



VirtualBrainCloud

Personalized Recommendations for Neurodegenerative Disease



www.VirtualBrainCloud-2020.eu

Public Deliverable Report

M18 deliverable D3.13 (formerly D3.5.1):

“Complete, updated semantic framework for neurodegeneration research documented”.

Date	May 2020
Authors	Stephan Gebel, Sumit Madan, Stephan Springstube (Fraunhofer) Petra Ritter (CHARITE) © VirtualBrainCloud consortium
Dissemination level	public
Website	www.VirtualBrainCloud-2020.eu



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826421



Table of content

1. Preface.....	3
2. Introduction.....	4
3. Referential Ontology Hub for Applications within Neurosciences (ROHAN)	5
3.1. Description of ROHAN service	5
3.2. Clinical Trial Ontology.....	7
3.3. Excerpt of ontology curation guideline	9
3.4. Harmonization of Alzheimer’s disease clinical trial variables with extCTO.....	10
4. SCAIView-Neuro, a Semantic Search Engine	10
5. NeuroImaging Feature Terminology (NIFT).....	14
6. NeuroImaging Data Model (NIDM)	15
7. Brain Imaging Data Structure (BIDS)	17
8. Minimal Information Neuroscience Data Set (MINDS)	18
9. Conclusions.....	19
10. Glossary	20



1. Preface

A key objective of Task 3.4 “Workflows for clinical data curation and processing” and Task 3.5 “A metadata framework for unified metadata annotations and Data Catalogues” within the wider range of the Work Package 3 “Data Processing, Standardization and Data-FAIRification” is to deliver a semantic framework for neurodegenerative diseases, that serves as a central resource for controlled vocabularies and shared ontologies to access and use them within different TVB-Cloud work packages.

Within TVB-Cloud two systems will be implemented and made available for the consortium members:

- The Ontology Lookup Service (OLS) from Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), and
- the CTS2 system from University of Genova (UNIGE).

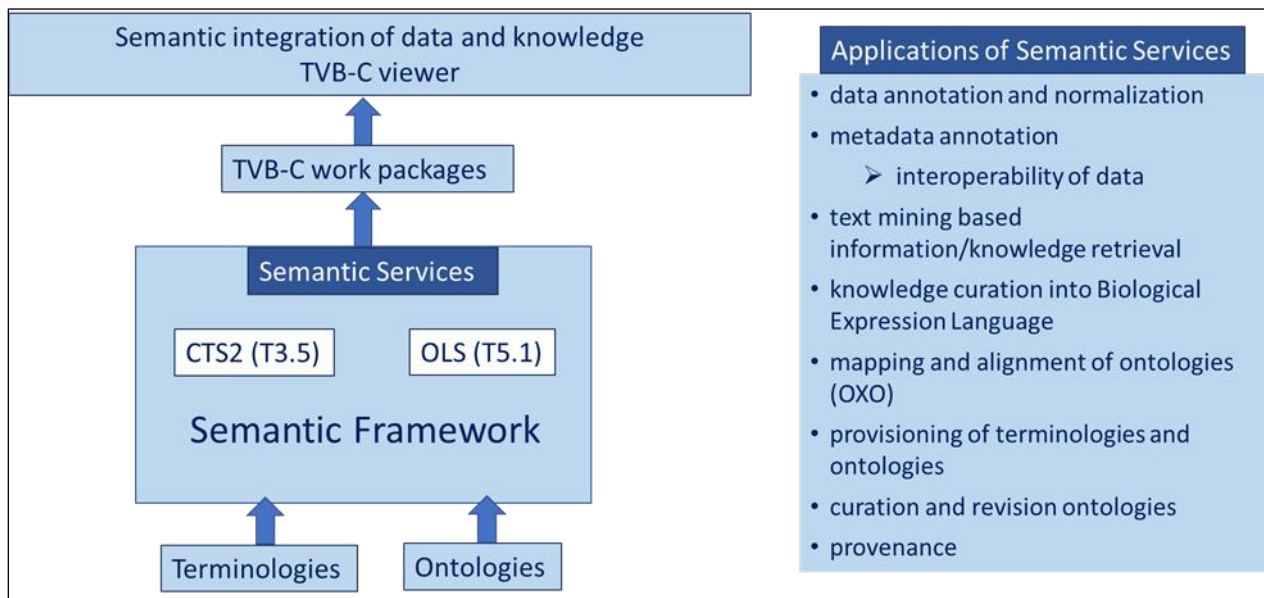


Figure 1: Schema of the Semantic Framework

The two deliverables share the same premise, but one was mainly developed by Fraunhofer SCAI and the other by the University of Genova. With these two deliverables two demonstrators of the websites are made available at the following addresses:

- **OLS**, the ontology lookup service software is being used to host the Referential Ontology Hub for Applications within Neurosciences (ROHAN), which is accessible at <https://rohan.scai.fraunhofer.de/ols/index>.
- **CTS2** compliant interface: The read and query part are open to everyone, while the access to the maintenance part is limited by a login. This access can be given to any member that wants to make a trial. The address is: http://www.medinfo.dibris.unige.it/VBC_CTS2 (for more information see delivery D3.14: “Integrated, open source terminology versioning and management tool CTS2 compliant application shared with partners”).



2. Introduction

With the TheVirtualBrain Cloud(TVB-Cloud) project, we are building a reference infrastructure for sharing and processing Health and Biomedical research data, specifically data in the domain of neurodegenerative disease (NDD). To enable users to use the data stored in TVB-Cloud platform, the data need to be annotated – that is they receive labels with a certain meaning. The labels are the metadata.

Metadata is the data that provides information about the data contained in the infrastructure. Metadata contain the information that can be searched and queried by users to find out what data are available, how they can be accessed and what data formats can be expected. A metadata standard provides a common understanding of the meaning or semantics of the data, to ensure correct and proper use and interpretation of the data. Metadata is "data about data." Many distinct types of metadata exist and there are specialized and well-accepted models to specify types of metadata. E.g. one differentiates between descriptive, structural, and administrative metadata – but also reference metadata and statistical metadata. Standardized concepts to assemble metadata – so called metadata schemes comprise: metadata standards and metadata models. Metadata standard refers to a common understanding of the meaning or semantics of the data. The metadata model defines the relations between classes of meta data.

