metrics are then used to generate RDF/OWL files (henceforth, “RDF data”), which contain a few basic assertions about each pattern and link them to the ontology.

As a result of formally encoding the ERP pattern classes, class descriptions (rules), and instance-level data, we can now classify real ERP data using a reasoner such as HermiT [6]. This is a major milestone for the NEMO project.

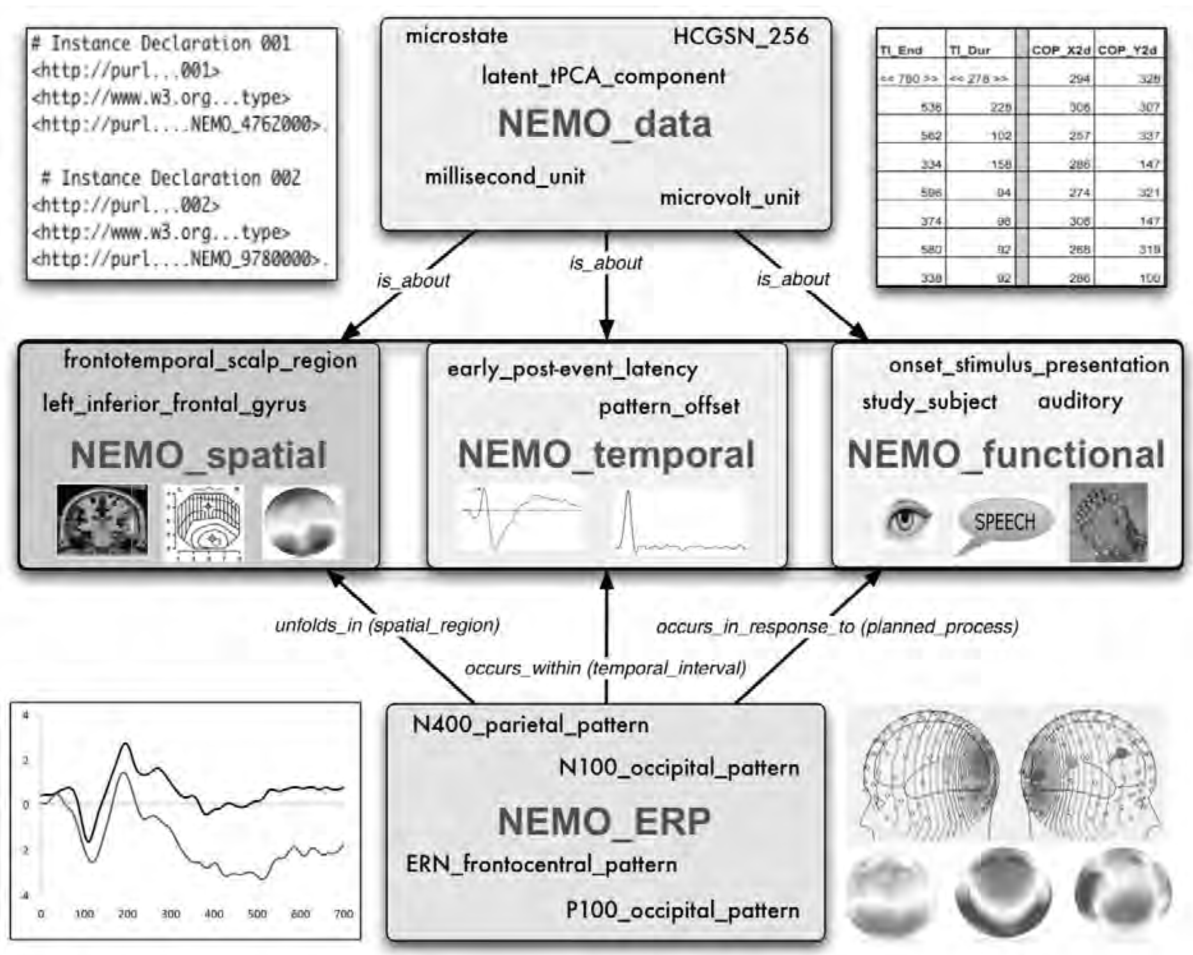

Figure 1. Core subdomains of the NEMO ontology.

