University of West Bohemia
Department of Computer Science and Engineering
Univerzitni 8
30614 Pilsen
Czech Republic

Semantic Web and Classic Data Structures

The State of the Art and the Concept of Ph.D. Thesis

Petr Brůha

Technical Report No. DCSE/TR-2013-03
July, 2013

Distribution: public

Technical Report No. DCSE/TR-2013-03
July 2013

Semantic Web and Classic Data Structures

Petr Brůha

Abstract

The aim of this work is to explore the data structures, data standards, terminologies, ontologies, and existing Semantic Web solutions in electrophysiology and neurophysiology community at the present time. The last part is devoted to the design of the structure and objectives of the dissertation thesis.

This work was supported by the UWB grant SGS-2013-039 Methods and Applications of Bio-and Medical Informatics.

Copies of this report are available on
http://www.kiv.zcu.cz/publications/
or by surface mail on request sent to the following address:

University of West Bohemia in Pilsen
Department of Computer Science and Engineering
Univerzitni 8
30614 Pilsen
Czech Republic

Contents

1 Introduction

A large amount of not described data, various data formats, standards and non-standardized domain descriptions lead to incompatible results and interpretations of electrophysiological experiments and to difficult communication between laboratories and research groups. A recent trend is to annotate data and create domain ontologies to share them within web-based systems. Laboratories and research groups need to follow best ontology practices, data formats, use existing ontologies and terminologies. Storing of well described data can increase the lifetime of acquired data and domain ontologies shared within web-based systems and allow better cooperation between laboratories and research groups.

Our research group, a member of the Czech INCF[1] National Node, specializes in research in electroencephalography (EEG) and event-related potentials (ERP). The group has developed the EEG/ERP portal (containing over 200 experiments) that allows groups of researchers to store, update and download data and metadata from EEG/ERP experiments. Our group collaborates with Electrophysiology ontology workgroup (EPhys). This workgroup is a newly formed group which aims is to define the practical and technical details for the development of an ontology for annotating electrophysiological data. I am a member of this group as a core ontology developer. The current primary objective of our research group is to define the ontologies to represent our EEG/ERP experiment setup.

In this work, Section 2 introduces semantic web technologies, projects and a state of the art of ontologies. Section 3 describes electrophysiological and neurophysiologic communities. It also provides details of our research group and EPhys ontology workgroup. Current ontologies, possible improvements for new ontologies and best ontology practices are described in Section 4. Current terminologies and data formats are described in Sections 5 and 6. The tools used for sharing ontologies are described in Section 7. The conclusion and the aims of Ph.D. thesis are given in Section 8.

[1] International Neuroinformatics Coordinating Facility

2 Semantic web

The Semantic Web enables computers to do more useful work and to develop systems that can support trusted interactions over the network. The Semantic Web refers to W3Cs vision of the Web of linked data. Semantic Web technologies enable people to create data stores on the Web, build vocabularies, and write rules for handling data [1].

2.1 Linked Data

The primary hypothesis is to have a huge amount of data on the Web available in a standard format, reachable and manageable by Semantic Web tools. Furthermore, not only does the Semantic Web need access to data, but relationships among data should also be made available to create a Web of Data. This collection of interrelated datasets on the Web can also be referred to as Linked Data. A common format (RDF[2]) is used to create Linked Data, to access to existing databases (relational, XML, HTML, etc). The next step is to be able to setup query endpoints to access that data more conveniently. W3C provides a palette of technologies (RDF, GRDDL[3], POWDER[4], RDFa[5], the upcoming R2RML[6], RIF[7], SPARQL[8]) to get access to the data [1].

2.2 Vocabularies

The vocabularies define the concepts and relationships used to describe and represent an area of concern. Vocabularies are used to classify the terms that can be used in a particular application, characterize possible relationships, and define possible constraints on using those terms. There is no clear division between what is referred to as "vocabularies" and "ontologies". The ontology is used for more complex and formal collection of terms. Vocabularies are the basic building blocks for inference techniques on the Semantic Web [1].

[2]RDF Resource Description Framework
[3]GRDDL - Gleaning Resource Descriptions from Dialects of Languages
[4]POWDER - The Protocol for Web Description Resources
[5]RDFa - Resource Description Framework in attributes
[6]R2RML - RDB to RDF Mapping Language
[7]RIF - Rule Interchange Format
[8]SPARQL - SPARQL Protocol and RDF Query Language (Recursive acronym)

2.3 Metadata

One of the challenges of using data is understanding how the various data were measured or obtained. This information (the data about data) is often called metadata and represents knowledge about how the data can be used [2].

2.4 W3C semantic web technology stack

Fig. 1 shows the W3C semantic web technology stack illustrating the architecture of the Semantic Web[2].

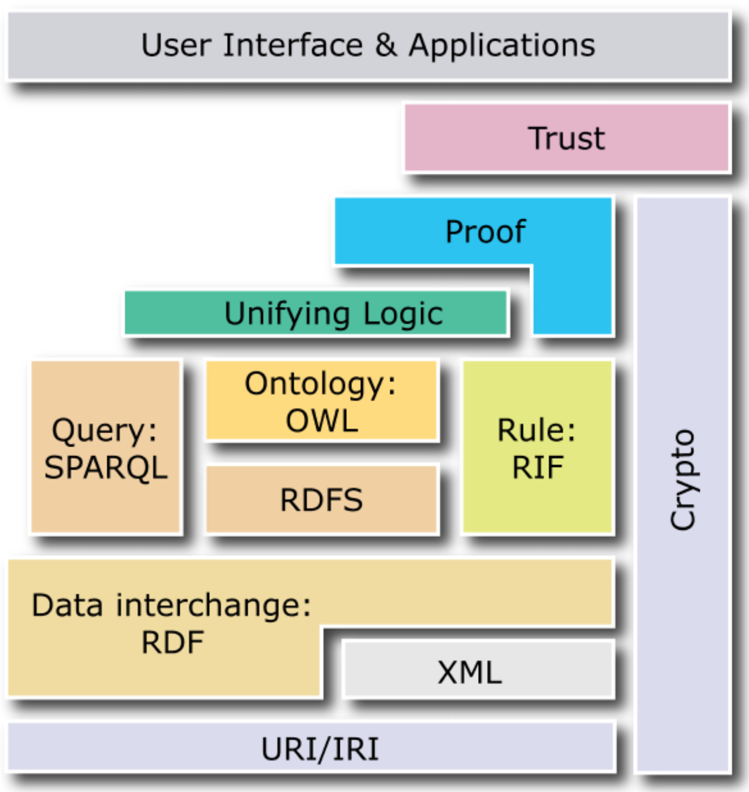

Figure 1: W3C semantic web technology stack [2].

Basic semantic web standards and technologies:

- **XML (eXtensible Markup Language)** is a markup language used to describe the hierarchical structure of text documents via tags. The tag is a structure which is used for the initial and final border along a defined element. The tag can be understood as a means for delivering metadata to text structure, which encloses.

- **RDF (Resource Description Framework)** is a directed, labeled graph data format for representing information in the Web. RDF is used to write the metadata [3].
- **RDFS (Resource Description Framework Schema)** works as a basic language for creating ontologies with very simple semantics. This scheme extends the RDF language to express properties of objects, design objects, classes and their hierarchy description. RDFS allows to express relationships between sources and can be divided into two groups - classes and properties [4].
- **OWL (Ontology Web Language)** is a markup language used for working with ontologies. Ontologies define the relations between individual items of RDF.
- **SPARQL (SPARQL Protocol and RDF Query Language)** is the query language for RDF. RDF specification defines the syntax and semantics of the SPARQL [3].
- **RIF (Rule Interchange Format)** is a format which is used to define a standard for exchanging rules among rule systems, in particular among Web rule engines [5].

2.5 Query

The Semantic Web is viewed as a global database, then we need a query language for that data. The query methods have looked for patterns within a single triple by setting the subject, predicate, or object to a wildcard. This is useful, but by treating each triple independently, we aren't able to easily query across a graph of relationships. It is these graph relationships, spanning multiple triples, that we are most interested in working with [2]. SPARQL is the query language for the Semantic Web.

2.5.1 SPARQL

The SPARQL provides a standardized query language to create queries over RDF graphs using the models RDF triples together with logical operations conjunction and disjunction. The construction of these queries allows us to obtain complex information from the source ontology. [3] SPARQL query consists of three basic parts:

- **prologue** - defines namespaces and prefixes that are used in other parts of the query.

- **query header** - specifies the type of query to be performed in the next section.
- **main part** - is used to define the target ontology (RDF graph) to query and variables including the very search terms in the form of graph patterns.

For illustration, below the notation is a simple SPARQL query which is used to find the titles and types of all devices in implicitly given graph.

```
PREFIX eegdatabase: <http://eegdatabase.kiv.zcu.cz/device>
SELECT ?title ?type
WHERE  {
     ?device x eegdatabase:Device.
     ?device eegdatabase:Title ?title.
     ?device eegdatabase:Type ?type.
}
```

2.6 Resource Description Framework

Technological base of the Semantic Web is the language RDF, which serves as a general framework for the description, exchange and reuse of metadata. This framework provides a simple model used for describing resources that is independent of the specific implementation. The actual details of the object are made through predication that are referred to as the triples. Each of the triples forms together a subject, a predicate and an object. The subject is any object identified by URI and has the property, which we describe by triple. The property is described by the predicate. The target object is the value that the previous object acquires for the predicate. This principle can be illustrated on a simple prediction, received the sentence "Amplifier has property sampling frequency." The subject is the *Amplifier*, predicate is a property "*has property*" and object is the *sampling frequency*, as shown in Fig. 2.

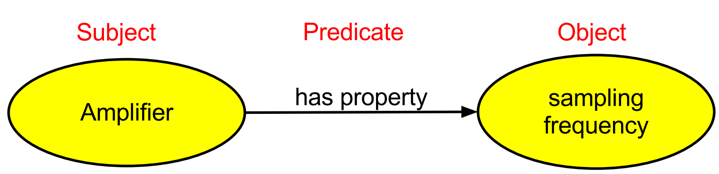

Figure 2: Example of RDF triple.

For the components of each triples graph applies the following rules:

- **subject** - can be formed by a URI identifier or the anonymous list

- **predicate** - can be formed by the URI identifier
- **object** - can be formed by the URI identifier, anonymous list or literal

The graphic design or XML syntax is used for write RDF triples graph. Below a sample code represents the triple from fig. 2.