Terminology refers to a group of specialized words or meanings relating to a particular domain, e.g. NDD. Ontologies define formal naming and definition of classes and the relations between the classes for a given domain.

For the type of data that is collected in TVB-Cloud, i.e. domain specific health and biomedical research data those standards for metadata and metadata models do not exist to the full extent that would be required. Tools such as controlled vocabularies, taxonomies, thesauri, data dictionaries, and metadata registries can be used to apply standardization to the metadata. So far, such controlled vocabularies in the domain of brain imaging and neurodegenerative disease are incomplete and therefore, as part of TVB-Cloud we are advancing existing terminologies and ontologies.

The major effort of this WP is related to the updating of the ontologies that will be used in semantic framework within TVB-Cloud. These ontologies comprise the majority of relevant entities and their relationships that are required to describe and harmonize data and knowledge about neurodegenerative diseases. The newly developed or revised ontologies are included in an ontology store that provides end-users the browsing and searching ability. Concepts and terms from the integrated ontologies will be used for metadata annotation and mapping of variables thereby enabling interoperability of data from clinical trials (see Task 4.1 “Characterization of clinical data sources for modelling”). In addition, a dedicated literature mining environment for brain research is part of the semantic framework.



3. Referential Ontology Hub for Applications within Neurosciences (ROHAN)

3.1. Description of ROHAN service

The ROHAN service is based on the software ‘Ontology Lookup Service’¹ (OLS) that is developed by European Bioinformatics Institute (EBI) for exploring ontologies. This hub is a repository for biomedical resources that aims to provide a single point of access to the latest ontology and terminology versions for translational neurodegeneration disease research. ROHAN is developed in the context of Medical Informatics Platform of Human Brain Project and is being extended in TVB-Cloud (Figure 2). The service provides a web-based user interface for exploring and visualizing the ontologies (Figure 3), and additionally, a flexible RESTful API to access the resources programmatically. It also provides an utility that allows to regularly update ontologies with ease. Furthermore, the service includes a search index for terms and synonyms with autocomplete functionality (Figure 4). To manage ontologies, the service also uses a flexible configuration system.

The ROHAN service is available at <http://rohan.scai.fraunhofer.de/ols>.

Referential Ontology Hub for Applications within Neurosciences (ROHAN)

Home Resources About Contact

Welcome to the ROHAN Service

Search the Terminology Service... Search!

Examples: alzheimer Looking for a particular resource?

Data Content
Updated 25 Mai 2020 15:32

- 14 ontologies and terminologies
- 56,984 terms
- 2,306 properties
- 1,204 individuals
- Version 3.2.1-SNAPSHOT

About:

This service is a repository for biomedical resources that aims to provide a single point of access to the latest ontology and terminology versions for translational neurodegeneration disease research in the context of [Human Brain Project](#). You can browse the resources through the website as well as programmatically via our API. This service is part of the [Medical Informatics Platform](#) of the [Human Brain Project](#). It is developed and maintained by [Fraunhofer Institute SCAI](#).

About the service: [Terms and Conditions](#), [Contact](#), [Publishing notes](#), [Data Protection](#)

Maintainers: **Fraunhofer SCAI**

Project Partners: **CHUV**

Funding: **Human Brain Project**
This project has received funding from the Human Brain Project.

This service is based on the [Ontology Lookup Service \(OLS\)](#).
[Learn about OLS](#)
[View OLS on GitHub](#)

Figure 2: Referential Ontology Hub for Applications within Neurosciences (ROHAN)

¹ Source code is available at <https://github.com/EBISPOT/OLS>



The screenshot shows the OLS Clinical Trials ontology interface. At the top, there are navigation tabs for 'Home', 'Resources', and 'About', along with a 'Contact' button. Below the navigation, the page title is 'Early Phase 1 clinical trial' with a search bar and a 'Search CTO' button. The main content area is divided into two sections: 'Tree view' and 'Term info'. The 'Tree view' section shows a hierarchical tree of concepts, with 'Early Phase 1 clinical trial' selected and highlighted in blue. The 'Term info' section provides details for the selected term, including its label, alternative terms, and a definition.

Label
Early Phase 1 clinical trial

alternative term
Pre-Clinical Trial, Phase 0 trial, Phase 0 clinical trial

definition
Exploratory trials, involving very limited human exposure, with no therapeutic or diagnostic intent (e.g., screening studies, microdose studies). (Formerly listed as "Phase 0"), A clinical trial that is at Early Phase I or Phase 0, which is designed to use an investigational agent that is available only in very limited quantities and which has never previously given to humans or for which there is extremely limited human experience. Phase 0 clinical trials are intended to enable researchers to understand the path of the drug in the body and its efficacy. Adverse event reporting in Phase 0 trials is expedited. [def-source: NCI]

Figure 3 Screenshot of a concept page from OLS Clinical Trials ontology

The screenshot shows the ROHAN autocomplete functionality. A search bar contains the text 'alzheimer'. Below the search bar, a list of suggestions is displayed, including 'Alzheimer disease 1', 'Alzheimer disease 18', 'Alzheimer disease 2', 'Alzheimer disease', and 'Alzheimer disease 4'. A red arrow points from a yellow 'Autocompletion' box to the first suggestion, 'Alzheimer disease 1'. Below the list, there is a 'Jump to' section with a table of suggestions and their corresponding DOIDs.

Jump to	
Alzheimer's disease 3	DOID:0110042
Alzheimer's disease 2	DOID:0110035
Alzheimer's disease 19	DOID:0110051
Alzheimer's disease 10	DOID:0110043
Alzheimer's disease 8	DOID:0110041

Search OLS for alzheimer

Figure 4: Screenshot of the autocomplete functionality of ROHAN

By May 20, 2020 the repository contains 14 terminologies and ontologies. Figure 5 shows a screenshot of the resources page of ROHAN service that lists the ontologies. These ontologies comprise the majority of relevant entities and their relationships that are required to describe and harmonize data and knowledge about neurodegenerative diseases. Updating and revision of the ontologies is an ongoing process within WP3 and WP5. In the next section, we describe the updating of the Clinical Trial Ontology. Other ontologies will be updated accordingly, if necessary.