Following, we outline these new developments and discuss how they can lead to scientific breakthroughs in the ERP domain. We see the development of ERP pattern classes and class descriptions (rules) as an ongoing project: researchers will ideally generate new results using the NEMO ERP analysis pipeline and refine the knowledge base to reflect new findings. We therefore emphasize the importance of considering both top-down (knowledge-driven) and bottom-up (data-driven) methods in ontology development.

2 NEMO Ontology

ERPs are measures of brain electrophysiology ("brainwaves"). ERPs provide a powerful means for studying brain function, because they are acquired noninvasively and can therefore be used in a variety of populations (e.g., children and patients, as well as healthy adults). In addition, they provide detailed information about the time dynamics, as well as the spatial distribution, of neural activity during various cognitive and behavioral tasks.

The NEMO ontology is a domain-specific knowledge base that is built on top of the Basic Formal Ontology (BFO) [8]. As described in previous work [1, 3-5], NEMO has been designed in general to comply with OBO Foundry best practices [7]. For example, we make every effort to re-use existing ontologies. To this end, NEMO imports concepts from other ontologies, including the Ontology for Biomedical Investigations (OBI; [0]), Neuroscience Information Framework (NIF; [10-11]), the Foundational Model of Anatomy (FMA; [12]), and the Cognitive Paradigm Ontology (cogPO; [13]).

NEMO includes five core domains (see Fig.

1). The NEMO_spatial domain includes concepts representing spatial regions (e.g., brain and scalp locations) and qualities (e.g., dorsal/ventral), and anatomical entities that correspond to the locations of interest (e.g., brain, scalp, skull). NEMO_temporal comprises temporal intervals (e.g., time periods referenced to ERP experiment events, which are critical for analysis) and temporal qualities, as well as some physiological concepts. NEMO_functional includes concepts related to cognitive and behavioral processes and paradigms that are relevant during experimentation. Finally, NEMO_data includes concepts related to measurement and analysis of data (e.g., "peak latency," "mean amplitude"). These five domains are separated only in theory. In practice, all classes and all relations are encoded in a single file.

In earlier versions we maintained separate files for each domain. However, a practical issue emerged as we started to define class restrictions: definitions often reference concepts from multiple domains. For example, NEMO defines an ERP as a type of process (NEMO_temporal), which unfolds in some spatial region (NEMO_spatial). In order to represent this information, it was therefore necessary to re-assert concepts in multiple files, to import these files (which caused performance problems), or to create bridge files [10] that would require additional work to maintain.

The latest release of the NEMO ontology can be browsed and downloaded from the BioPortal website.
(http://bioportal.bioontology.org/ontologies)

All versions, including the most recent ("working") and older (legacy) versions can be accessed from our SVN repository.
(http://purl.bioontology.org/NEMO)

3 ERP Pattern Rule Representation

An important goal for the NEMO ontology is to represent formally the spatiotemporal ERP patterns that have been frequently studied in cognitive and language-related ERP experiments over the past several decades. To this end, we have coded ~40 ERP pattern classes in the current version of NEMO (v. 1.60). Most recently, we have added equivalent class descriptions for each of these pattern rules. Each pattern rule specifies three sets of criteria (see Figure 2):

(1) ***Temporal***. The peak latency of a particular pattern falls within a certain time range (in milliseconds).

(2) ***Spatial.*** A pattern is characterized by surface-positive and negative voltages (in microvolts), which are distributed over certain scalp regions-of-interest.

(3) ***Functional.*** A particular pattern occurs within a certain experimental context, in response to specific types of experimental stimuli, response and task requirements.

```
visual_occipital_P100_pattern EquivalentTo scalp_recorded_ERP_extracted_pattern
        (1) and (has_proper_part some (peak_latency_measurement_datum
                  that (has_numeric_value some (decimal[>= "70"] and decimal[<= "140"]))))
        (2) that ((has_proper_part some (intensity_measurement_datum
                  that (is_quality_measurement_of some (intensity
                            that (inheres_in some (scalp_recorded_ERP
                                      that (unfolds_in some occipital_scalp_surface_region)))))
                  and (has_numeric_value some decimal[>= ".4"^^decimal])))
        (3) and (proper_part_of some (averaged_EEG_data_set
                  that (is_about some (scalp_recorded_ERP
                            that (occurs_in_response_to some (onset_stimulus_presentation
                                      that (has_object some (object
                                                that (has_quality some visual)
                                                and (has_role some stimulus_role))))))))))
```

Figure 2. Example of an ERP pattern rule.
Part (1) of this assertion expresses the ***temporal criterion*** for the visual_occipital_P100_pattern.
Part (2) expressed the ***spatial criterion***.
Part (3) expresses the ***functional (experimental) criterion***.