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description rdf:about="http://www.w3.org/Device/Amplifier">
    <dc:has property>sampling frequency</dc:has property>
  </rdf:Description>
</rdf:RDF>
```

RDF is used for creating the knowledge base (RDF triples tree). The knowledge base is a repository that provides a means for specific area (e.g. electrophysiology and experiments to be collected, organized, shared, searched and utilized). The knowledge base must be depend on ontology.

2.7 Ontologies

At the heart of all Semantic Web applications are ontologies. T. Gruber formulated the definition of an ontology: "An ontology is an explicit specification of a conceptualization" [6]. W. Borst made modification of the definition and its wording is: "a formal specification of shared conceptualization" [7]. In other words, ontologies are formalized vocabularies of terms. Yann Le Franc defined ontologies in an alternative way: "Ontologies are formal models of knowledge in a particular domain and composed of classes that represent concepts defining the field as well as the logical relations that link these concepts together" (see Fig. 3) [8]. Designing an ontology is a long process in which it is necessary to understand the area and compile a list of used terms.

Domain ontology is the most widely used ontology. It describes each specific area, which is defined in general (medicine, neuroinformatics) or specifically (the issue of certain diseases, measuring ERP). The OWL language is used for the implementation of the ontology.

There is a large number of tools for creating ontologies. The most widespread and most elaborate tool is Protege[9].

[9] http://protege.stanford.edu/

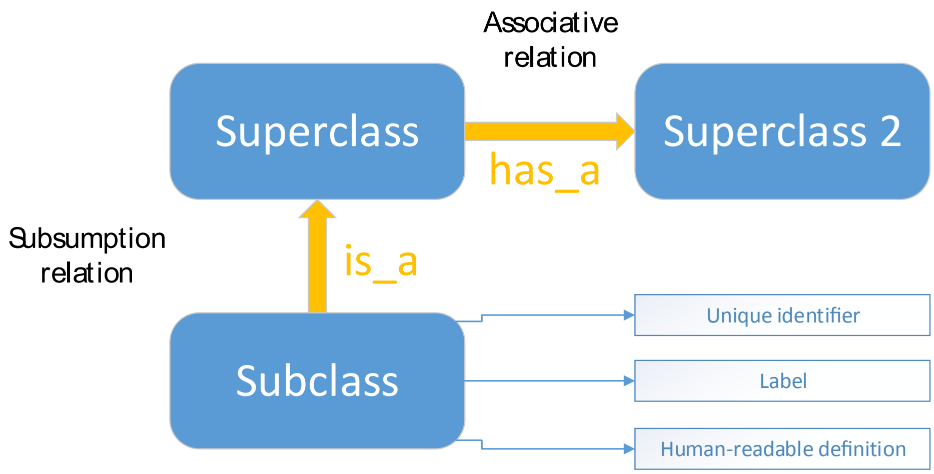

Figure 3: Model of ontology [8].

2.7.1 OWL

OWL (Web Ontology Language) provides three different languages (see Fig. 4) intended for use in specific cases:

- **OWL Lite** allows users to define a hierarchical class system with simple constraints links [9]. It also provides resources for notation the symmetric, transitive and inverse properties, and also a simple restriction on the size of the selected set of model objects - cardinality, but which in this version is limited to acceptable values 0 and 1 [10].
- **OWL DL** offers all OWL language constructs which are subject to certain restrictions on their use (e.g. it is not possible to use cardinality constraints for properties defined as transitive) [10]. It allows to define the union, disjunction and supplements classes or any constraints cardinality [9].
- **OWL FULL** is a variant that provides maximum expressivity, which uses all OWL language constructs. This option will not impose any restrictions on the evaluation rules, thus it is impossible to guarantee the computational satisfiability [10].

OWL 2 is an extension and revision of OWL developed by the W3C working group which was published in 2004. Fig. 5 gives an overview of the language OWL 2. It shows the basic construction blocks and the relationships between them [12].

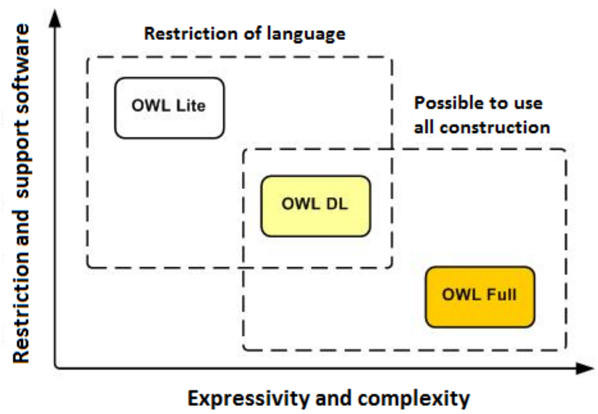

Figure 4: Dialects of OWL language [11].

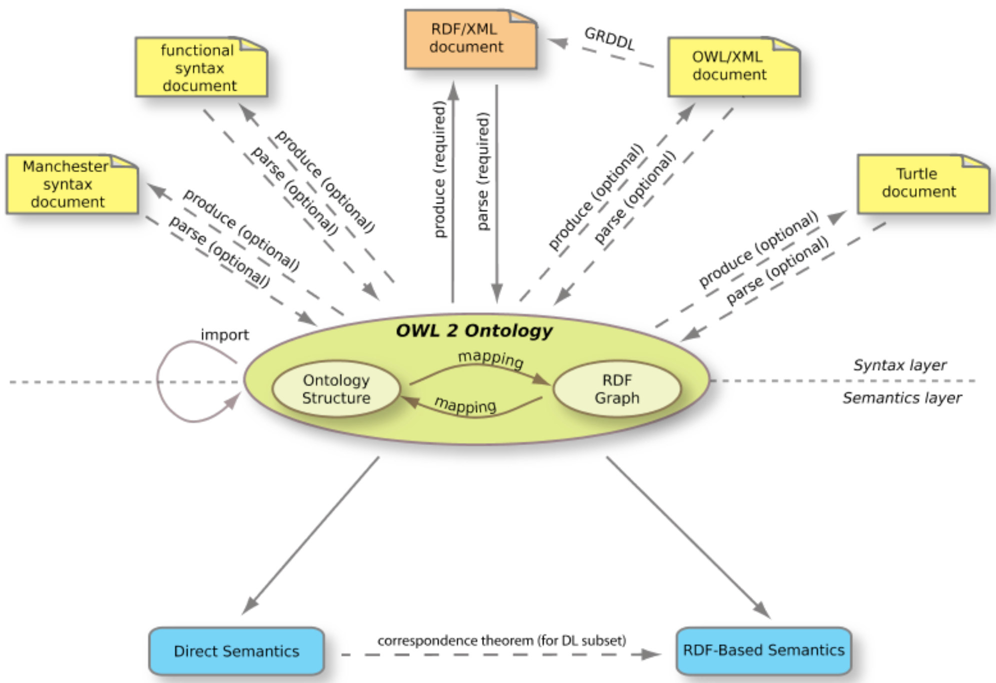

Figure 5: The structure of the language OWL 2 [12].

2.8 Project

There are many projects and tools for Semantic Web solutions [13].

2.8.1 DBpedia

This project is an effort to publish structured data extracted from Wikipedia: the data is published in RDF and made available on the Web for use under the GNU Free Documentation License, thus allowing Semantic Web agents to provide inferencing and advanced querying over the Wikipedia-derived dataset and facilitating interlinking, re-use and extension in other data-sources [14].

2.8.2 Friend of a Friend

A popular vocabulary on the semantic web is Friend of a Friend (FOAF). This project uses RDF to describe people, the relationships between them and the things around them (work, school and etc.). FOAF defines an open, decentralized technology for connecting social Web sites. [15]

2.8.3 Semantically-Interlinked Online Communities project

This project (SIOC) provides a vocabulary of terms and relationships that model web data spaces (e.g. discussion forums, blogs, blogrolls / feed subscriptions, mailing lists, shared bookmarks and image galleries) [16]. SIOC reuses the Dublin Core model to define various attributes of created content and FOAF. Interlinking FOAF and SIOC also provides a model for identity federation on the Web. SIOC has become an inevitable core element of the Social Semantic Web [17].

2.8.4 GoPubMed

GoPubMed is a knowledge-based search engine (ontology-based browser) for biomedical texts (for PubMed). The Gene Ontology (GO) and Medical Subject Headings (MeSH) serve as "Table of contents" in order to structure the millions of articles of the MEDLINE database [18].

2.8.5 NextBio

A database consolidating high-throughput life sciences experimental data tagged and connected via biomedical ontologies. Nextbio is accessible via a search engine interface. Researchers can contribute their findings for incorporation to the database. The database currently supports gene expression or protein expression data and sequence centric data and is steadily expanding to support other biological data types [19].

2.9 State of the Art of ontologies in the semantic web

A survey (published in 2006) of ontologies available on the web collected 688 OWL ontologies. Of these, 199 were OWL Lite, 149 were OWL DL and 337 OWL Full (by syntax). They found that 19 ontologies had in excess of 2 000 classes, and that 6 had more than 10 000. The same survey collected 587 RDFS vocabularies.[20] DARPA Agent Markup Language (DAML) ontology library contains about 250 examples written in OWL or DAML+OIL (a converter from DAML+OIL to OWL is available on the web). In addition, several large ontologies have been released in OWL. These include a cancer ontology developed by the US National Cancer Institute's Center for Bioinformatics, which contains about 17,000 cancer related terms and their definitions, and an OWL version of the well-known GALEN medical ontology, developed at the University of Manchester. [21]

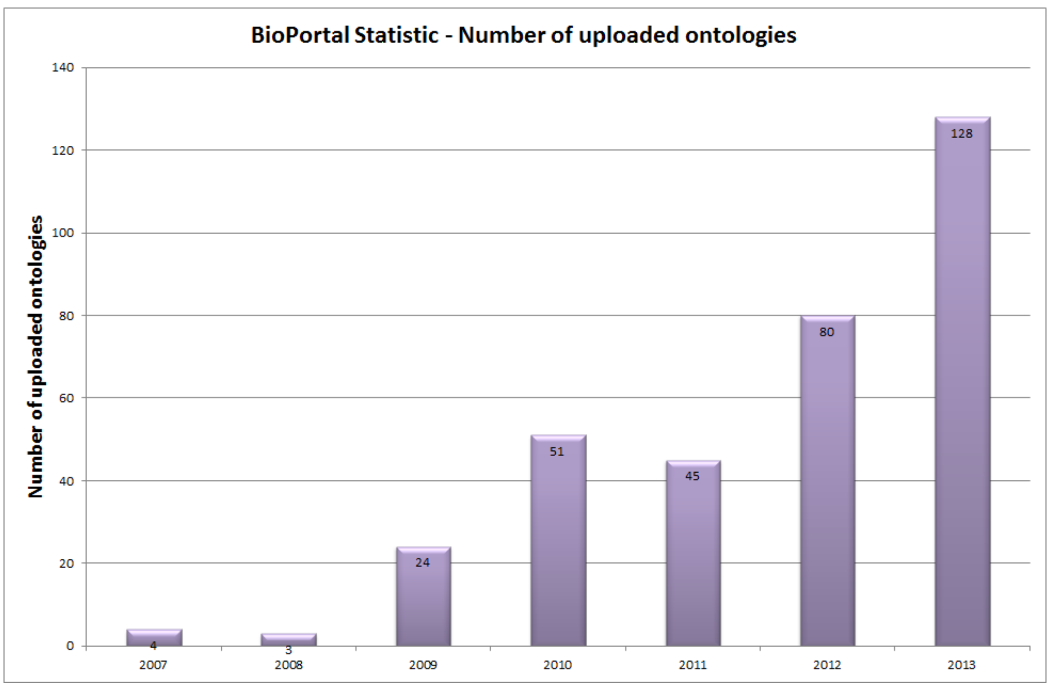

Figure 6: BioPortal Statistic - Number of uploaded ontologies.

Fig. 6 shows a graph with number of uploaded ontologies from year 2007 to present time. There are currently more than 350 biomedical ontologies, 6 080 782 terms, 38 resources indexed, 4 739 989 indexed records, 2 152 326 315 direct annotations, 26 050 710 506 direct plus expanded annotations in the National Center for Biomedical Ontology (NCBO) BioPortal[10] [22].

[10]http://bioportal.bioontology.org/

Table 1: Availability of ontologies within web-based systems

Web-based systems	Available Ontologies
BioPortal	352
OntoFox	127
Ontobee	125
Neuroscience Information Framework	56
Ontology Lookup Service	84

There are a lot of web-based systems to support ontology reuse. Tab. 1 shows the availability of ontologies within web-based systems.

3 Electrophysiology and neurophysiology community

Electrophysiology is a technology used to study the electrical properties of biological cells and tissues. Electrophysiology typically involves the measurements of voltage change or electrical current flow on a wide variety of scales from single ion channel proteins to whole tissues [23]. Neurophysiology is the study of nervous system function. The primary tools of basic neurophysiological research include electrophysiological recordings such as patch clamp and calcium imaging, as well as some of the common tools of molecular biology. Neurophysiology is connected with electrophysiology and other brain sciences. [24]

3.1 Our research group

Our research group at the Department of Computer Science and Engineering, University of West Bohemia, a member of the Czech INCF National Node, specializes in the research into electroencephalography (EEG) and event-related potentials (ERP). These techniques were expected to become obsolete when hemodynamic methods (e.g. PET and (f)MRI) were developed, but due to their high temporal resolution they are currently viewed as important complements to them. [25]

3.1.1 EEG/ERP experiments

Our research group focuses on exogenous event-related potentials that are emitted by the brain as it makes the decision or initiates the response. [26]

The EEG/ERP data measured at our department are usually based on odd-ball paradigm. These experiments usually try to elicit the P3 response by presenting a surprising and task-relevant stimulus [27]. The following EEG/ERP experiments are performed in our laboratory:

- Driver's attention during monotonous driving (visual and auditory stimulation)
- Traditional oddball EEG/ERP experiments (visual stimulation, e.g. based on the LED stimulation)
- Attention of children with developmental coordination disorder (auditory stimulation)

The following EEG/ERP experiment are performed in the laboratory of Faculty of Medicine in Pilsen:

- Visual cortex of mice (Steady State Visually Evoked Potentials)

3.1.2 EEG/ERP laboratory

For our experiments, the EEG/ERP laboratory at the Department of Computer Science and Engineering at the University of West Bohemia is used. The laboratory is located on the fourth floor of the building. It contains a soundproof cabin, standard EEG caps placed according to 10-20 system with 19 electrodes, an amplifier, different kinds of stimulation devices and a computer with the software for storing data [28]. The diagram of commonly used devices in our laboratory is shown in Fig. 7.

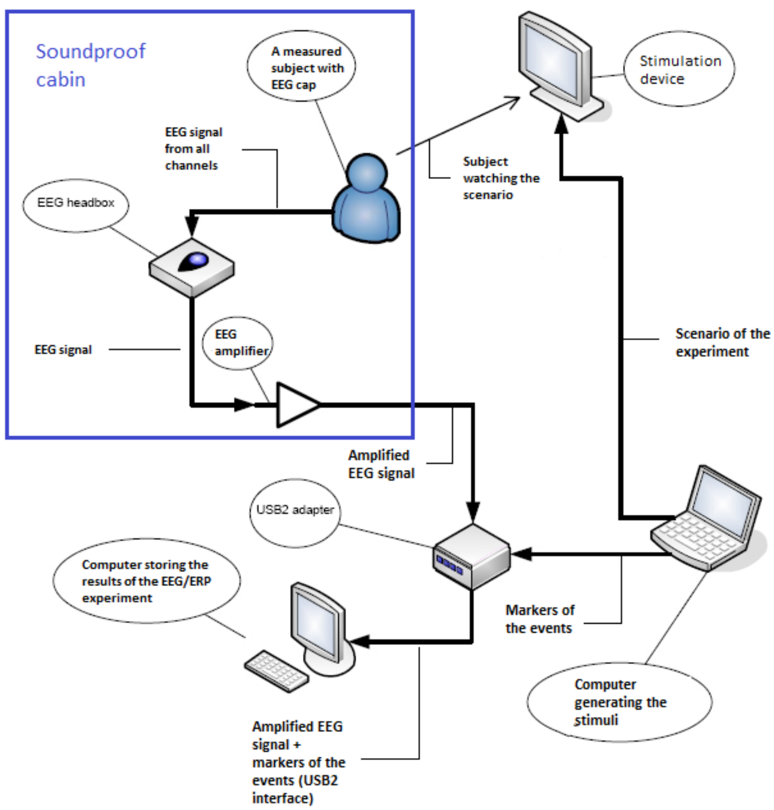

Figure 7: EEG laboratory - Diagram [28].

3.1.3 Portal for Research in Electrophysiology

The EEG/ERP Portal allows groups of researchers to store, update, download data and metadata from the EEG/ERP experiments measured in our laboratory. On the basis of activities that a user can perform we recognize four user roles

(Reader, Experimenter, Group Administrator, Supervisor). User who wants to upload a custom experiment or download experiments he/she is forced to create an account and to be a member of at least one group. The user can create an account within the EEG/ERP Portal directly or can use a custom Facebook or LinkedIn account (such account is synchronized with the Portal after first login). Fig. 8 shows the main web page of the portal [29].

Figure 8: EEG/ERP Portal - User Interface Preview [30].

3.2 EPhys ontology workgroup

The analysis of the existing ontological resources reveals a lack of terms for accurate and unambiguous annotation of electrophysiological data and metadata. With the development of different resources for describing and sharing this particular type of data, the community needs controlled vocabularies to describe the different types of electrophysiology recording (from single cell to EEG recordings). [31]

As electrophysiological recordings are often used in multi-modal neurophysiological experiments, the scope of the ontology has been extended to cover any type of neurophysiological experiments. The development of OEN[11] has been separated along two main branches: the device/method branch and the neurophysiological concept branch (describing action potential, action potential features, etc.). [31]

The device branch terminology is built upon existing ontologies related to neurophysiological experiments or investigation, namely the Ontology for Biomedical Investigation[12](OBI) and the Neural ElectroMagnetic Ontologies[13](NEMO). Existing terms describing neurophysiological set-ups and investigations have been imported using the MIREOT format [32] and the web service Ontofox[14]. The granularity of the ontology has been extended using the odML terminology[15] and terms from the EEGBase. To test the ontology, the group created a simple knowledge base to describe the content of the EEGBase database. [31]

In parallel, the members of the group are developing a terminology to describe neurophysiological concepts such as action potentials for instance. This work is done in collaboration with other relevant ontologies such as the Phenotypic Quality Ontology, PATO[16] and the Gene Ontology, GO[17]. [31]

[11]OEN - a formal ontology for describing experimental neurophysiology
[12]http://obi-ontology.org/
[13]http://purl.bioontology.org/ontology/NEMO
[14]http://ontofox.hegroup.org/
[15]http://www.g-node.org/projects/odml/terminologies
[16]http://obofoundry.org/wiki/index.php/PATO:Main_Page
[17]http://www.geneontology.org

4 Current ontologies

4.1 Neural ElectroMagnetic Ontology

The main aims of this project (NEMO) is to create EEG and MEG ontologies and ontology based tools. These resources will be used to support representation, classification, and meta-analysis of brain electromagnetic data. The three pillars of NEMO are:

1. DATA - consist of raw EEG, averaged EEG (ERPs), and ERP data analysis results.
2. ONTOLOGY - describes classes of event-related brain potentials (ERP) and their properties, including spatial, temporal, and functional (cognitive/behavioral) attributes, and data-level attributes (acquisition and analysis parameters).
3. DATABASE - is a large repository that stores NEMO consortium data, data analysis results, and data provenance.