Referential Ontology Hub for Applications within Neurosciences (ROHAN)
Human Brain Project

Home Resources About Contact

List of all resources in the Terminology Service

Show 10 entries Search:

Resource Name	Short name	Description	Loaded	Action
Parkinson's Disease Ontology	PDON	The Parkinson's disease ontology (PDO) as a comprehensive semantic framework covering the whole breadth of the Parkinson's knowledge domain. This ontology with a subclass-based taxonomic hierarchy not only covers the broad spectrum of major biomedical concepts from molecular to clinical features of the disease, but also the different views on disease features held by molecular biologists, clinicians and drug developers. This resource has been created for use in the IMI-funded AETIONOMY project (www.aetionomy.org).	Mon Mar 02 14:21:11 CET 2020	Search Terms Properties Individuals Download
Pathway Terminology System	PTS	Pathway terminology systems integrates biological Events and various Pathway types.	Mon Mar 02 14:21:45 CET 2020	Search Terms Properties Individuals Download
Schizophrenia Ontology	SCHIZO	The SCAI Schizophrenia Ontology. This is a draft	Mon Mar 02 14:58:52 CET 2020	Search Terms Properties Individuals Download
The core Clinical Trial Ontology (CTO)	CTO	The core Ontology of Clinical Trials (CTO) will serve as a structured resource integrating basic terms and concepts in the context of clinical trials. Thereby covering clinicaltrials.gov . CoreCTO will serve as a basic ontology to generate extended versions for specific applications such as annotation of variables in study documents from clinical trials.	Mon May 25 15:32:37 CEST 2020	Search Terms Properties Individuals Download

Showing 11 to 14 of 14 entries Previous 1 2 Next

Figure 5: Screenshot of terminologies and ontologies integrated into the ontology hub².

3.2. Clinical Trial Ontology

To model the field of clinical trials, the Clinical Trial Ontology for Neurodegeneration Diseases (CTO-NDD) was developed for the IMI-funded AETIONOMY³ project. During the TVB-Cloud project, this ontology has been updated. The overall structure of the updated ontology is based on community-agreed standards, like basic formal ontology (BFO) and, in addition, the best practices as well as the principles of Open Biological and Biomedical Ontology (OBO) Foundry are applied. The development of the ontology is undertaken by using Protégé OWL editor⁴ in ontology web language (OWL) format.

The updated version contains now two ontologies, a core Clinical Trials Ontology (coreCTO) and an extended Clinical Trials Ontology (extCTO). The coreCTO will serve as a structured resource integrating basic terms and concepts in the context of all clinical trials (Figure 6 and Figure 7). Therefore, terms and concepts from ClinicalTrials.gov as dedicated data provider to the WHO International Clinical Trials Registry Platform (ICTRP) were integrated⁵. To ensure alignment with WHO and International Committee of Medical Journal Editors (ICMJE) recommendations, terms outlined in WHO ICTRP were considered⁶. The coreCTO is being developed together with external partners from NCBI, FDA and University of Michigan. Whereas, the extCTO will be used to support implementation of specific

² <https://rohan.scai.fraunhofer.de/ols/ontologies>

³ <https://www.aetionomy.eu/>

⁴ <https://protege.stanford.edu/>

⁵ <https://prsinfo.clinicaltrials.gov/definitions.html>

⁶ <https://www.who.int/ictrp/network/trds/en/>



applications such as annotation of variables in documents derived from neurodegeneration disease clinical trials and further text mining use cases.

An application case for this coreCTO has been developed, the related manuscript is in preparation⁷. The current versions of both ontologies are available in ROHAN service^{8,9}.

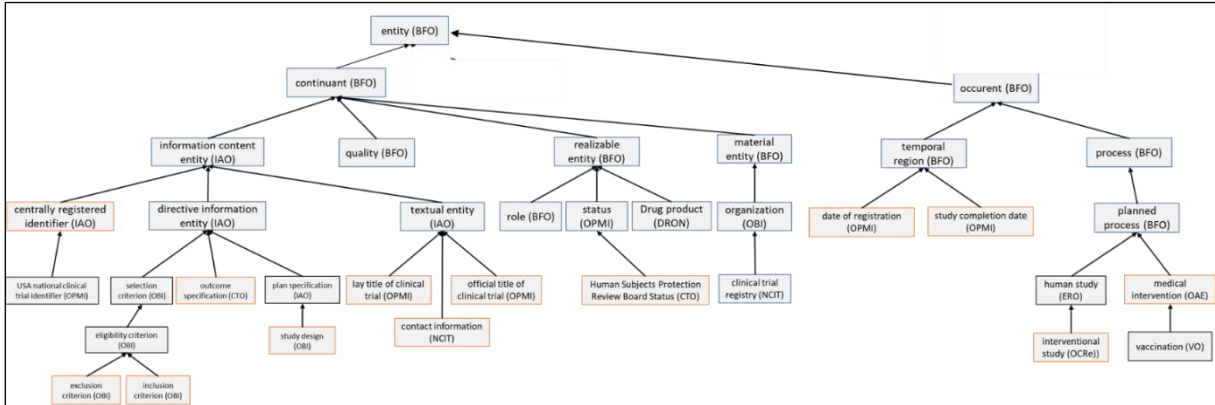


Figure 6: Top hierarchical architecture of coreCTO aligned to basic formal ontology (BFO). Concepts that are defined in WHO ICTRP are framed in red. Name of the OBO ontologies of reused concepts are given in brackets after the concept name.