Figure 2 illustrates how these three criteria are used to express the pattern rule as an equivalent class description for the “visual occipital P100,” a well-known pattern in ERP research on visual perception [3-5].

4 ERP Data Representation

The NEMO ERP Analysis Toolkit is a suite of tools that provides an pipeline for ERP analysis, statistical measure generation, and creation of instance-level data that are linked to the NEMO ontology. The Toolkit uses a MATLAB class-based architecture, which allows for re-use of common objects, such as data provenance, which are referenced at every stage of the processing pipeline.

The analysis pipeline itself includes the three main steps (1-3 in Figure 3, below): (1) Step 1, ERP pattern extraction, (2) Step 2, ERP metric extraction, and (3) Step 3, RDF code generation. After initializing the script for pattern extraction (Step 1), the rest of the process is entirely automated.

Step 1: ERP Pattern Extraction. ERP pattern analysis is the process of transforming complex spatiotemporal ERP data into discrete patterns, which are used for analysis of experimental (condition) effects on the latency, amplitude, and topographic distribution of neural activity. The NEMO toolkit includes two types of pattern analysis: Decomposition, which includes various implementations of Principal Components Analysis, or PCA, and Independent Components Analysis or ICA and Windowing, or Segmentation. In contrast with conventional methods for ERP component analysis, all of the methods in NEMO are data-driven (See Ref. [1] for details). As a result, the extraction (and subsequent definition) of a particular ERP pattern is not subject to experimenter bias. Further, data can be batch-processed for efficiency.

Step 2: ERP Metric Extraction. The ERP patterns that are extracted in Step 1 are input to the ERP Metric Extraction tool, which computes summary measures of time course (e.g., peak latency, duration) and scalp distribution (e.g., average intensity over each scalp region of interest) for each of the patterns in a particular dataset (See Ref. [1] for details).

Step 3: RDF Data Generation. Finally, the latest version of the NEMO toolkit (v. 1.18) automatically writes out the results of the metric extraction script to RDF. RDF generation is new to this project. Therefore, this process is detailed in the following section.

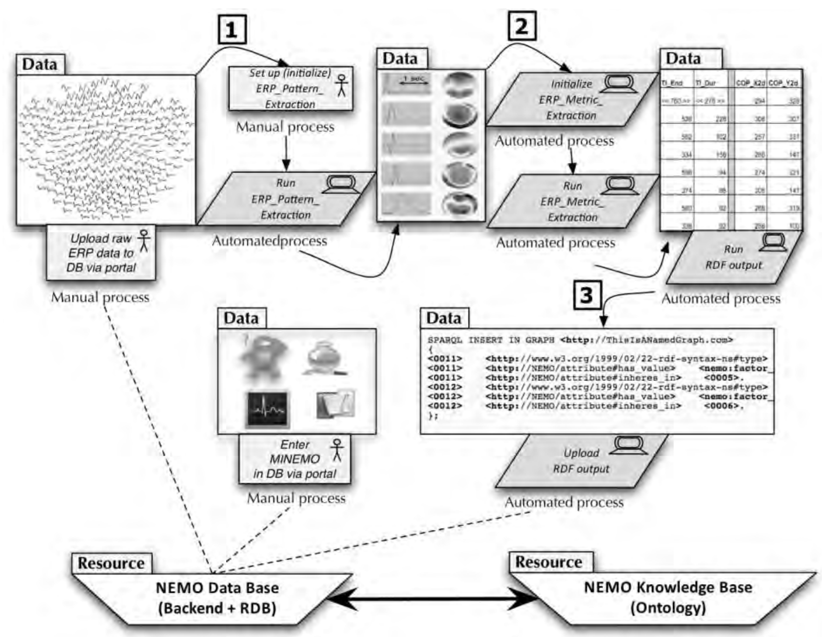

Figure 3. ERP Data processing pipeline.