Minimal Information for Neural ElectroMagnetic Ontologies (MINEMO) is the minimal set of experiment meta-data that is required for datasets that are used in the NEMO project. It is modeled on MINI (Minimal Information about a Neuroscience Investigation), which was developed by Frank Gibson and colleagues for the CARMEN. NEMO has 1817 classes, 119 individuals and 98 properties. [33] [23]

4.1.1 MINEMO

This project contains three categories:

1. **Standardized checklist** - specifies guidelines for reporting eight sets of fields: General features of experiments, Study subject, Anatomical location of electrophysiological recording, Experimental task, Experimental stimuli, Behavioral response data, Recording specifications and Electrical (time series) data.
2. **Controlled vocabularies or lexicons for data annotating** - contain the BrainMap lexicon which has enjoyed widespread use in connection with their database in human neuroscience. The BrainMap database is a repository with results from thousands of functional brain imaging studies. The BrainMap lexicon covers a range of metadata, including stimuli, tasks (instructions), and protocols for measurement of behavioral and brain responses.

3. **Formal ontology** - contains semantic categories or classes that refer to well-defined entities (e.g., stimulus, response). Each class has a uniform resource identifier, or URI, which is globally unique, in addition to a human-readable label. Ontology specifies the semantic relations between classes and these relations are called object properties and impart much of the power behind ontologies. NEMO has adopted many of the recommended practices outlined by the OBO Foundry, including reuse of existing resources, modularity or orthogonality, human-readable annotations and using of Basic Formal Ontology as an upper ontology and Ontology of Biological Investigations as a mid-level ontology. [34]

4.2 Gene Ontology

The Gene Ontology (GO) project provides an ontology of defined terms representing gene product properties. The ontology covers three domains: cellular component, the parts of a cell or its extracellular environment; molecular function, the elemental activities of a gene product at the molecular level, such as binding or catalysis; and biological process, operations or sets of molecular events with a defined beginning and end, pertinent to the functioning of integrated living units: cells, tissues, organs, and organisms. The GO ontology is structured as a directed acyclic graph, and each term has defined relationships to one or more other terms in the same domain, and sometimes to other domains. The GO vocabulary is designed to be species-neutral, and includes terms applicable to prokaryotes and eukaryotes, single and multicellular organisms [35]. Gene Ontology has 39 464 classes, 0 individuals and 8 properties.

4.3 Phenotype And Trait Ontology

PATO is an ontology of phenotypic qualities, intended for use in a number of applications, primarily defining composite phenotypes and phenotype annotation, Phenotypic qualities (properties). This ontology can be used in conjunction with other ontologies such as GO or anatomical ontologies to refer to phenotypes. Examples of qualities are red, ectopic, high temperature, fused, small, edematous and arrested [36]. PATO has 2378 classes, 0 individuals and 10 properties.

4.4 The open biomedical ontologies Foundry

The open biomedical ontologies (OBO)[18] Foundry collects and coordinates ontologies in the biological domain.

[18] www.obofoundry.org

The OBO Foundry[19] is a collaborative experiment involving developers of science-based ontologies who are establishing a set of principles for ontology development with the goal of creating a suite of orthogonal interoperable reference ontologies in the biomedical domain. The groups developing ontologies who have expressed an interest in this goal are listed below, followed by other relevant efforts in this domain. [37].

OBO Foundry Principles

1. The ontology is open in the sense that it is available to be used by all under the following two constraints (1) its origin must be acknowledged and (2) it is not to be altered and subsequently redistributed under the original name or with the same identifiers.

2. The ontology is in, or can be expressed in, a common formal language. (A provisional list of languages supported by OBO is provided at http://obo.sf.net/.)

3. The ontology possesses a unique identifier space within OBO.

4. The ontology provider has procedures for identifying distinct successive versions.

5. The ontology has a clearly specified and clearly delineated content.

6. The ontology includes textual definitions for all terms.

7. The ontology uses relations which are unambiguously defined following the pattern of definitions laid down in the OBO Relation Ontology.

8. The ontology is well documented.

9. The ontology has a plurality of independent users.

10. The ontologies in the OBO Foundry will be developed in a collaborative effort. [38]

4.5 OBO Relations Ontology

OBO RO is a collection of relations intended primarily for standardization across ontologies in the OBO Foundry and wider OBO library. It incorporates core upper-level relations such as part of from BFO. [38]

[19] www.obofoundry.org

4.6 NIFSTD

NIF Standard ontology (NIFSTD) is a core component of the NIF. The ontology includes a set of modular ontologies that provide a comprehensive collection of terminologies to describe neuroscience data and resources. NIFSTD relies on these existing biomedical ontologies as Chemical Entities of Biological Interest (ChEBI), Gene Ontology, Protein ontology, Ontology for Biomedical Investigation and The Ontology of Phenotypic Qualities. The NIFSTD is constructed according to the best practices closely followed by the Open Biological Ontology (OBO) community. NIFSTD has 90 064 classes, 548 individuals and 366 properties.

4.6.1 Basic structure and design

NIFSTD follows OBO Foundry principles [38] to developing highly interoperable and reusable reference ontologies. **Modularity** NIFSTD is founded on a modular fashion, each module covering orthogonal domain of Neuroscience concepts (macroscopic anatomy, cell types, techniques, nervous system function, molecules etc.). Following a modularization ontology pattern (ODPs[20]), NIFSTD promotes easy extendibility towards its evolution [39] [40]. Each of the modules is standardized to the same upper level ontologies such as the Basic Formal Ontology and OBO Relations Ontology. [41]

Representation Language NIFSTD is expressed in W3C standard OWL-DL dialect. [41]

Reuse of Available Ontology structure NIFSTD reuses existing available knowledge sources, terminologies and ontologies [41]. The process of importing or adapting a new ontology or vocabulary source varies depending upon its state [42].

- If a source already uses OWL, the OBO-RO and the BFO and is orthogonal to existing modules, the import simply involves adding an owl:import statement to the main ontology file (nif.owl). [41]
- If an existing orthogonal ontology is in OWL but does not use the same foundational ontologies as NIFSTD, then an ontology-bridging module (explained later) is constructed declaring the deep level semantic equivalencies such as foundational object and processes. [41]
- If an external source is satisfiable by the above two rules but observed to be too large for NIF's scope of interests, a relevant subset is extracted as suggested by NIF's domain experts (e.g., CHEBI, OBI). MIREOT principle is used here because allows extracting a required part of larger ontology. [41]

[20]http://odps.sourceforge.net/

- If the external source has not been represented in OWL, or does not use the same foundation as NIFSTD, then the terminology is adapted to OWL/RDF in the context of the NIFSTD foundational layer ontologies. [41]

Single Inheritance NIFSTD follows the simple inheritance principle for the hierarchy of named classes; i.e., an asserted named class can have only one named class as its superclass; however, a named class can have multiple anonymous superclasses. This principle promotes the named classes to be univocal and to avoid to ambiguities. The classes with multiple superclass are derivable via automated classification on defined NIFSTD classes with necessary and sufficient conditions. [41]

Unique Identifier and Annotation Properties NIFSTD entities are identified by a unique identifier and accompanied by a variety of annotation properties from Dublin Core Metadata (DC) and the Simple Knowledge Organization System model. [41]

Object Properties and Bridge Modules NIFSTD object properties are mostly drawn from the OBO Relations Ontology (OBO-RO). The cross-domain relations are specified in separate bridging modules - modules that only contain logical restrictions and definitions on a required set of classes assigned between multiple modules [42]. The bridging modules allow the main domain modules to remain independent of one another without the bridging modules.

Versioning Various annotation properties are associated with versioning different levels of content within NIFSTD. These include creation and modification dates for each of the classes; file level versioning for each of the modules, annotations for retiring antiquated concept definitions, tracking former ontology graph position and replacement concepts. [41] [42]

Accessing NIFSTD Ontologies NIFSTD is available in OWL format and can be loaded by the OWL editing tools (e.g. Protege). It is also available through NCBO BioPortal[21]. NIFSTD is served through an ontology management system called OntoQuest[22]. NIFSTD has a SPARQL endpoint[23] and is also available in RDF. [41]

4.7 Basic Formal Ontology

Basic Formal Ontology (BFO) grows out of a philosophical orientation which overlaps with that of DOLCE and SUMO. Unlike these, however, it is narrowly focused on the task of providing a genuine upper ontology which can be used

[21]http://bioportal.bioontology.org/ontologies/40510
[22]OntoQuest Web Services - http://ontology.neuinfo.org/ontoquestservice.html
[23]NIFSTD SPARQL Endpoint - http://ontology.neuinfo.org/sparqlendpoint.html

in support of domain ontologies developed for scientific research, as for example in biomedicine within the framework of the OBO Foundry. Thus BFO does not contain physical, chemical, biological or other terms which would properly fall within the special sciences domains. [43]

Many of the members of the Open Biomedical Ontologies (OBO) Foundry initiative, including the Gene Ontology, the Foundational Model of Anatomy, the Protein Ontology, and the Ontology for Biomedical Investigations (http://www.obofoundry.org/) are utilizing Basic Formal Ontology to assist in the categorization of entities and relationships in their respective domains of research. BFO adopts a view of reality as comprising (Fig. 9) (1) such as object, qualities, and functions, and (2) occurents, the events or happenings in which continuants participate.

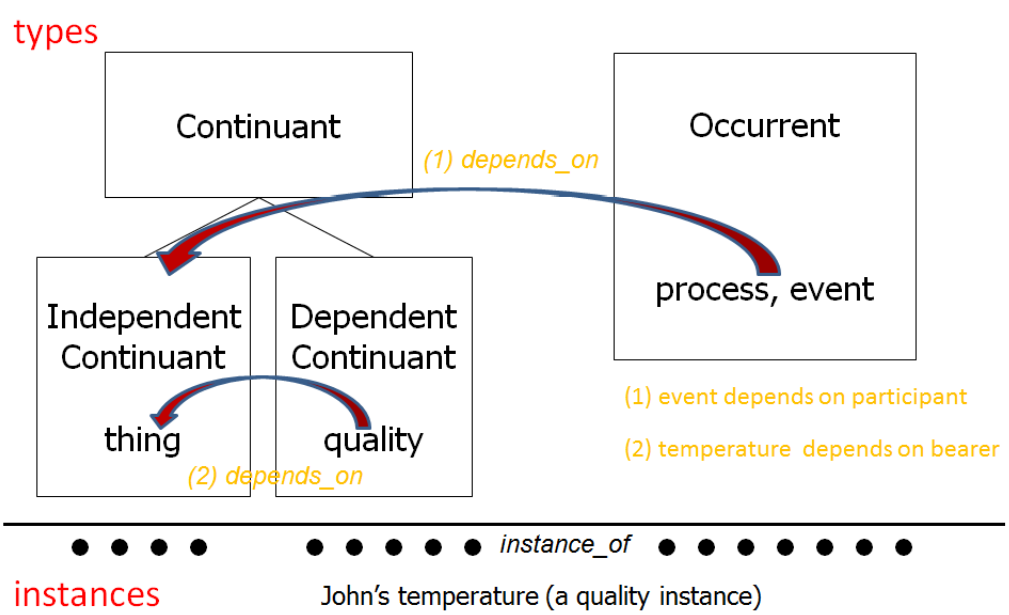

Figure 9: BFO view [44].

BFO has 39 classes, 0 individuals and 0 properties. The subtypes of continuant (thing, quality, . . .) and occurrent (process, event, . . .) categories represented in BFO are shown in Sections 4.7.1 and 4.7.2. [45] Fig. 10 shows rationale of OBO Foundry coverage.

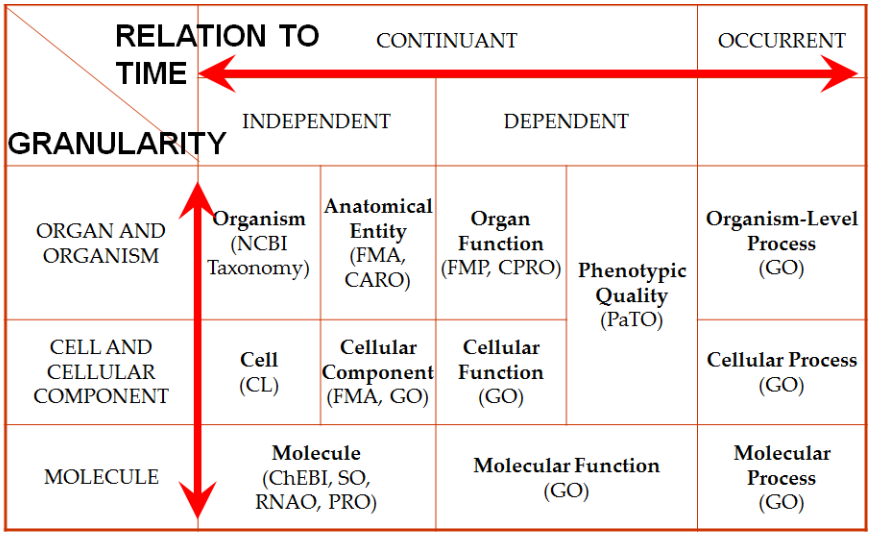

Figure 10: BFO - rationale of OBO Foundry coverage [44].

4.7.1 The continuant categories of BFO

There are represented the subtypes of continuant in BFO:

- BFO:entity
  - continuant
    - independent continuant
      - object
      - object boundary
      - object aggregate
      - fiat object part
      - site
    - dependent continuant
      - generically dependent continuant
      - specifically dependent continuant
        - quality
        - realizable entity
          - function
          - role
          - disposition
    - spatial region
      - zero-dimensional region
      - one-dimensional region
      - two-dimensional region
      - three-dimensional region

4.7.2 The occurrent categories of BFO

There are represented the subtypes of occurrent in BFO:

BFO:entity
- processual entity
  - process
  - process boundary
  - process aggregate
  - fiat process part
  - processual context
- spatiotemporal region
  - scattered spatiotemporal region
  - connected spatiotemporal region
    - spatiotemporal instant
    - spatiotemporal interval
  - temporal region
    - scattered temporal region
    - connected temporal region
      - temporal instant
      - temporal interval

4.8 Computational Neuroscience Ontology

Computational Neuroscience Ontology (CNO) is a controlled vocabulary of terms used in Computational Neurosciences to describe models of the nervous system. This first release of CNO is an alpha version and should be further aligned with other ontologies accessible on BioPortal and should be made compliant with the OBO foundry recommendations. [46].

4.9 Information Artifact Ontology

The Information Artifact Ontology (IAO) is a new ontology of information entities, originally driven by work by the OBI digital entity and realizable information entity branch [47]. IAO has 173 classes, 20 individuals and 55 properties.

4.10 Ontology for Biomedical Investigations