⁷ Lin, A., Gebel, S. et al. CTO: a community-based Clinical Trial Ontology and Its Applications in PubChemRDF and SCAIView. 2020 (submitted)

⁸ coreCTO version: <https://rohan.scai.fraunhofer.de/ols/ontologies/cto>

⁹ extCTO version: <https://rohan.scai.fraunhofer.de/ols/ontologies/ecto>



The screenshot displays the Protégé editor interface for the Clinical Trials Ontology (CTO). The left pane shows a class hierarchy starting from 'owl:Thing' and descending through 'entity', 'continuant', 'generically dependent continuant', 'information content entity', 'centrally registered identifier', 'condition', 'Study Condition', 'Condition or Disease under Study', 'data item', 'clinical finding', 'clinical history', 'data about an ontology part', 'curation status specification', 'diagnosis', 'number of participants', 'preclinical finding', 'study result', 'outcome measure result', 'Primary Outcome Measure', 'Secondary Outcome Measure', 'directive information entity', 'action specification', 'objective specification', 'clinical trial primary purpose specification', 'outcome specification', 'other outcome measure information', 'primary outcome measure information', 'secondary outcome measure information', 'plan specification', 'protocol', 'Study Protocol', 'Clinical Study Protocol', 'study design', 'factorial design', 'intervention design', and 'Allocation'. The 'Primary Outcome Measure' class is highlighted. The right pane shows the 'Annotations: Primary Outcome Measure' section, which includes a 'label' annotation with the value 'Primary Outcome Measure', a 'definition' annotation with a detailed text description, and several 'has_exact_synonym' annotations with values like 'OUTMSPRI', 'PRIMARY OUTCOME MEASURE', and 'Primary Outcome Measure'. Below the annotations, there is a 'Description: Primary Outcome Measure' section and a 'SubClass Of' section showing that 'Primary Outcome Measure' is a subclass of 'study result'.

Figure 7: Screenshot of the core Clinical Trials Ontology showing hierarchal structure and annotations of the concepts (displayed with Protégé editor)

3.3. Excerpt of ontology curation guideline

The scientific community defines several standards and best practices to create and develop ontologies. Open Biological and Biomedical Ontology (OBO) Foundry has defined in the past principles to “develop interoperable ontologies, that are both logically well-formed and scientifically accurate”. Hence, where possible, we have used the best practices. That includes the usage of terms that were already described in an existing OBO conform ontology (Figure 4). Furthermore, to make the curation reproducible and transparent, several annotations have been added to the ontologies. For example:

- `rdfs:label` annotation was added for each concept as exactly same as the concept name.
- `oboInOwl:hasDefinition` property contains the definitions for a concept, which is mandatory for all concepts.
- `rdfs:isDefinedBy` contains a reference to the source of the definition, if the definition couldn't be imported from an existing ontology.
- `importedFrom` annotation is used if a concept is reused from another ontology.
- `rdfs:seeAlso` is used to capture any additional relevant references.
- `oboInOwl:hasExactSynonym` includes exact synonyms. Source of the synonyms are the terms from articles or research papers.



- `oboInOwl:hasRelatedSynonym` includes related synonyms.
- `oboInOwl:hasDbXRef` is used to add additional link from PubMed/NCBI.

The curation guidelines contain several further rules that ontology experts follow during the curation process.

3.4. Harmonization of Alzheimer's disease clinical trial variables with extCTO

Annotation of data from diverse clinical studies considered to be used in TVB-C and integration into a common data model is fundamental with respect to interoperability, comparability, and integrative analysis of data in relation to the process of the FAIRification of clinical data. Therefore, a data model that was developed by FRAUNHOFER SCAI in the IDSN¹⁰ project was utilized and adapted to map variables derived from Alzheimer's Disease Neuroimaging Initiative (ADNI) and AddNeuroMed study (ANM) into the model (poster submitted for AAIC 2020 conference: S. Schaaf et al., "Let the Data Speak! In a Common Language..."). In the further progression of this task effort, extCTO will be used for the annotation of distinct variables including the variable space and the trial design principles for clinical trials in neurodegeneration.

4. SCAIView-Neuro, a Semantic Search Engine

SCAIView is an advanced text mining system and semantic search engine that addresses questions of interest to general biomedical and life science researchers. Most of the current knowledge exists as unstructured text (publications, text fields in databases) and SCAIView provides users with full-text and biomedical concept searches, which are supported by large biomedical terminologies and outstanding text mining technologies. Using machine learning and rule-based named entity recognition, SCAIView extracts information about genes, drugs, SNPs and other life science entities. SCAIView provides various functionalities and further have been developed within TVB-Cloud together with corresponding view components (see D6.1 and D6.2).

Corpus selection and advanced search utility: SCAIView provides access to various corpora, for instance, it includes documents from PubMed, PubMed Central databases (but also arbitrary text corpora). Using the SCAI text mining pipeline, these corpora are annotated with ontological concepts that occur in the documents. The search and query utility allows a user to 'fire' free text search, search for a specific ontological concept, or search documents that are tagged with a specific ontology in all available corpora. Figure 8 and Figure 9 show two screenshots of the corpus selection page and an exemplary search query.

¹⁰ IDSN (Integrative Data Semantics for Neurodegenerative research) project is a German Federal Ministry of Education and Research (BMBF) funded project. More details are available at <https://www.idsn.info/en/>.



Figure 8: User Interface and the functionality of the semantic search engine SCAIView

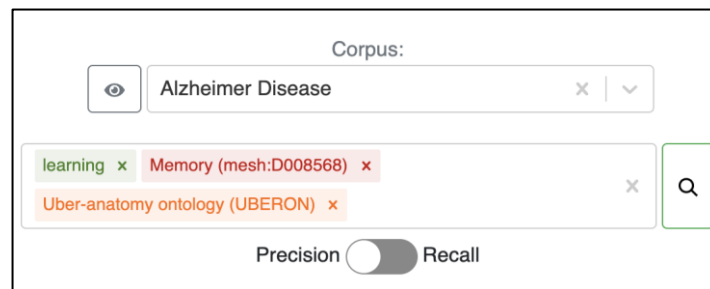


Figure 9: Query example: one can search for 'learning' as a free text (green) term, memory as a concept (red) and 'Uber-anatomy ontology' as an ontology (orange) in the 'Alzheimer Disease' literature collection

Document list view is available to visualize the results of a search query. Figure 10 shows the document list view that highlights the search query and the 843 retrieved documents. Each document is represented by its title, abstract, authors, and the text mined annotations. The view also allows the user to change the sorting of the results.