The NEMO ERP Metric Extraction (Step 2 in the processing pipeline; see Fig. 2) yields a set of spatial and temporal metrics that capture the main features of ERP pattern instances. These metrics are subsequently used to classify instances, using NEMO ERP ontology rules (see following section for details on these rules).

In order to represent ERP data within the NEMO ontology, we have created a MATLAB script that writes out the summary ERP metrics to an RDF (Resource Description Framework) file. The MATLAB RDF generation treates all input/output files as distinct resources, each with a Uniform Resource Identifier reference (URIref). Similarly, each ERP data file, its attributes, the elements of its provenance, the parameters governing its transformation, the transformed data, and the file contents (ERP summary metrics) are assigned URIs that are linked to the NEMO ontology.

The MATLAB RDF runtime script assigns a set of RDF triples, or descriptive statements, to each resource. A triple consists of a *subject–predicate–object* structure, in which the predicate specifies a binary relationship between the subject and object, for example, *value001 – is_a – mean_intensity_LFRONT*. All subjects, predicates and objects (except for the typed literals) are RDF resources, which are indexed by NEMO concept URIs. Thus, the RDF generation effectively "annotates" ERP data using a small set of concepts from the NEMO ontology.

RDF triples represent the minimal information that is needed to link the data to the ontology. For example, the class *mean_intensity_LFRONT* represents the average intensity over left frontal electrodes, a concrete and uncontroversial concept. Our goal was to generate data representations that are likely to be stable and uncontroversial, and are therefore unlikely to change over time. These data-related classes are linked to other parts of the NEMO ontology through a chain of assertions that are more complicated and abstract, as shown in Figure 4. Note that this more complex assertion is not part of the RDF representation. As a result, changes in scientific knowledge should not require that we re-annotate existing data. Rather, the RDF representation of the data can simply be reclassified using a new version of the ontology.

```
mean_intensity_LFRONT EquivalentTo intensity_measurement_datum
        that (is_quality_measurement_of some (intensity
              that (inheres_in some (scalp_recorded_ERP
                    that (unfolds_in some (left_frontocentral_scalp_surface_region))))))
```

Figure 4. Class restriction for mean_intensity_LFRONT (an ERP metric class).

5 Classifying and Labeling ERP Data

After a data set has been fully processed using the NEMO ERP Analysis Toolkit, the resulting data (RDF file) can be opened and processed in Protégé [15]. The first several lines of the RDF file import the NEMO ontology. Thus, the data and ontology are both available within the file. The data can then be classified using a reasoner such as HermiT [6].

Figure 5 illustrates the classification results for one such analysis. The instance-level datum ('NW_ERP_0352') is an ERP pattern that has a pronounced surface-negativity at around 352 ms and occurs in response to a visually presneted nonword stimulus. Based on the spatial, temporal, and functional properties of this pattern, it was classified as a member of the medial_frontal_negativity (MFN).

The example in Figure 5 also illustrates an interesting scenario, which is likely to appear rather frequently in real applications: If distinct ERP pattern classes have overlapping spatial and temporal criteria, then a particular ERP observation can be classified as a member of more than one class.