The Ontology for Biomedical Investigations (OBI) Consortium is developing an integrated ontology for the description of biological and clinical investigations, written in OWL DL. [48] OBI has 3689 classes, 163 individuals and 114 properties.

OBI uses the Basic Formal Ontology as its upper-level ontology. OBI structure was separated into 10 sections (biomaterial, data transformation, digital entity, function, instrument, plan, protocol applications, qualities, role and relations) called branches. These branches allow independent development by different groups. Each branch is maintained in separate OWL file, and contains closely related terms and definitions.

4.10.1 OBI development practices

Minimal Information to Reference External Ontology Terms (MIREOT) - OBI built-in import mechanism is to copy only parts of external ontology into *obi.owl* using MIREOT. [48]

Releasing OBI - is a mechanism that allows the release of the public version of OBI on a monthly basis. The release process includes checks for content quality (e.g., annotations compliant with our policy), syntax (e.g., OWL species validation), and reporting candidate release status to the ontology developers. [48]

Quality check and reports - Jena-based script to read branch files and indentify missing elements, duplicates, or misuse of any of their metadata properties is used. The reports are simple HTML pages displaying terms and associated issues. [48]

Identifier maintenance policy - Having a stable and consistent ID policy is a fundamental OBO Foundry principle. In OBI, identifier are prefixed with "OBI_" and followed by seven digits. [48]

Managing disjoints - is a mechanism of manually added disjoints to classes as building the ontology [48].

Distributing OBI with inferred superclasses - The defined classes are used and an easy-to-use file is provided that does not require the use of reasoner on the end-user side. Therefore they assert, via script, the inferred superclasses to their OWL file. [48]

Assuming that all classes have instance - The OWL reasoners (Pellet and Fact++) are used. The mechanism of scripting the addition of anonymous individuals of each type named in the ontology as part of the release process is used. This mechanism is used for each leaf class and before computing the inferrend superclasses. [48]

Increasing the readability of the RDF/XML - Numerical identifiers for all entities are used. Numeric identifiers ensure that a human-readable label can be changed without needing to change the URI, and establishes an unbiased basis for internationalization. However, they sometimes need to edit the OWL RDF/XML directly, which is cumbersome because IDs are not easily remembered.

To increase human readability they post-process the RDF/XML and generate XML comments for the released version of the file. [48]

4.11 Summary

4.11.1 Current state

Semantic web solutions for EEG/ERP data management focus mainly on building ontologies. The new terminology is build upon existing ontologies related to neurophysiological experiments or investigation, namely the NEMO and OBI. The NEMO project provides formal semantic definitions of concepts in ERP research, including ERP patterns, spatial, temporal, functional (cognitive/behavioral) attributes of these patterns, and data acquisition and analysis methods and parameters.

4.11.2 Possible improvements

A recent trend in bio-informatics and neuro-informatics is creation of domain ontologies [49]. The granularity of the new ontologies has been extended using the odML terminology and terms from the EEGBase. Since there is no ontology describing EEG/ERP experiments in our laboratory [25], including visual evoked potential experiment with mice, conditions and circumstances during EEG/ERP experiments (e.g. weather conditions, laboratory environment, measured subjects' emotional state etc.), it is necessary to define the related terminology and to design and create new ontologies.

4.11.3 Best Ontology Practices

The new ontologies must have the recommended practices outlined by the OBO Foundry [42], including reuse of existing resource (terms, ontologies, modularity or orthogonality, human-readable annotations and perhaps most important - use of the Basic Formal Ontology as an upper ontology and the Ontology of Biological Investigations as a mid-level ontology. The next practice is cooperation with community of researchers who design and create ontologies (EPhys ontology workgroup, NIF), and develop web-based systems providing access to a library of biomedical ontologies and terminologies (BioPortal, NIF, OntoFox, Ontobee). [41] [34]

Benefits of coordination:

- No need to reinvent the wheel
- Can profit from lessons learned through mistakes made by others
- Can more easily reuse what is made by others
- Can more easily inspect and criticize results of others work (PATO)
- Leads to innovations (e.g. Mireot) in strategies for combining ontologies [44]

5 Current terminology

5.1 Minimum Information about a Neuroscience Investigation for Electrophysiology

A MINI:Electrophysiology minimum represents the minimum requirements that should be reported about a dataset to facilitate computational access and analysis to allow a reader to interpret and critically evaluate the processes performed and the conclusions reached, and to support their experimental corroboration. In practice a MINI comprises a checklist of information that should be provided (for example about the protocols employed) when a data set is submitted to the CARMEN system[24]. [23]

5.2 Reporting requirement for description in neurophysiological domain

This section describes requirements for description in neurophysiological domain. Each checklist term has a definition. The first MINI checklist is shown in Appendix A and the second checklist of EEG/ERP portal terminology is shown in Appendix B.

[24] https://portal.carmen.org.uk/

6 Current data formats

6.1 odML

G-node[25] proposed a simple format, the "open metaData Markup Language" (odML), for collecting and exchanging metadata in an automated, computer-based fashion. In odML arbitrary metadata information is stored as extended keyvalue pairs in a hierarchical structure. Central to odML is a clear separation of format and content, i.e., neither keys nor values are defined by the format. This makes odML flexible enough for storing all available metadata instantly without the necessity to submit new keys to an ontology or controlled terminology. Common standard keys can be defined in odML-terminologies for guaranteeing interoperability. [32]

For annotation and sharing of data it is necessary to have a format that fulfills certain requirements:

1. Easy to use and ideally human readable.
2. Can be implemented into any recording, analysis or management tool.
3. Open and freely available.
4. Inherently extensible and flexible for changes.
5. More or less unresticted i.e. it must/should not restrict the user or strictly require entries. [50]

6.1.1 Description of the odML data model

Data exchange requires that also annotations, metadata, are exchanged. In order to allow interoperability we need both a common (meta) data model, the format in which the metadata are exchanged, and a common terminology. The data model of the odML is based on the idea of key-value pairs like sampling frequency = 1000Hz. The group tried to keep the model as simple as possible while being flexible, allowing interoperability, and being customizable. The model defines four entities (Property, Section, Value, RootSection) who's relations and elements are shown in the Fig. 11. More information and a detailed description can be (soon) found on the documentation pages. Property and Section entities are the core of the odml. A Section contains Properties and can further have subsection thus building a tree-like structure. The model further does not control the content which is a risk, on the one hand, but offers the flexibility we consider essential. So far, this model has been implemented as a XML schema. [50] [32]

[25] German Neuroinformatics Node

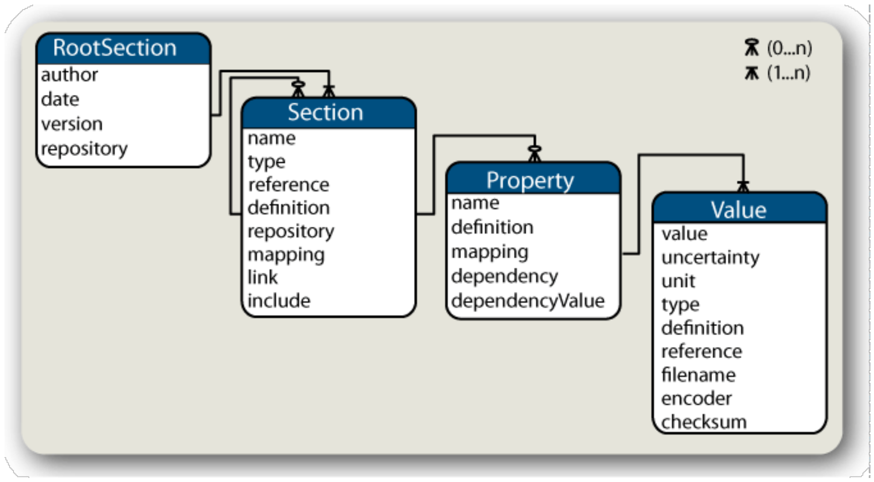

Figure 11: odML ERA model [50].

6.2 Other ML formats

6.2.1 SignalML

Signal Markup Language[26] (SignalML) provides simple and effective process of encoding the metainformation, needed for a proper interpretation of digital time series, stored in different formats. SignalML allows describing other formats for storing biomedical data (including EEG data). Unlike the actual software, created in thousands of instances for conversions, display or analysis of data in particular data formats, only one instance (metainformation SignalML file) is needed for a given format to make it readable by any compliant software [51]. Fig. 12 shows SignalML 2.0 layers.

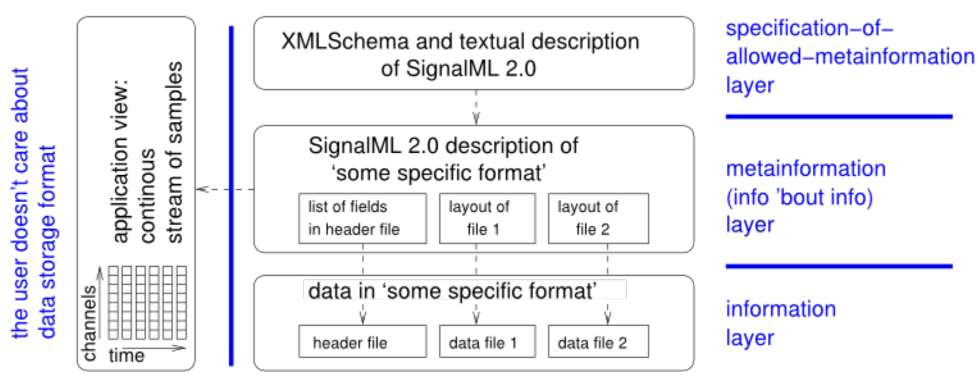

Figure 12: SignalML 2.0 layers [51].

[26] http://bci.fuw.edu.pl/wiki/SignalML

6.2.2 NeuroML

NeuroML is a language for describing detailed models of neural systems [52].

The aims of the NeuroML initiative are:

- To create specifications for a language in XML to describe the biophysics, anatomy and network architecture of neuronal systems at multiple scales.
- To facilitate the exchange of complex neuronal models between researchers, allowing for greater transparency and accessibility of models.
- To promote software tools which support NeuroML and support the development of new software and databases.
- To encourage researchers with models within the scope of NeuroML to exchange and publish their models in this format. [52]

6.3 Hierarchical Data Format

Hierarchical Data Format (HDF5) is a file format, data model and an I/O library for storing and managing data. It supports an unlimited variety of datatypes, and is designed for flexible and efficient I/O and for high volume and complex data. HDF5 is portable and is extensible, allowing applications to evolve in their use of HDF5. The HDF5 Technology suite includes tools and applications for managing, manipulating, viewing, and analyzing data in the HDF5 format. [53] [54]

The HDF5 technology suite includes:

- A versatile data model that can represent very complex data objects and a wide variety of metadata.
- A completely portable file format with no limit on the number or size of data objects in the collection.
- A software library that runs on a range of computational platforms, from laptops to massively parallel systems, and implements a high-level API with C, C++, Fortran 90, and Java interfaces.
- A rich set of integrated performance features that allow optimizations of storage space and access time.
- Tools and applications for managing, manipulating, viewing, and analyzing the data in the collection.
- The HDF5 data model, file format, API, library, and tools are open and distributed without charge.[53]

7 Current tools

7.1 Minimum information to reference an external ontology term

The Minimum Information to Reference an External Ontology Term (MIREOT[27]) allows us to do so by providing a way to import external terms from ontologies not yet using BFO as an upper ontology, or not yet using OWL DL [47].

Challenges of imports:

- **Large overhead** - using large ontologies, such as NCBI Taxonomy or Foundational Model of Anatomy (FMA)
- **True Alignment** - Ontologies constructed using a different design, or not using BFO as upper level ontology prevents full integration
- **Fluid development** - Resources under development [47]

7.1.1 Defining minimum information

The first step is to define URI of the class, URI of the source ontology and position in the target ontology. This minimal set allows us to ambiguously identify a term. The second step is to capture additional information (label, definition, other annotation - "human-readable", superclasses: e.g. NCBI taxonomy, ...). [47]

7.1.2 Implementation

This section describes an implementation of the MIREOT guidelines. The implementation was performed in the context of the OBI project, and can be decomposed into a two-step process:

1. Gather the minimum information for the external class.
2. Use this minimum information to fetch additional elements, like labels and definitions. [47]

Once the external term is identified for import, the first step is to gather the corresponding minimum information set. This set is stored in a OWL file - external.owl. A Perl script, *add-to-external.pl*[28] is used to automatically append the

[27]http://obi-ontology.org/page/MIREOT
[28]http://obi.svn.sourceforge.net/svnroot/obi/trunk/src/tools/add-to-external.pl

minimum information set to the external.owl file. This script takes as arguments the identifiers of the external class to be imported and its parent class in the target hierarchy, in this case in the OBI hierarchy. [47]

In addition, a mapping mechanism between the prefix used in the identifier and the external source ontology URI is built into the script. Curators therefore need only specify the ID of the external class to import and the ID of the class it should be imported under, within the target ontology. [47]

Additional elements can be obtained programmatically via SPARQL CONSTRUCT queries, as described in Fig. 13. These queries specify which extra information about the class to gather, such as the definition and preferred label, and how to map these into the corresponding OBI annotation properties. [47]