Found 843 documents

Document List MeSH Graph Citation Table

Sort by: Relevance Newest Oldest

< 1 2 3 4 5 ... 85 > 10 / page ▾

The neuroanatomical and neurophysiological basis of learning.
B J Boss
 PubMed Journal Article
The Journal of neuroscience nursing : journal of the American Association of Neuroscience Nurses – Oct 1986

MESH 17 GO 3 CHEBI 1 GO_COMPONENT 2 UBERON 2 BAO 1
MeSH_Terms 17
Identifiers: PMID:2945875

Learning-forgetting independence, unidimensional memory models, and feature models: comment on Bogartz (1990).
G R Loftus, D Bamber
 PubMed Journal Article
Journal of experimental psychology. Learning, memory, and cognition – Sep 1990

GO 2 MESH 12 GO_COMPONENT 2 UO 1 UBERON 1 BAO 1 hypothesis 1
MeSH_Terms 6
Identifiers: PMID:2147446

Figure 10: Document list view of SCAIView

Full text view: Visualization of a full text view of a document that highlights text mining results (annotations) using different colors for each ontology or terminology. It has been developed specifically to visualize the full-text of a single document. Figure 11 shows a screenshot of a PubMed article with its corresponding text mined results.



The screenshot displays the SCAIView interface for a PubMed article. At the top, the Fraunhofer SCAI logo is visible on the left, and the user 'admin@scaiview.com' is on the right. The search bar contains 'learning', 'Memory (mesh:DO08568)', and 'Uber-anatomy ontology (UBERON)'. Below the search bar, there are tabs for 'Document', 'Network', and 'Tree'. The article title is 'Postnatal calpeptin treatment causes hippocampal neurodevelopmental defects in neonatal rats.' The abstract text is visible, with several terms highlighted in yellow, including 'calpain', 'hippocampus', 'spatial learning', 'memory abilities', 'anxiety-like activities', 'Phalloidin staining', 'hippocampus', 'immunohistochemistry', 'hippocampal CA1 region', 'Dil', 'dendritic spines', 'actin polymerization', 'Western blot assay', 'Akt', 'Erk', 'p-Akt', 'p-Erk1/2', 'Erk1/2', 'SCOP', 'PTEN', 'mTOR', 'p-mTOR', 'CREB', and 'p-CREB'. A concept hierarchy viewer for 'MAPK3' is shown, with 'HGNC:6877' highlighted. The interface also includes an 'Annotations' panel on the right with 'MESH', 'SWISSPROT', 'UBERON', 'HGNC', and 'ENTREZGENE' tabs. At the bottom, there is a 'Fraunhofer SCAI' logo and a 'No description available' message.

Figure 11: Full-text view of the PubMed article with highlighted annotations and the concept hierarchy viewer

SCAView API: The API of SCAIView (available at <https://api.scaiview.com/swagger-ui.html>) provides programmatic access to the above-mentioned functionalities. The API is also consumed by the several SCAIView view components. The search-controller API can be used to search and retrieve documents for a given search query and a corpus ID. Whereas, the document-controller API is being used to retrieve the title, abstract, annotations, authors etc. by providing a document ID. Various other SCAIView functionalities can be applied through further API controllers. We have created a Jupyter notebook especially, to demonstrate the SCAIView functionalities by including examples of full-text queries, semantic queries and further API methods (see Figure 12).



SCAView as API - Create workflows via Jupyter Notebooks

<https://api.scaiview.com/swagger-ui.html>

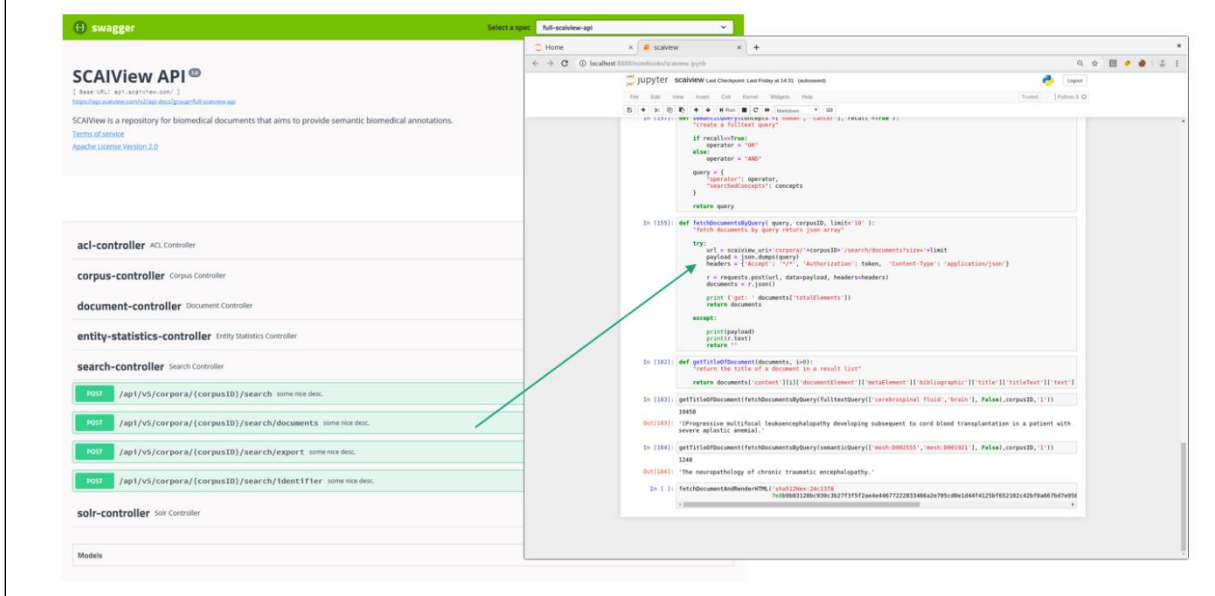


Figure 12: Usage example of the SCAView API in a Jupyter Notebook.

Authentication and Authorization: The access to SCAView GUI and API is secured with the OpenID Connect (OIDC) authentication standard. Only authorize users can access, upload and modify their corpora, documents and associated metadata. Furthermore, these services are also secured on the transport level via Hypertext Transfer Protocol Secure (HTTPS) technique. Hence, SCAView is capable to offer a secured access to sensitive documents.