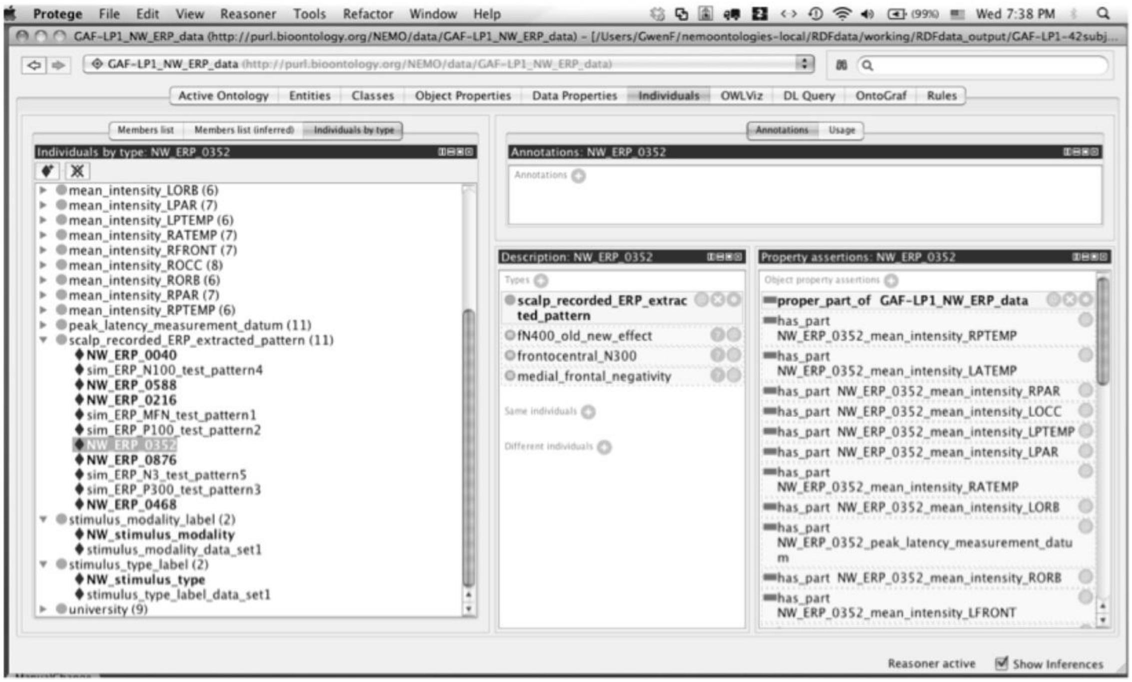

Figure 5. Classification results for one pattern (ERP response to a nonword) from a real ERP dataset.

Conversely, a pattern instance that satisfies none of the pattern rules will be classified as a member of the *undefined_ERP_pattern* class, which is defined in NEMO as the complement of all other (defined) ERP pattern classes. In each case, the classification results may challenge current definitions and, in doing so, raise a central issue for future applications: how to manage changes in the ontology over time.

To address this issue, we must first acknowlege that the most interesting parts of the ontology – that is, the pattern rules – are uncertain by their very nature. To capture this uncertainty, NEMO makes use of evidence codes, a type of annotation that has been used in GO [14] to flag the source of evidence for existence of a particular class or class definition. In NEMO, "author assertion" is considered the weakest source of evidence for a particular ERP pattern rule. The strongest evidence is a published set of results from a quantitative meta-analysis — evidence that will come with the application of NEMO tools to multiple datasets from our cross-laboratory ERP consortium.

6 Discussion

In conclusion, we have described a novel application of NEMO (Neural ElectroMagnetic Ontology), a formal ontology for the event-related potentials (ERP) domain. The ontology encodes knowledge about patterns that are commonly seen in ERP studies. The patterns are defined using equivalent class descriptions, which specify the spatial, temporal, and functional constraints that must be satisfied for an ERP instance, or datum, to belong to a particular pattern class. We have thereby attempted to capture ERP domain knowledge in a formal, explicit way. Naturally, this knowledge will evolve over time. Hence, it will be important to track and document the evidence that supports a particular pattern description and to curate this information over time.

Our hope is that this approach can help to resolve some long-standing controversies in the ERP literature. For example, the "N400" pattern has been described in more than 400 published papers, but it remains a point of controversy whether this pattern reflects automatic (e.g., unconscious) activation of word meanings or whether it is only seen in response to effortful processing of semantic information

[16]. Informally, it has been characterized as a surface-negative pattern peaking at around 400 ms over centroparietal sites [17]. However, its precise measurement and quantification vary widely, even across studies within the same research lab. This variability has made it hard even to achieve informal generalizations across ERP study results. To illustrate, Dombrowski and Heil [18] recently stated that "the interpretation of the N400 is far from being resolved." This state-of-affairs is somewhat surprising, given 30+ years of research and several hundred publications focused on N400 semantic effects. However, the reason for this state-of-affairs is evident: there are inconsistent definitions of core concepts, such as the "N400," in ERP research.