```
prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
prefix owl: <http://www.w3.org/2002/07/owl#>
prefix obi: <http://purl.obofoundry.org/obo/>
prefix obo: <http://www.geneontology.org/formats/oboInOwl#>

construct
{
  _ID_GOES_HERE_ rdf:type owl:Class.
  _ID_GOES_HERE_ alias:preferredTerm ?label.
  _ID_GOES_HERE_ rdfs:label ?label.
  _ID_GOES_HERE_ alias:definition ?definition.
}
where
{
  {   _ID_GOES_HERE_ rdfs:label ?label. }
  UNION
  {   _ID_GOES_HERE_ obo:hasDefinition ?blank.
      ?blank rdfs:label ?definition}
}
```

Figure 13: Template SPARQL query [47].

The external term is directly imported from the external resource, with the status and definition as defined by the external resource. Finally, a script, *create-external-derived.lisp*[29], iterates through the minimum information stored in external.owl. Depending on the source ontology URI of each of our imported terms, it then selects the correct SPARQL template and substitutes the relevant ID. The queries are then executed against the Neurocommons SPARQL endpoint. [47]

This supplementary information, which is prone to change as the source ontologies evolve, is stored in a second file, externalDerived.owl. This file can be removed on a regular basis, e.g., before release of OBI. [47]

7.2 ONTOFOX

OntoFox is a web-based system to support ontology reuse. It allows users to input terms, fetch selected properties, annotations, and certain classes of related terms

[29]https://obi.svn.sourceforge.net/svnroot/obi/trunk/src/tools/old-build/create-external-derived.lisp

from source ontologies and save the results using the RDF/XML serialization of the OWL. OntoFox follows and expands the MIREOT principle. Inspired by existing ontology modularization techniques, OntoFox also develops a new SPARQL-based ontology term extraction algorithm that extracts terms related to a given set of signature terms. In addition, OntoFox provides an option to extract the hierarchy rooted at a specified ontology term [55].

7.2.1 OntoFox workflow

The input data is parsed internally by the OntoFox web server. SPARQL queries are then constructed and used to query remote RDF triple stores, containing the RDF triples of source ontologies. After successful query execution, an OWL output file is generated and provided to the user for download. Fig. 14 shows the Ontofox workflow.

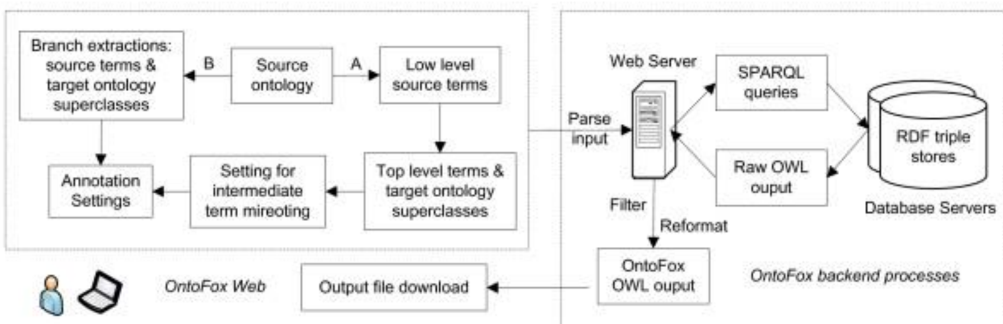

Figure 14: OntoFox Workflow [55].

7.3 BioPortal

The National Center for Biomedical Ontology (NCBO) has developed BioPortal, a web portal that provides access to a library of biomedical ontologies and terminologies (http://bioportal.bioontology.org) via the NCBO Web services. BioPortal enables community participation in the evaluation and evolution of ontology content by providing features to add mappings between terms, to add comments linked to specific ontology terms and to provide ontology reviews [56] [57] [58]. BioPortal allows researcher to:

- browse the library of ontologies
- search for a term across multiple ontologies

- browse mappings between terms in different ontologies
- receive recommendations on which ontologies are most relevant for a corpus
- annotate text with terms from ontologies
- search biomedical resources for a term
- browse a selection of projects that use BioPortal resources

7.4 Ontobee

A web server aimed to facilitate ontology visualization, query, and development. Ontobee provides a user-friendly web interface for displaying the details and its hierarchy of a specific ontology term. Meanwhile, Ontobee provides a RDF source code for the particular web page, which supports remote query of the ontology term and the Semantic Web [59].

7.5 Ontology Lookup Service

The Ontology Lookup Service[30] provides a web service interface to query multiple ontologies from a single location with a unified output format. The OLS can integrate any ontology available in the Open Biomedical Ontology (OBO) format.[60] Currently OLS has 84 ontologies and 1 950 654 terms.

7.6 Neuroscience Information Framework

The Neuroscience Information Framework (NIF) is a dynamic inventory of registered Web-based neuroscience resources containing data, materials, and tools. NIF maintains the largest searchable collection of neuroscience data, the largest catalog of biomedical resources, and the largest ontology for neuroscience on the web. It advances neuroscience research by enabling access to public research data and tools through an open source environment [61]. Currently over 6,000 resources are registered in the NIF. Data and metadata of individual research groups are contained in simple spreadsheets, a web page, or a queryable database. The NIF has many tools available to help to share data and metadata (e.g. Curation Interface CINDY for registering resources, LBGenTool6.1 for generating a DISCO file). The technical support and curators care about NIF portal management and assist with registration. Registration of neuroinformatics resource is divided into three levels [62]. I successfully registered EEG/ERP portal within NIF.

[30] http://www.ebi.ac.uk/ontology-lookup/

7.6.1 Level 1

At the lowest level,registering our source to the NIF requires providing NIF with the URL and basic information about the resource. This level places the resource in the NIF Registry, where it is available through direct query and through the NIF Web. The first step is to complete a registration form on the website NeuroLex, where basic information about registered resources is required (the user must be registered). The registration form includes a description of the source, source name, owner resources, cooperating entities, a grant title of the source, abbreviation, source URL, the link to the publication source, keywords and free text information about the source. The second step is to generate a sitemap which keeps our NIF Registry description up-to-date and inform search engines about the resource.The DISCO (Extensible Web resource DISCOvery, registration and interoperation framework) Resource Description (Sitemap) files have been created and "EEGbase" has been registered to the DISCO Dashboard (Fig. 15). Resource integration is managed by editing existing DISCO files to add more services. [62] [61] The last step is downloading DISCO XML file(disco.xml), which contains information about integration approach designed to facilitate interoperation among Internet resources, and downloading DISCO resource description (disco.rd.xml), which contains description of our resource. These two files "disco.rd.xml" and "disco.xml" are copied to the root directory of our web resource so that the DISCO crawler can find our resource. DISCO file consists of a set of tools and services that allow resource providers to share data and metadata with automated systems such as NIF. This step completes Level 1 registration. It provides information about the resource, but does not provide direct access to dynamic content or to the structure of the content. [62] [61]

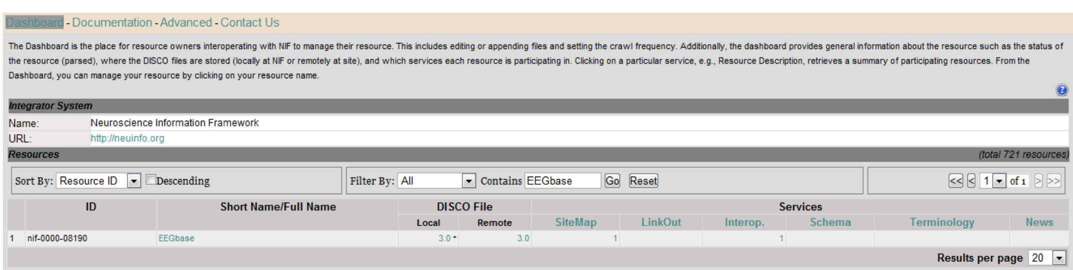

Figure 15: NIF - DISCO dashboard.

7.6.2 Level 2

Level 2 uses an XML-based script to provide a wrapper to a web site that allows NIF to search for key details about the web site and some information about dynamic content. The advantage of level 2 registration is that the content is dynamically updated from our source file, ensuring that all content is up-to-date.

7.6.3 Level 2.5

For resources registered with DISCO Web Interoperation, NIF developers may be able to create data views to extract LinkOut data for that resource. The Data Federation provides the ability to drill down into individual databases and data sets and returns relevant content. The NIF works with the following types of resources:

- Database with query API
- Database with web service
- Database dump
- XML data
- Structured web pages without API (e.g., HTML)
- Unstructured data files in several formats (Excel, CSV, PDF, etc.)
- RDF (Resource Description Framework) [61]

The NIF is then able to harvest information (e.g. experiments and scenarios) using implemented services every month and keep these sets of information up-to-date. Fig. 16 shows the UML class diagram of our solution. Our aim is to register EEG/ERP portal as a recognizable data source that can provide data in the CSV output; authors generate CSV files using implemented services within EEG/ERP portal. The input point of the harvesting mechanism is NIFMultiController called from the Web Browser. This controller transforms CSV files containing experiments or scenarios into the output stream returned to the Web Browser according to the input users request. The NIFMultiController using CSVFactory which gets the set of Plain Old Java Object (POJO) objects (experiments or scenarios) according to input users request and creates CSV file into the output stream and returns to the Web Browser. Our research group has obtained the registration at Level 2.5 which allows us to share our experiments and scenarios within the NIF. Privacy and security of electronic health information are guaranteed. [62]

7.6.4 Level 3

Level 3 integration utilizes a data integration framework to knit independently maintained databases into a virtual data federation through registration of

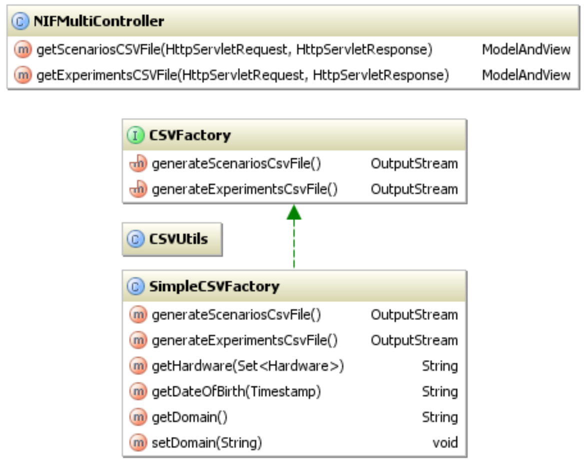

Figure 16: UML diagram - NIF services.

schema information and database views with the NIF mediator. A concept mapping tool is available to map tables, fields and values to the Neuroscience Information Framework Standard ontology (NIFSTD). This ontology includes a set of modular ontologies that provide a comprehensive collection of terminologies to describe neuroscience data and resources. Resource providers do not need to change their resource in any way and may control the content that is exposed to the NIF database mediator. Level 3 provides several advantages to resource providers. The first advantage, the mapping to the NIF vocabularies, provides the means to provide a standardized terminology and also to search through the relationships contained in the NIF ontologies. The second advantage, data within a source database can be combined with that from other databases by defining an integrated view across databases. We cooperate with the NIF mediator and their main goal is a full registration (Level 3) of the EEG/ERP portal as a recognizable data source within the NIF portal. [62] [61]

7.6.5 Registration overview

Level 1 registration was completed. It provides information about the resource, but does not provide direct access to dynamic content or to the structure of the content. The EEG/ERP portal was registered as a neuroscience resource within the NIF (Fig. 17). Fig. 17 shows resource "EEGbase" within the *Registry* tab which contains basic information about our resource.