Availability: SCAView and SCAView API are available at <https://ui.scaiview.com> and <https://api.scaiview.com/>. Registration is needed to access the SCAView service and the API.

5. NeuroImaging Feature Terminology (NIFT)

The NeuroImaging Feature Terminology¹¹ (NIFT) was developed to organize the knowledge domain of measured brain features in association with neurodegenerative diseases by imaging technologies. Annotating raw and derivative imaging data (e.g. from large cohort data sets like ADNI or the UK biobank) using NIFT allows to apply automated workflows to link imaging features with existing ontologies, e.g. for entities at the sub-cellular level.

¹¹ Iyappan A, Younesi E, Redolfi A, Vrooman H, Khanna S, Frisoni GB, et al. Neuroimaging Feature Terminology: A Controlled Terminology for the Annotation of Brain Imaging Features. Journal of Alzheimer's Disease. 2017;59:1153-69



With the bridge of The Virtual Brain simulator (TVB; thevirtualbrain.org) that mechanistically links neuronal population behavior (modulated by the subcellular signaling cascades) to the imaging modality we provide the missing element that can characterize the complex relationship between the different brain scales. Extracting imaging features from a database that is annotated in NIFT would enable us for example to link an imaging dataset to a unified pathway knowledge graph such as PathMe¹².

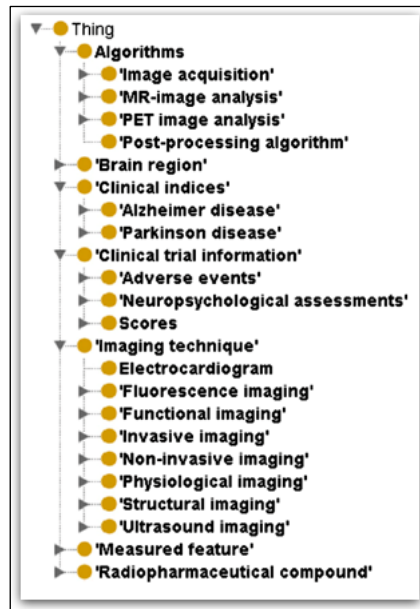


Figure 13: Neuroimaging Feature Terminology (NIFT)

6. NeuroImaging Data Model (NIDM)

In contrast to NIFT, the NeuroImaging Data Model (NIDM) is an ontology that describes the dataset (for example, who is the owner, how were output files generated). NIDM is a Semantic Web based metadata standard that helps capture and describe experimental data, analytic workflows and statistical results via the provenance of the data. NIDM uses consistent data descriptors in the form of machine accessible terminology, and a formal, extensible data model, which enables rich aggregation and query across datasets and domains^{13, 14}.

¹² Domingo-Fernández D, Mubeen S, Marín-Llaó J, Hoyt CT, Hofmann-Apitius M. PathMe: merging and exploring mechanistic pathway knowledge. BMC Bioinformatics. 2019;20(1):243.

¹³ <https://github.com/incf-nidash/nidm-specs/tree/master/nidm/nidm-experiment/terms>

¹⁴ Karl Helmer, Keator David, Tibor Auer, Satrajit Ghosh, Camille Maumet, et al. Development of an Ontology for the INCF Neuroimaging Data Model (NIDM). *OHBM 2018 - 24th Annual Meeting of the Organization for Human Brain Mapping*, Jun 2018, Singapore, Singapore. pp.1-2. ([insefm-01932994](#))



NIDM-Terms

Project Description

THE NEUROIMAGING DATA MODEL: FAIR DESCRIPTORS OF BRAIN INITIATIVE IMAGING EXPERIMENTS NIH RF1 MH120021

Reuse of existing neuroscience data relies, in part, on our ability to understand the experimental design and study data. Historically, a description of the experiment is provided in textual documents, which are often difficult to search, lack the details necessary for data reuse, and are hampered by differences in terminologies across related fields of neuroscience. Our vision is to build on existing resources to create annotation and discovery tools that are based on a metadata standard expressive enough to provide unambiguous descriptions of the experimental methods and metadata. In this proposal, we develop the Experiment component of the Neuroimaging Data Model (NIDM-E), a metadata format leveraging techniques from the semantic web, capable of precisely describing information about the design and intent of an experiment, experimental subject characteristics, and the acquired data. The deliverables of this project focus on developing the terminologies to support NIDM-E and description of BRAIN datasets.

Core Project Team

- [David Keator](#) - UC Irvine
- [Karl Helmer](#) - MGH / Harvard
- [Jeff Grethe](#) - UC San Diego
- [Jean-Baptiste Poline](#) - McGill
- [Satra Ghosh](#) - MIT
- [Theo G.M. VanErp](#) - UC Irvine

Figure 14: Neuroimage data model (NIDM)

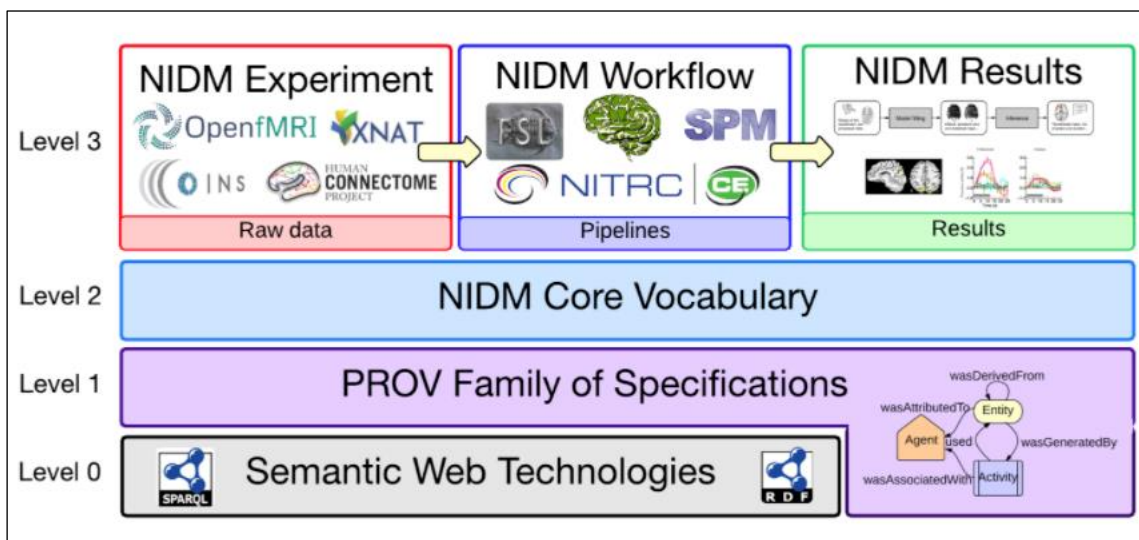


Figure 15: NIDM formalizes the metadata description and required input from the user to define metadata and provenance details.