The ambiguity of natural-language definitions in ERP research has important implications for ERP research. In particular, it suggests that data mining from text may give unreliable results, since natural-language terms are used inconsistently and thus cannot be assumed to pick out the same real-world entitites. This implies, in turn, that the inputs to ERP data mining and cross-laboratory analysis should ideally consist of either structured or semi-structured ERP data, rather than natural-language descriptions of these data. To this end, we have created a unique workflow for ERP analysis, which has several key features. First, it seamlessly combines ERP analysis, metric extraction, and RDF file generation. Thus, it fills an important gap in available tools for ERP research. Second, the workflow is fully automated, which removes the need for manual selection of spatial and temporal variables. Thus, the results of ERP analysis and metric extraction (which are inputs to pattern classification) are guaranteed to be compatible with the ontology. Third, the variables themselves are noncontroversial (measures of onset, offset, and peak latency and distribution of positive and negative potentials over different regions of the scalp). Fourth, the set of variables is extensible (for example, we are adding spectral measures to the next release), so the system can support users who wish to do something novel. Likewise, the Toolkit allows flexibility in selection of pattern extraction methods, so users are not bound to one approach. In this sense, our system is not complete, but this is true by design: we fully expect that methods for ERP analysis will continue to evolve. Finally, the ontology and ontology-based resources for NEMO were developed in collaboration with an international group of ERP researchers, who represent different approaches and different kinds of experience with ERPs.

Our next step is to apply the NEMO ontology-based workflow to data from a variety of studies from across our 8 consortium sites. We are focusing on three related paradigms that have generally been studied in isolation from one another: (1) word and nonword recognition, (2) semantic priming, and (3) episodic memory for familiar and newly learned words. These three paradigms all evoke surface-negativities that have been related to semantic memory. Our hope is to discover similiarities and differences in the brain's response to semantic memory in these different experimental contexts. This work has strong significance for reading and language development and interventions for clinical conditions, such as dyslexia and language deficits due to traumatic brain injury and stroke.

Finally, as in prior work, we emphasize the importance of both top-down (knowledge-driven) and bottom-up (data-driven) methods in ontology development [1, 3-5]. Previously, we have suggested this approach could lead to robust descriptions of ERP pattern classes. Our current approach is consistent this top-down/bottom-up framework: whereas the initial ("seed") versions of the ERP pattern rules are based on published literature (top-down), our approach to ERP pattern analysis is data-driven (bottom-up). The challenge is what to do when classification results suggest inconsistencies or gaps in the ontology. This question is likely to be a central topic of ongoing and future research in biomedical ontologies.

Acknowledgements

This work is supported by a grant from the National Institutes of Health, award #R01EB007684. We would also like to thank Timothy Redmond (Stanford Center for Biomedical Informatics) for helpful suggestions on coding RDF files and combining the instance-level information with the ontology.