Level 2 and Level 2.5 registration were completed to provide direct access to implemented services within EEG/ERP portal. Experiments and scenarios stored in the EEG/ERP portal are available also within NIF. Fig. 18 shows our exper-

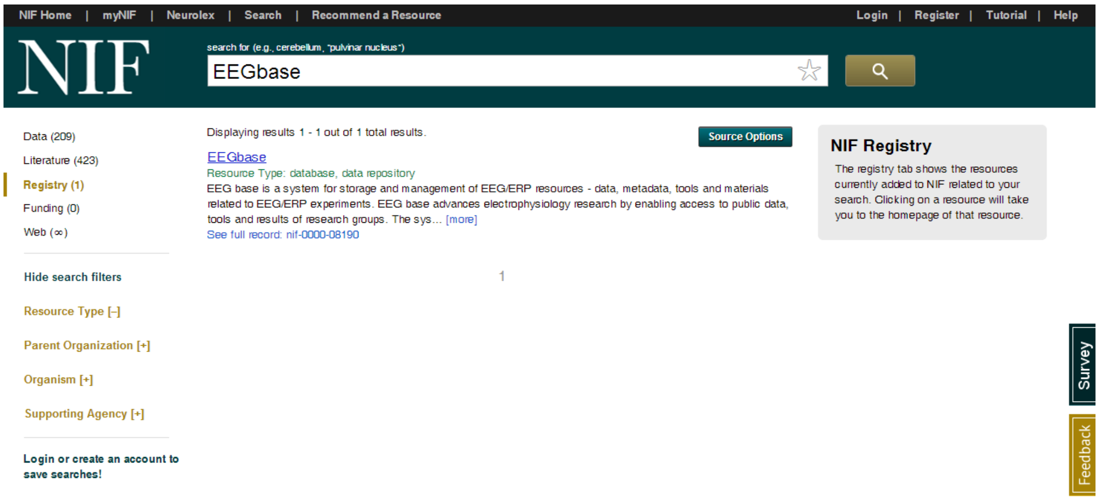

Figure 17: NIF Registry - EEG/ERP portal.

iments within the *Data* tab which contains experiments and scenarios. Privacy and security of electronic health information are guaranteed. All data and metadata are anonymous. All of the tested subjects signed that they agree with the conditions of the experiment and with the sharing of their EEG/ERP data. [62] The following data and metadata were shared within NIF:

- measured subject (gender, date of birth)
- experiment data
- used hardware
- used scenario (scenario title, description, length, source file)
- other parameters of the experiment (length of recording) [62]

7.7 Other tools

7.7.1 NeuroElectro

This project[31] deals with the organizing information on cellular neurophysiology. Specifically, project is focused on extracting information about the electrophysiological properties (e.g. resting membrane potentials and membrane time constants) of diverse neuron types from the existing literature and place it into

[31] http://neuroelectro.org/

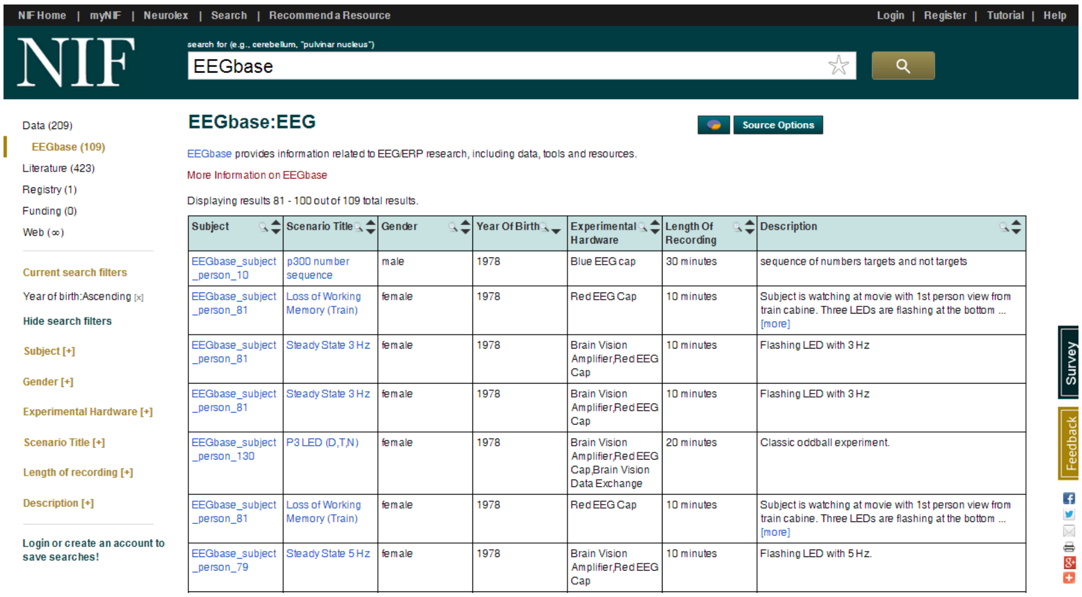

Figure 18: Data - EEG/ERP portal.

a centralized database. Project goal is to facilitate the discovery of neuron-to-neuron relationships and better understand the role of diversity across neuron types. The database and website are still in development [63].

7.7.2 OpenElectrophy

OpenElectrophy[32] is a python module for electrophysiology data analysis (intra- and extra-cellular). OpenElectrophy is build on top of Neo [64] :

- It includes the powerful Neo IO that can read a quantity of data formats (Plexon, NeuroExplorer, Spike2, TDT, Axon, BlackRock, . . .)
- neo object ready for analyses (AnalogSignal, SpikeTrain, RecordingChannel, Segment, Block, . . .)

But OpenElectrophy also provide:

- A GUI for exploring dataset.
- A complete off-line spike-sorting tool chain = GUI and/or command line.
- A time-frequency toolbox = fast wavelet scalogram plotting + transient oscillation in LFP detection.

[32] http://neuralensemble.org/OpenElectrophy/

- Viewers for neo objects.
- A database for storage. [64]

8 Conclusion

This work briefly describes the current Semantic Web (W3C standards, projects, ontologies), our research group, current biomedical ontologies, terminologies, data formats and tools. It proposes a complex summary of available ontologies and best practises of their creating. The main challenge of this work was to get acquainted with ontologies from biomedical environment and best ontology practices outlined by the OBO Foundry, including reuse of existing terms, ontologies, modularity, orthogonality, human-readable annotations and use of the BFO as an upper ontology and the OBI as a mid-level ontology. It is common to cooperate with community of researchers who design and create ontologies (EPhys ontology workgroup, NIF). This work brought the summary of current state of the electrophysiological domain (MINI, NEMO, EEG/ERP portal), innovations in strategies for combining ontologies (e.g. MIREOT), and the simple format for collecting and exchanging metadata (odML). The web-based systems providing access to a library of biomedical ontologies and terminologies in electrophysiology and neurophysiology (NIF, BioPortal, Ontobee, OntoFox and Ontology Lookup Service) are described in this work.

8.1 Current state in EPhys ontology workgroup

The device branch terminology is built upon existing ontologies related to neurophysiological experiments or investigation, namely the OBI and the NEMO. Existing terms have been imported using the MIREOT format and the web service Ontofox. The granularity of the ontology has been extended using the odML terminology and terms from the EEG/ERP portal. To test the ontology, the group created a simple knowledge base to describe the content of the EEG/ERP portal. [31]

8.2 Current state in the EEG/ERP portal

The EEG/ERP portal contains over 200 experiments. The experiment dealing with visual cortex of mice (Visual evoked potential measurement) is performed at the Department of Pathophysiology of the Faculty of Medicine using our devices (stimulator, BrainAmp amplifier), but the measured data and metadata are not stored in the EEG/ERP portal. It is necessary to define the terminology and to design and create ontology for this experiment. EEG/ERP data model is necessary to be updated within the EEG/ERP portal on the base of the proposed terminology and ontologies. In conclusion, the knowledge base has to be created for EEG/ERP experiments.

8.3 Aims of Ph.D. Thesis

- Define the terminology for EEG/ERP experiments (including the experiment dealing with Visual cortex of mice).
- Design and create ontologies for describing metadata for EEG/ERP experiments (including the experiment dealing with Visual cortex of mice).
- Create the knowledge base (RDF triple graph) for EEG/ERP experiments.
- Implement a process for obtaining metadata in the EEG/ERP laboratory and a storage for existing and proposed ontologies.
- Update the data model of the EEG/ERP portal using existing and proposed ontologies.
- Verify the proposed solution within the EEG/ERP portal.

References

[1] W3C. (2013, March) Semantic web. [Online]. Available: http://www.w3.org/standards/semanticweb/

[2] T. Segaran, C. Evans, and J. Taylor, *Programming the semantic web.* O'Reilly Media, 2009.

[3] E. Prud'hommeaux and A. Seaborne. (2008, January) Sparql query language for rdf. [Online]. Available: http://www.w3.org/TR/rdf-sparql-query/

[4] D. Brickley and R. Guha. (2004, February) Rdf vocabulary description language 1.0: Rdf schema. [Online]. Available: http://www.w3.org/TR/rdf-schema/

[5] R. W. Group. (2010, June) Rule interchange format (rif). [Online]. Available: http://www.w3.org/2001/sw/wiki/RIF

[6] T. R. Gruber, "A translation approach to portable ontology specifications," *KNOWLEDGE ACQUISITION*, vol. 5, pp. 199–220, 1993.

[7] W. N. Borst, "Construction of engineering ontologies for knowledge sharing and reuse," Ph.D. dissertation, Enschede, September 1997. [Online]. Available: http://doc.utwente.nl/17864/

[8] Y. L. Franc. (2012, March) Introduction to CNO. [Online]. Available: http://www.neuroml.org/files/NeuroML2012/YleFranc_CNO.pdf

[9] D. L. McGuinness and F. van Harmelen. (2004, February) Owl web ontology language overview. [Online]. Available: http://www.w3.org/TR/owl-features/

[10] V. Thomas, "Semantic Web," Ph.D. dissertation, VUT, Czech Technical University in Prague, December 2011, habilitation thesis. [Online]. Available: http://www.cvut.cz/pracoviste/odbor-rozvoje/stranky/habilitace-a-inaugurace/habilitacni-prednasky/lecture.pdf

[11] L. W. Lacy, *OWL : representing information using the Web Ontology Language*. Trafford Publishing, Jan. 2005. [Online]. Available: http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/1412034485

[12] W. O. W. Group. (2013, January) Owl 2 web ontology language document overview (second edition). [Online]. Available: http://www.w3.org/TR/owl2-overview/

[13] M. K. Bergman. (2013, May) "sweet tools". ai3; adaptive information, adaptive innovation, adaptive infrastructure. [Online]. Available: http://www.mkbergman.com/sweet-tools-simple-list/

[14] C. Bizer, J. Lehmann, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak, and S. Hellmann, "Dbpedia - a crystallization point for the web of data," *Web Semant.*, vol. 7, no. 3, pp. 154–165, Sep. 2009. [Online]. Available: http://dx.doi.org/10.1016/j.websem.2009.07.002

[15] D. Brickley and L. Miller. (2013, May) The friend of a friend project. [Online]. Available: http://www.foaf-project.org/

[16] A. Passant, U. Bojars, J. Breslin, and S. Decker, "The sioc project: Semantically-interlinked online communities, from humans to machines," in *Coordination, Organizations, Institutions and Norms in Agent Systems V*, ser. Lecture Notes in Computer Science, J. Padget, A. Artikis, W. Vasconcelos, K. Stathis, V. Silva, E. Matson, and A. Polleres, Eds. Springer Berlin Heidelberg, 2010, vol. 6069, pp. 179–194.

[17] J. Breslin, U. Bojars, A. Passant, S. Fernandez, and S. Decker, "Sioc: Content exchange and semantic interoperability between social networks," 2009.

[18] A. Doms and M. Schroeder, "Gopubmed: exploring pubmed with the gene ontology." *Nucleic Acids Res*, vol. 33, no. Web Server issue, pp. W783–6, 2005.

[19] I. Kupershmidt, Q. J. Su, A. Grewal, S. Sundaresh, I. Halperin, J. Flynn, M. Shekar, H. Wang, J. Park, W. Cui, G. D. Wall, R. Wisotzkey, S. Alag, S. Akhtari, and M. Ronaghi, "Ontology-Based Meta-Analysis of Global Collections of High-Throughput Public Data," *PLoS ONE*, vol. 5, no. 9, pp. e13 066+, Sep. 2010. [Online]. Available: http://dx.doi.org/10.1371/journal.pone.0013066

[20] T. D. Wang, B. Parsia, and J. Hendler, "A survey of the web ontology landscape," in *In Proc. of the International Semantic Web Conference, ISWC*, 2006.

[21] J. Hendler and the W3C Communications Team. (2008, May) Frequently asked questions on w3c's web ontology language. [Online]. Available: http://www.w3.org/2003/08/owlfaq

[22] P. L. Whetzel, N. H. Shah, N. F. Noy, B. Dai, M. Dorf, N. B. Griffith, C. Jonquet, C. H. Youn, C. Callendar, A. Coulet, D. L. Rubin, B. Smith, M.-A. Storey, C. G. Chute, and M. A. Musen, "Bioportal: Ontologies and integrated data resources at the click of the mouse," in *International Conference on Biomedical Ontology, ICBO'09*, 2009, p. 197. [Online]. Available:

http://www.lirmm.fr/ jonquet/publications/documents/PosterICBO09-NCBO.pdf

[23] F. Gibson, P. Overton, T. Smulders, S. Schultz, S. Eglen, C. Ingram, S. Panzeri, P. Bream, E. Sernagor, M. Cunningham, C. Adams, C. Echtermeyer, J. Simonotto, M. Kaiser, D. Swan, M. Fletcher, and P. Lord, "Minimum Information about a Neuroscience Investigation (MINI) Electrophysiology : Nature Precedings," *Nature Precedings*, Mar. 2008. [Online]. Available: http://precedings.nature.com/documents/1720/version/1

[24] A. S. Blum and S. B. Rutkove, *The clinical neurophysiology primer*. Springer, 2007, vol. 388.

[25] J. Štebeták, P. Bruha, and R. Moucek, "Neuroinformatics-data management and analytic tools for eeg/erp research," *Beyond AI: Interdisciplinary Aspects of Artificial Intelligence*, p. 91.

[26] T. W. Picton, O. G. Lins, and M. Scherg, "The recording and analysis of event-related potentials." in *Handbook of Neuropsychology*, F. Boller and J. Grafman, Eds. Amsterdam: Elsevier, 1995, vol. 10, pp. 3–73.