NIDM is a collection of specification documents that define extensions the W3C PROV¹⁵ standard for the domain of human brain mapping. NIDM uses provenance information as means to link components from different stages of the scientific research process from dataset descriptors and computational workflow, to derived data and publication. NIDM has recommended Brain Imaging Data Structure (BIDS) terms that will be used in TVB-Cloud.

¹⁵ <https://www.w3.org/TR/prov-primer/>



7. Brain Imaging Data Structure (BIDS)

In the brain imaging community, the BIDS – brain imaging data structure¹⁶ – standard is increasingly accepted. BIDS is a standard prescribing a formal way to name and organize neuroimaging data and metadata in a file system that enables easier data processing through using consistent paths and naming for data files.

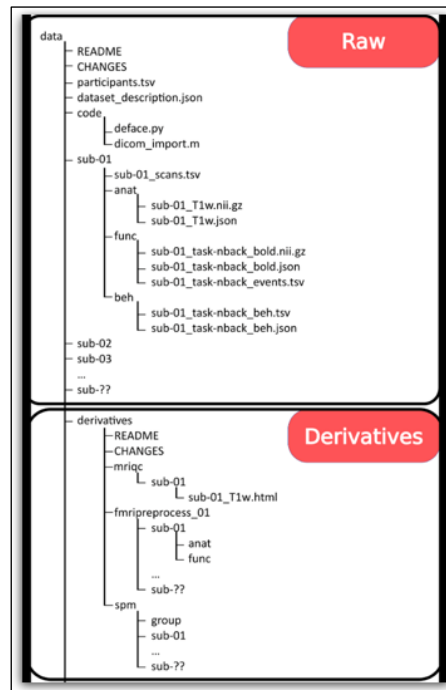


Figure 16: Structure of BIDS

BIDS is focused on practicality minimizing the need to specify metadata. Much of relational information is implicit in the organization of the files in folder structures.

BIDS is strict regarding file organization, naming, and file metadata but, in order to support wide adoption, permits substantial flexibility in the details of how other dataset metadata are described within the standard.

BIDS & NIDM, both of these elements, data organization and data semantics, are critical. The integrated use of both of these standards BIDS plus NIDM can be defined as a “SemanticBIDS” representation. BIDS MRI conversion to NIDM convert a BIDS MRI dataset to a NIDM-Experiment RDF document. It will parse phenotype information and store variables/values and link to the associated json data dictionary file.

PyNIDM reads BIDS TSV files or CSV files, links to online “dictionaries” where users can select which preexisting “term” best matches each TSV or CSV heading. Output is a “turtle” file saved in the root-level BIDS directory. The program iterates over the header variable names performing an elastic search of <https://scicrunch.org/nidm-terms> for NIDM-ReproNim tagged terms that fuzzy match the variable names. The user will then interactively pick a term to associate with the variable name. The resulting annotated CSV data will then be written to a JSON-LD collection – a data format for handling metadata.

¹⁶<https://bids.neuroimaging.io/>

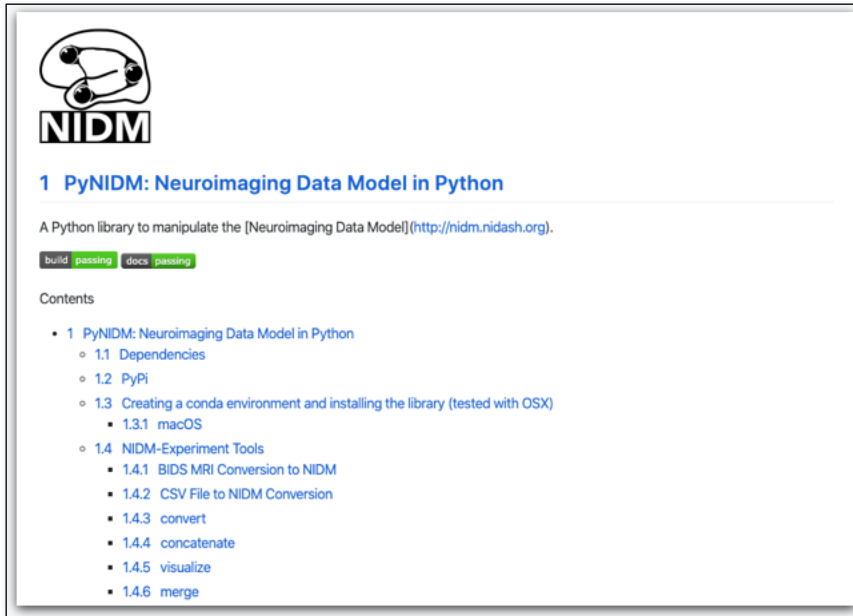


Figure 17: Neuroimaging Data Model in Python

8. Minimal Information Neuroscience Data Set (MINDS)

Minimal information for neuroscience dataset, the open Metadata Initiative for Neuroscience Data Structures powered by the Human Brain Project (HBP) and its European Brain ReseArch InfraStructure (EBRAINS), is a collection of ontology-based metadata schemas for neuroscience data structures¹⁷.

MINDS describes specimen or subject details (including species, age and strain), contributors, brain location (using a standard atlas or brain parcellation ontology), methods (including method types, analysis methods, equipment, parameters and specific protocols), data category (such as electroencephalogram, intracellular recordings, MRI or functional MRI), data-file format and persistent identifiers, showing where the data is stored.