References

1. Frishkoff G, Frank R, Rong J, Dou D, Dien J, Halderman L. A framework to support automated classification and labeling of brain electromagnetic patterns. *Comput Intell Neurosci* 2007:14567.
2. Laird, A. R., Eickhoff, S. B., Kurth, F., Fox, P. M., Uecker, A. M., Turner, J. A., et al. (2009). ALE Meta-Analysis Workflows Via the Brainmap Database: Progress Towards A Probabilistic Functional Brain Atlas. *Front Neuroinformatics*, 3, 23.
3. Dou D, Frishkoff G, Rong J, Frank R, Malony A, Tucker, D. Development of NeuroElectro-Magnetic Ontologies (NEMO): A Framework for Mining Brainwave Ontologies. Proceedings of the *13th International Conference on Knowledge Discovery & Data Mining (KDD'07)* 2007:270-279.
4. Frishkoff, G. A., LePendu, P., Frank, R. M., Liu, H., & Dou, D. (2009). Development of Neural Electromagnetic Ontologies (NEMO): Ontology-based Tools for Representation and Integration of Event-related Brain Potentials. Paper presented at the *International Conference on Biomedical Ontologies (ICBO'09)*, Buffalo, NY.
5. Frishkoff, G., Sydes, J., Mueller, K., Frank, R., Curran, T., Connolly, J. F., Kilborn, K., Molfese, D., Perfetti, C. A. and Malony, A. D.: Minimal information for neural electromagnetic ontologies (minemo): A standards-compliant method for analysis and integration of event-related potentials (erp) data. In: *Standards in Genomic Sciences,* (2011, under review).
6. Shearer, R., Motik, B., & Horrocks, I. (2008). HermiT: A Highly-Efficient OWL Reasoner. Paper presented at the *OWL: Experiences and Directions (OWLED 2008).*
7. Simon, J., Dos Santos, M., Fielding, J. and Smith, B.: Formal ontology for natural language processing and the integration of biomedical databases. In: *Int J Med Inform*, vol. 75(3-4), pp. 224-231 (2006).
8. Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L. J., Eilbeck, K., Ireland, A., Mungall, C. J., Leontis, N., Rocca-Serra, P., Ruttenberg, A., Sansone, S. A., Scheuermann, R. H., Shah, N., Whetzel, P. L. and Lewis, S.: The OBO foundry: Coordinated evolution of ontologies to support biomedical data integration. In: *Nat Biotechnol*, vol. 25(11), pp. 1251-1255 (2007).
9. Brinkman, R. R., Courtot, M., Derom, D., Fostel, J. M., He, Y., Lord, P., Malone, J., Parkinson, H., Peters, B., Rocca-Serra, P., Ruttenberg, A., Sansone, S. A., Soldatova, L. N., Stoeckert, C. J., Jr., Turner, J. A. and Zheng, J.: Modeling biomedical experimental processes with obi. In: *J Biomed Semantics*, vol. 1 Suppl 1(2011, in press).
10. Bug, W. J., Ascoli, G. A., Grethe, J. S., Gupta, A., Fennema-Notestine, C., Laird, A. R., Larson, S. D., Rubin, D., Shepherd, G. M., Turner, J. A. and Martone, M. E.: The nifstd and birnlex vocabularies: Building comprehensive ontologies for neuroscience. In: *Neuroinformatics*, vol. 6(3), pp. 175-194 (2008).
11. Bug, W. J., Astahkov, V., Boline, J., Fennema-Notestine, C., Grethe, J., Gupta, A., Kennedy, D. N., Rubin, D. L., Sanders, B., Turner, J. and Martone, M.: Data federation in the biomedical informatics research network: Tools for semantic annotation and query of distributed multiscale brain data. In: *AMIA Annu Symp Proc*, vol. pp. 1220 (2008).
12. Martin, R. F., Rickard, K., Mejino, J. L., Jr., Agoncillo, A. V., Brinkley, J. F. and Rosse, C.: The evolving neuroanatomical component of the foundational model of anatomy. In: *AMIA Annu Symp Proc*, vol. pp. 927 (2003).
13. Turner, J. A. and Laird, A. R.: The cognitive paradigm ontology: Design and application. In: *Neuroinformatics*, (2011, in press).
14. Dwight, S. S., Harris, M. A., Dolinksy, K., Ball, C. A., Binkley, G., Christie, K. R., et al. (2002). Saccharomyces Genome Database (SGD) provides secondary gene annotation using the Gene Ontology (GO). *Nucleic Acids Res,* 30(1), 69-72.
15. Rubin, D. L., Noy, N. F., & Musen, M. A. (2007). Protégé: a tool for managing and using terminology in radiology applications. *J Digit Imaging*, 20 Suppl 1, 34-46.
16. Mari-Beffa, P., Catena, A., Valdes, B., Cullen, D. and Houghton, G.: N400, the reference electrode, and the semantic activation in prime-task experiments: A reply to dombrowski and heil (2006). In: *Brain Res*, vol. 1147(pp. 209-212 (2007).
17. Kutas, M. and Federmeier, K. D.: Electrophysiology reveals semantic memory use in language comprehension. In: *Trends Cogn Sci*, vol. 4(12), pp. 463-470 (2000).
18. Dombrowski, J. H. and Heil, M.: Semantic activation, letter search and n400: A reply to mari-beffa, valdes, cullen, catena and houghton (2005). In: *Brain Res*, vol. 1073-1074(pp. 440-443 (2006).