[27] S. Luck, *An introduction to the event-related potential technique*, ser. Cognitive neuroscience. MIT Press, 2005. [Online]. Available: http://books.google.cz/books?id=J_QgAQAAIAAJ

[28] P. Bruha, L. Vareka, and R. Moucek, "Sharing the eeg/erp experiments based on a modified odd-ball paradigm," may 2013.

[29] P. Jezek, J. Stebetak, P. Bruha, and R. Moucek, "Model of software and hardware infrastructure for electrophysiology," in *6th International Conference on Health Informatics*, 2013, pp. 1099–1103.

[30] J. Pergler, P. Borik, P. Jezek, R. Moucek, P. Bruha, and J. Stebetak. (2009, jul) Eeg/erp portal. [Online]. Available: http://eegdatabase.kiv.zcu.cz/home.html

[31] A. Brandowski, P. Bruha, J. Greweand, Y. Le Franc, R. Moucek, V. Papez, S. Tripathy, and T. Wachtler, "Oen, a formal ontology for describing experimental neurophysiology." june 2013.

[32] J. Grewe, T. Wachtler, and J. Benda, "A bottom-up approach to data annotation in neurophysiology," *Frontiers in Neuroinformatics*, vol. 5, no. 16, 2011. [Online]. Available: http://www.frontiersin.org/neuroinformatics/10.3389/fninf.2011.00016/abstract

[33] N. research team. (2013, May) Neural electromagnetic ontologies. [Online]. Available: http://nemo.nic.uoregon.edu/wiki/NEMO

[34] G. Frishkoff, J. Sydes, K. Mueller, R. Frank, T. Curran, J. Connolly, K. Kilborn, D. Molfese, C. Perfetti, and A. Malony, "Minimal information for neural electromagnetic ontologies (minemo): A standards-compliant method for analysis and integration of event-related potentials (erp) data," *Standards in genomic sciences*, vol. 5, no. 2, p. 211, 2011.

[35] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock, "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium." *Nature genetics*, vol. 25, no. 1, pp. 25–29, May 2000. [Online]. Available: http://dx.doi.org/10.1038/75556

[36] G. Gkoutos. (2010, oct) Phenotype and trait ontology. [Online]. Available: http://www.eie.gr/nhrf/institutes/ibrb/ontology-en.html

[37] B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L. J. Goldberg, K. Eilbeck, A. Ireland, C. J. Mungall, N. Leontis, P. Rocca-Serra, A. Ruttenberg, S.-A. Sansone, R. H. Scheuermann, N. Shah, P. L. Whetzel, and S. Lewis, *Nat Biotech*, no. 11, pp. 1251–1255, nov.

[38] B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L. J. Goldberg, K. Eilbeck, A. Ireland, C. J. Mungall *et al.*, "The obo foundry: coordinated evolution of ontologies to support biomedical data integration," *Nature biotechnology*, vol. 25, no. 11, pp. 1251–1255, 2007.

[39] M. E. Aranguren, E. Antezana, M. Kuiper, and R. Stevens, "Ontology design patterns for bio-ontologies: a case study on the cell cycle ontology," *BMC bioinformatics*, vol. 9, no. Suppl 5, p. S1, 2008.

[40] M. Egaña, A. Rector, R. Stevens, and E. Antezana, "Applying ontology design patterns in bio-ontologies," in *Knowledge Engineering: Practice and Patterns*. Springer, 2008, pp. 7–16.

[41] F. T. Imam, S. D. Larson, J. S. Grethe, A. Gupta, A. Bandrowski, and M. E. Martone, "Nifstd and neurolex: Comprehensive neuroscience ontology development based on multiple biomedical ontologies and community involvement," 2011.

[42] W. J. Bug, G. A. Ascoli, J. S. Grethe, A. Gupta, C. Fennema-Notestine, A. R. Laird, S. D. Larson, D. Rubin, G. M. Shepherd, J. A. Turner *et al.*, "The nifstd and birnlex vocabularies: building comprehensive ontologies for neuroscience," *Neuroinformatics*, vol. 6, no. 3, pp. 175–194, 2008.

[43] H. Stenzhorn. (2013, March) Basic formal ontology. [Online]. Available: http://www.ifomis.org/bfo

[44] B. Smith and A. Ruttenberg. (2012, July) Bfo 2.0. [Online]. Available: https://bfo.googlecode.com/svn/trunk/docs/BFO_Tutorial_Graz_2012-Smith.ppt

[45] R. Arp and B. Smith, "Function, role, and disposition in basic formal ontology," *Nature Preceedings*, pp. 1–4, 2008.

[46] Y. L. Franc. (2013, may) Computational neuroscience ontology. [Online]. Available: http://datahub.io/cs_CZ/dataset/bioportal-cno

[47] M. Courtot, F. Gibson, A. L. Lister, J. Malone, D. Schober, R. R. Brinkman, and A. Ruttenberg, "Mireot: The minimum information to reference an external ontology term," *Appl. Ontol.*, vol. 6, no. 1, pp. 23–33, Jan. 2011. [Online]. Available: http://dl.acm.org/citation.cfm?id=1971674.1971680

[48] M. Courtot, W. Bug, F. Gibson, A. L. Lister, J. Malone, D. Schober, R. Brinkman, and A. Ruttenberg, "The owl of biomedical investigations," in *OWLED Workshop on OWL: Experiences and Directions, collocated with the 7th International Semantic Web Conference (ISWC-2008)*, 2008.

[49] S. D. Larson and M. E. Martone, "Ontologies for neuroscience: what are they and what are they good for?" *Frontiers in neuroscience*, vol. 3, no. 1, p. 60, 2009.

[50] T. Wachtler. (2013, Jul.) German neuroinformatics node webpage. [Online]. Available: http://www.g-node.org/projects/odml

[51] P. J. Durka. (2012, March) Signalml. [Online]. Available: http://bci.fuw.edu.pl/wiki/SignalML

[52] P. Gleeson, S. Crook, R. C. Cannon, M. L. Hines, G. O. Billings, M. Farinella, T. M. Morse, A. P. Davison, S. Ray, U. S. Bhalla *et al.*, "Neuroml: a language for describing data driven models of neurons and networks with a high degree of biological detail," *PLoS computational biology*, vol. 6, no. 6, p. e1000815, 2010.

[53] T. H. Group. (2013, January) Hdf5 home page. [Online]. Available: http://www.hdfgroup.org/HDF5/

[54] A. E. Buleu, W. Benger, and M. S. Venkataraman, "An ontological scheme for specifying time in hdf5," in *Proceedings of The National Conference On Undergraduate Research (NCUR)*. Citeseer, 2007.

[55] Z. Xiang, M. Courtot, R. Brinkman, A. Ruttenberg, and Y. He, "OntoFox: web-based support for ontology reuse," *BMC Research Notes*, vol. 3, no. 1, pp. 175+, Jun. 2010. [Online]. Available: http://dx.doi.org/10.1186/1756-0500-3-175

[56] P. L. Whetzel, N. F. Noy, N. H. Shah, P. R. Alexander, C. Nyulas, T. Tudorache, and M. A. Musen, "Bioportal: enhanced functionality via new web services from the national center for biomedical ontology to access and use ontologies in software applications." *Nucleic Acids Research*, vol. 39, no. Web-Server-Issue, pp. 541–545, 2011. [Online]. Available: http://dblp.uni-trier.de/db/journals/nar/nar39.html#WhetzelNSANTM11

[57] M. Salvadores, M. Horridge, P. Alexander, R. Fergerson, M. Musen, and N. Noy, "Using sparql to query bioportal ontologies and metadata," in *The Semantic Web ISWC 2012*, ser. Lecture Notes in Computer Science, P. Cudr-Mauroux, J. Heflin, E. Sirin, T. Tudorache, J. Euzenat, M. Hauswirth, J. Parreira, J. Hendler, G. Schreiber, A. Bernstein, and E. Blomqvist, Eds. Springer Berlin Heidelberg, 2012, vol. 7650, pp. 180–195.

[58] M. A. Musen, N. F. Noy, N. H. Shah, P. L. Whetzel, C. G. Chute, M.-A. D. Storey, and B. Smith, "The national center for biomedical ontology." *JAMIA*, vol. 19, no. 2, pp. 190–195, 2012.

[59] Z. Xiang, C. Mungall, A. Ruttenberg, and Y. He, "Ontobee: A linked data server and browser for ontology terms." in *ICBO*, ser. CEUR Workshop Proceedings, O. Bodenreider, M. E. Martone, and A. Ruttenberg, Eds., vol. 833. CEUR-WS.org, 2011. [Online]. Available: http://dblp.uni-trier.de/db/conf/icbo/icbo2011.html#XiangMRH11

[60] R. Côté, F. Reisinger, L. Martens, H. Barsnes, J. A. Vizcaino, and H. Hermjakob, "The ontology lookup service: bigger and better," *Nucleic acids research*, vol. 38, no. suppl 2, pp. W155–W160, 2010.

[61] D. Gardner, H. Akil, G. A. Ascoli, D. M. Bowden, W. J. Bug, D. E. Donohue, D. H. Goldberg, B. Grafstein, J. S. Grethe, A. Gupta, M. Halavi, D. N. Kennedy, L. N. Marenco, M. E. Martone, P. L. Miller, H.-M. Mller, A. Robert, G. M. Shepherd, P. W. Sternberg, D. C. V. Essen, and R. W. Williams, "The neuroscience information framework: A data and knowledge environment for neuroscience." *Neuroinformatics*, vol. 6, no. 3, pp. 149–160, 2008. [Online]. Available: http://dblp.uni-trier.de/db/journals/ni/ni6.html#GardnerAABBDGGGGHKMMMMRSSEW08

[62] P. Bruha and R. Moucek, "Portal for research in electrophysiology - data integration with neuroscience information framework," in *Biomedical Engineering and Informatics (BMEI), 2012 5th International Conference on*, 2012, pp. 1099–1103.

[63] R. G. Shreejoy Tripathy. (2012, January) Neuroelectro: organizing information on cellular neurophysiology. [Online]. Available: http://neuroelectro.org/

[64] N.-T. S.Garcia. (2012, January) Openelectrophys documentation. [Online]. Available: http://pythonhosted.org/OpenElectrophy/

[65] A. Brazma, P. Hingamp, J. Quackenbush, G. Sherlock, P. Spellman, C. Stoeckert, J. Aach, W. Ansorge, C. A. Ball, H. C. Causton *et al.*, "Minimum information about a microarray experiment (miame)toward standards for microarray data," *Nature genetics*, vol. 29, no. 4, pp. 365–371, 2001.

[66] P. Jezek, "Ontology development in eeg/erp domain," Ph.D. dissertation, University of West Bohemia, Faculty of Applied Sciences, Univerzitni 22, 306 14 Pilsen, March 2012.

List of Figures

9 List of Abbreviations

Abbreviation	Explanation of the abbreviation
BFO	Basic Formal Ontology
CNO	Computational Neuroscience Ontology
CSV	A comma separated values
DISCO	Extensible Web resource DISCOvery
EEG	Electroencephalography
ERP	Even-Related Potentials
FMA	Foundational Model of Anatomy
fMRI	Functional Magnetic Resonance Imaging
FOAF	Friend of a Friend
GO	Gene Ontology
GRDDL	Gleaning Resource Descriptions from Dialects of Languages
GUI	Graphical User Interface
HDF	Hierarchical Data Format
ChEBI	Chemical Entities of Biological Interest
IAO	Information Artifact Ontology
INCF	International Neuroinformatics Coordinating Facilities
MEG	Magneto Encephalography
MeSH	Medical Subject Headings
NCBO	National Center for Biomedical Ontology
NEMO	Neural ElectroMagnetic Ontologies
NIF	Neuroinformatics Information Framework
NIFSTD	Neuroinformatics Information Framework Standard ontology
MINEMO	Minimal Information for Neural ElectroMagnetic Ontologies
MINI	Minimum Information about a Neuroscience Investigation
MIREOT	Minimal Information to Reference External Ontology Terms
OBI	Ontology for Biomedical Investigation
OBO	The Open Biomedical Ontologies
odML	open metadata Markup Language
OEN	Ontology for describing Experimental Neurophysiology
OWL	Ontology Web Language
PATO	Phenotypic Quality Ontology
PET	Positron-Emission Tomography
POJO	Plain Old Java Object
POWDER	The Protocol for Web Description Resources
RDF	Resource Description Framework
RDFa	Resource Description Framework in attributes
RDFS	Resource Description Framework Schema
RIF	Rule Interchange Format
R2RML	RDB to RDF Mapping Language
SIOC	Semantically-Interlinked Online Communities
SQL	Structure Query Language

Abbreviation	**Explanation of the abbreviation**
SPARQL	SPARQL Protocol and RDF Query Language(Recursive acronym)
UML	Unified Modeling Language
URI	Unified Resource Identifier