The goal is to develop schemas for data generated in experimental laboratories across all neuroscience modalities as well as the code of computational models and software. Presently few schemas are available to the community. More development work is needed by the community.

The schemas are the architectural building blocks of the Knowledge Graph (KG) database, the data sharing framework by HBP. openMINDS metadata are thought to be linked to existing neuroscience terminologies or ontologies.

¹⁷ Hill SL. How do we know what we know? Discovering neuroscience data sets through minimal metadata. *Nature Reviews Neuroscience*. 2016;17(12):735-6.



A	B	C	D
3	Dataset	Marinazzo	_has_contributor
4	Dataset	HBP_SGA2	created as
5	Dataset	Marinazzo	_has_custodian
6	Dataset	input to TVB (connectomes). In TVB BIDS format, so all the info is machine readable	description
7	Dataset	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	DOI
8	Dataset	not-embargoed	embargo status
9	Dataset	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	intended release date
10	Dataset	CC-BY	license
11	Dataset	Marinazzo	_has_main_contact
12	Dataset	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	_has_main_file_bundle
13	Dataset	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	title
14	EthicsApproval	Ethics approval obtained before HBP - data collected outside HBP	_ID
15	EthicsApproval	Ghent University	authority
16	EthicsApproval	Belgium	country
17	EthicsApproval	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	ID
18	EthicsApproval	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	SP12 approval
19	File	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	_ID
20	File	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	description
21	File	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	format
22	File	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	title
23	File	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	URL
24	FileBundle	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	_ID
25	FileBundle	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	description
26	FileBundle	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	format
27	FileBundle	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	tag
28	FileBundle	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	title
29	FileBundle	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	URL
30	FundingInformation	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	_ID
31	FundingInformation	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	grant ID
32	FundingInformation	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	name
33	MethodParadigm	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	_ID
34	MethodParadigm	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	abbreviation
35	MethodParadigm	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	description
36	MethodParadigm	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	experimental type
37	MethodParadigm	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	full name
38	MethodParadigm	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	type
39	ModelInstance	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	_ID
40	ModelInstance	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	_is_modelling_sub_cellular_target
41	ModelInstance	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	abstraction level
42	ModelInstance	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	alias
43	ModelInstance	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	_is_modelling_brain_structure
44	ModelInstance	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	_has_contributor
45	ModelInstance	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	_has_custodian
46	ModelInstance	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	description
47	ModelInstance	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	_has_main_contact
48	ModelInstance	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	model format
49	ModelInstance	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	model format version compatibility
50	ModelInstance	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	model scope
51	ModelInstance	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	species
52	ModelInstance	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	title
53	ModelInstance	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	version
54	Person	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	_ID
55	Person	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	email
56	Person	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	first name
57	Person	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	last name
58	Person	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	ORCID
59	PLAComponent	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	_ID
60	PLAComponent	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	associated task (ID)
61	PLAComponent	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	ID
62	PLAComponent	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	_has_owner
63	PLAComponent	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	phase
64	Project	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	_ID
65	Project	<< enter corresponding_ID in order to assign metadata key and value (+unit) >>	_has_coordinator

Figure 18: openMINDS form to enter metadata manually

All openMINDS schemas are defined in JSON-Schema and can be serialized in JSON-LD. The latter can also be directly digested as machine-readable data descriptions into the database of the KG. In Figure 18 we see an excel sheet that is also used to enter metadata manually in HBP.

To make the data model richer and granularity finer, metadata standards for specific categories and subcategories of data are required.

This can be achieved by advancing present MINDS schemas by keys from the Neuroimaging Data Model (NIDM), which was developed by the International Neuroinformatics Coordinating Facility (INCF) task force or by the Neuroimaging feature terminology NIFT.

Notably, HBPs KG so far in is good for high level labeling of data. It doesn't help to model causal effects on a mechanistic level.

9. Conclusions

Standards do not imply a ban of variety. Variable vocabularies for the same matter can be integrated by mapping tools. D3.14 is further elaborating on tools for creating interoperability. Given the fast developments in the fields in terms of data acquisition and analyses techniques, requirements for those standards will rapidly change with time. Therefore, we will take an evolutionary approach in TVB-Cloud that allows for continuous updates and advancements of existing standards. Ideally, we can automate this process, of a continuous evolutionary development to large degrees. The benefit of using standardized vocabularies in TVB-Cloud will be not only in making data FAIR but also in automating the



processing of the data and hence being able to process big data yielding generalizable results that can be applied in clinics.

10. Glossary

AAIC – Alzheimer's Association International Conference

ADNI – Alzheimer's Disease Neuroimaging Initiative

ANM – AddNeuroMed study

API – Application Programming Interface

BFO – Basic Formal Ontology

BIDS – Brain Imaging Data Structure

CSV – Comma separated values

CTO – Clinical Trials Ontology

EBRAINS – European Brain ReseArch InfraStructure

FDA – Food and Drug Administration

GUI – Graphical User Interface

HBP – Human Brian Project

ICTRP – International Clinical Trials Registry Platform

ICMJE – International Committee of Medical Journal Editors

IDSN – Integrative Data Semantics for Neurodegenerative research

JSON-LD – JSON for linking data

MRI – Magnetic resonance imaging

NCBI – National Center for Biotechnology Information

NDD – Neurodegenerative disease

NIDM – Neuroimage data model

NIFT – Neuroimage feature terminology

OBO – Open Biological and Biomedical Ontology

OLS – Ontology Lookup Service



Ontology – refers to a “knowledge graph”-type web which specify terms and the relationships between these terms

PathMe – Tool that combines and visualizes information from three pathway databases

RDF – Resource description framework

ROHAN – Referential Ontology Hub for Applications within Neurosciences

SNP – Single Nucleotide Polymorphism

Terminology – refers to hierarchically listed terms as subtypes of other terms

Tsv – tab separated values

TVB-Cloud – The Virtual Brain Cloud

WHO – World Health Organization

W3C – The World Wide Web Consortium (W3C) is the main international standards organization for the World Wide Web.