W3C	World Wide Web Consortium
XML	eXtensible Markup Language

A Appendix - Minimum Information about a Neuroscience Investigation for Electrophysiology

The checklist is completely overtaken from [23].

1. **General features**

   - *Data and time* - The date and time on which the work described was initiated given in the ISO:8601 representation. YYYY-MM-DDThh:mm:ss
   - *Responsible person or role* - The (stable) primary contact person for this data set; this could be the experimenter, lab head, line manager etc. Where responsibility rests with an institutional role (e.g. one of a number of duty officers) rather than a person, give the official name of the role rather than any one person. In all cases give affiliation and stable contact information, which consists of (i) Name, (ii) Postal address and (iii) Email address.
   - *Experimental context* - The name of the project, study or wider investigation of which the "experiment" is a part (if appropriate).
   - *Electrophysiology type* - The type of electrophysiology recording reported as "extra cellular" or "intra cellular".

2. **Study subject** - derived from Minimum information about a microarray experiment (MIAME) 1.1 [65].

   - *Genus* - The genus classification of the study subject according to the NCBI taxonomy classification.
   - *Species* - The species classification of the study subject according to the NCBI taxonomy classification.
   - *Strain* - The strain, genetic variant classification of the study subject, if appropriate.
   - *Cell line* - The identifier for the immortalised cell line, if appropriate.
   - *Genetic characteristic* - The genotype of the study stubject. Genetics characteristics include polymorphisms, disease alleles and haplotypes.
   - *Genetic variation* - The genetic modification introduced in addition to strain, if appropriate.
   - *Disease state* - The name of the pathology diagnosed in the subject. The disease state is "normal" if no disease state has been diagnosed.

- *Clinical information* - A link, summary or reference to additional clinical information, if appropriate.
- *Sex* - The sex of the subject, in terms of either male, female or hermaphrodite.
- *Age* - The time period elapsed since an identifiable point in the life cycle of an organism. If a developmental stage is specified the identifiable point would be the beginning of that stage. Otherwise the identifiable point must be specified.
- *Development stage* - The developmental stage of the study subjects life cycle.
- *Subject label* - If the subject has been chemically labeled or stained; state the label name.
- *Subject identifier* - The type and value of the identifier assigned to the subject.
  - *Type* - The type and value of the identifier assigned to the subject. For example, vendor or patient identifier. For patients, the identifier must be approved by an Institutional Review Board or appropriate body.
  - *Value* - The unique string which corresponds to the identifier type.
- *Associated subject details* - The organisation (e.g vendor) or individual responsible for the subject.
- *Preparation protocol* - The surgical procedure or the preparation protocol implemented to obtain the specific sample for recording.
- *Preparation date* - The date the surgical procedure or the preparation protocol was performed to obtain the specific sample for recording. Given in the ISO:8601 representation. YYYY-MM-DDThh:mm:ss

3. **Recording Location**

- *Location structure* - The anatomical part or structure of the subject under investigation or recorded from. For example brain or cell culture.
- *Brain area* - If the anatomical structure under study from 3.(Location structure) is the brain then state the location. If the anatomical structure under study is the mammalian brain then state the location using Neuronames. (http://braininfo.rprc.washington.edu/aboutfolder/aboutbi.html).
- *Slice thickness* - The thickness of the recording slice in millimeters.
- *Slice orientation* - State the planes of the slice, in terms of either i) coronal (width ways), ii) saggital (lengthways parallel to midline) or iii) tangential (lengthways perpendicular to midline).

- *Cell type*
  - *Target cell type* - The cell type of the anatomical structure given in 3. (Location structure) under investigation if non mixed. If mixed the target cell type should be provided.
  - *Confirmed cell type* - Reported as "anatomy confirmation","estimation" or chemical label (this includes antibodies and staining). Additional information such as recordings or image files which also confirm the location can be referenced here.

4. **Task** - *if appropriate*

   - *Protocol* - A description of the task protocol undertaken by the subject.
   - *Sensory conditions* - The sensory conditions during the task protocol.
   - *Equipment* - The Model Name, Model Number and Manufacturer for equipment used in the task protocol.
   - *Recording* - If the task is recorded state how and what data types are being recorded.

5. **Stimulus** - *if appropriate*

   - *Protocol* - A description of the stimulus protocol undertaken by the subject.
   - *Sensory conditions* - The sensory conditions during the stimulus protocol.
   - *Solutions* - Description of the solutions used in terms of name, components with concentrations,(if appropriate).
   - *Equipment* - The Model Name, Model Number and Manufacturer for specialised equipment used during the stimulus protocol.
   - *Recording* - If the stimulus is recorded state how and what data types are being recorded.

6. **Behavioral event** - *if appropriate*

   - *Event* - A description of the behavioural event observed.
   - *Equipment* - The equipment use to record the behavioural event, if recorded in terms of The Model Name, Model Number and Manufacturer.
   - *Recording* - The type of recording of the behavioural event, the file format and the format encoding.

7. **Recording**

- *Protocol* - A description of the recording protocol.
- *Conditions* - The subject conditions during the recording. Invivo or invitro preparation If invivo was it anathematised or awake? If awake what was the stimulus condition.
- *Containing device* - Containing device temperature of the subject or sample (for example, a bath): Include temperature if appropriate.
- *Solutions* - Description of the solutions used in terms of name, components with concentrations,(if appropriate).
- *Solution flow speed* - The flow speed of the solution described in terms of ml/min.
- *Equipment*
  - *Electrode* - The type of electrode and the Model Name, Model Number and Manufacturer for specialised equipment.
  - *Electrode configuration* - The configuration or arrangement of the electrode. For example, a 2-dimensional array. Also state the distance between each electrode. If the study uses voltage clamp in a patch configuration, state the access resistance (the resistance of the cell membrane, which is in series with the electrode resistance).
  - *Electrode impedance* - The electrode range or impedance of the electrode.
  - *Amplifier* - The Model Name, Model Number and Manufacturer of the amplifier.
  - *Filter* - The Model Name, Model Number and Manufacturer of the filter.
  - *Filter setting* - The settings or the parameters of the filter.
  - *Recorder* - The Model Name, Model Number and Manufacturer of the recorder.

8. **Time series data**

- *Data format* - The name of the data format of the time series data and specific encoding. For example, ASCII or binary encoding.
- *Sampling Rate* - The sampling rate of the recording.
- *File location* - The time series file location should be made available when the experiment is published, for example, using a Uniform Resource Identifier (URI) or a Digital Object Identifier (DOI). (Note this will be achieved automatically via submission to the CARMEN system).

B Appendix - EEGbase terminology

1. **Experiment** - When the experiment may be affected by surrounding conditions experimenters need to describe an environment and conditions where the experiment takes place. It includes weather, time of the day, environment note and temperature. [66]

    - *Start time*
    - *End time*
    - *Temperature*
    - *Environment note*
    - *Weather*

2. **Analysis** - The signal is analyzed using various techniques. The most often technique is averaging. The ontology describes the technique of analyzing the EEG/ERP signal. It includes the length of the pre- and post-stimuli part of the signal, the number of epochs or the verbal description of signal processing procedure. [66]

    - *Epoch number*
    - *Prestimulus time*
    - *Poststimulus time*

3. **Digitization** - Before the data from electrodes are stored, they are digitalized using an analogue-digital converter. This conversion is influenced by the set of parameters as filtration, sampling frequency and band-pass. The conditions of such digitalization process are described using the ontology as well. [66]

    - *Gain*
    - *Filter*
    - *Sampling rate*

4. **Electrode** - The brain activity is measured by the set of electrodes putted on the tested subject scalp. The proposed ontology describes their type, impedance, location, the used system and their fixation. [66]

    - *Impedance*
    - *System description image*
    - *Fixation*
    - *Location*

- *Type*

5. **Stimulus** - The experiment contains a set of stimuli. The stimuli types are defined for each scenario. [66]

   - *Type*

6. **Scenario** - This scenario includes information about video, pictures on the computer screen, or sound sloughed off into the headphone, or information about the instructions that the tested subject received when the experiment started. [66]

   - *Title*
   - *Description*
   - *Scenario length*
   - *MIME type*

7. **Person (Tested subject, Experimenter)** - Information about the tested subject or experimenter are stored (Laterality - left or right handed, Education, Date of birth, Gender, Diseases, Drugs, Optional note). [66]

   - *Given name*
   - *Surname*
   - *Date of birth*
   - *Gender*
   - *Email*
   - *Phone number*
   - *Person Note*
   - *Authority*
   - *Disease*
   - *Education level*
   - *Laterality*
   - *Note*

8. **Hardware (Amplifier, EEG cap)** - During the experiments various hardware equipment may be used. The description of the used hardware should be stored. We propose to store the type, producer and the serial number of the used hardware. [66]

   - *Title*
   - *Type*
   - *Description*

9. **Software (BrainVision Recorder 1.2, Presentation 16.3)** - The experiment is usually performed using supporting software equipment. This software includes software for running an experimental scenario or software for digitalizing data from electrodes. [66]