

NIDM-Experiment: An Ontology for Annotating Neuroscientific Data

Karl G. Helmer¹, Derek Chaplin¹, Nazek Queder², Satrajit Ghosh³, Camille Maumet⁴, Jean-Baptiste Poline⁶, Theo van Erp², David Keator²

¹ *Massachusetts General Hospital, Boston, MA, USA*

² *University of California, Irvine, Irvine, CA, USA*

³ *Massachusetts Institute of Technology, MA, USA*

⁴ *University of Rennes, Inria, CNRS, Inserm, IRISA, Rennes, France*

⁵ *McGill University, Montreal, Canada*

Abstract

NIDM-Experiment (NIDM-E) provides a collection of general and domain-specific terms that can be used to annotate data from neuroscientific experiments. NIDM-E reuses terms from existing ontologies and standards (such as DICOM and the Brain Imaging Data Standard) and adds new defined domain-specific terms. NIDM-E was created by annotating existing datasets and provides tools such as an online schema browser and term URI-resolution pages, and a GitHub-based workflow for users to propose new terms and edits existing ones.

Keywords

neuroscience, ontology, data annotation, neuroimaging

1. Introduction

Efficacious reuse of data [1] relies on the capture and availability of information describing the acquisition and processing of that data. Crucial to data annotation are the use of defined terms, a practice that avoids ambiguities in interpretation, and the reuse of terms from established, active vocabularies and ontologies [2]. In addition, terms should be resolvable through a URL so that automated methods and users are able to retrieve information about each term from a single resource that is kept up to date. It is also critical that any ontology allow users to have input on its development with the goal that community adoption will be provide a feedback loop that drives further development. We report here on recent developments in the Neuroimaging Data Model-Experiment (NIDM-E) [3], an ontology

used to describe neuroscience experiments and the resulting data artifacts.

The goal of NIDM-E is to provide semantic-web and other tools, a collection of defined terms that can be used to annotate data to an arbitrary level of detail. Other, often used standards, such as the Brain Imaging Data Structure (BIDS) [4], embed metadata into fixed directory structures and metadata file formats, and these can be restrictive when dealing with complicated study or experimental configurations, such as multiple acquisition modalities, multi-site studies, and cutting-edge acquisition methods. NIDM-E can provide a framework to both annotate complicated experiments and data, as well as accommodate terms for new modalities and acquisition methods. It also provides tools to find

ICBO 2022, September 22-25, 2022, Ann Arbor, MI, USA.

EMAIL: khelmer@mgh.harvard.edu (A. 1); dchaplin1@mgh.harvard.edu (A. 2); naqueder@gmail.com (A. 3); satra@mit.edu (A. 4); camille@inria.fr (A. 5); jbpoline@gmail.com (A. 6); tvanerp@hs.uci.edu (A. 7); dbkeator@hs.uci.edu (A. 8)

ORCID: 0000-0002-5113-6843 (A. 1); none (A. 2); 0000-0001-6268-862X (A. 3); 0000-0002-5312-6729 (A. 4); 0000-0002-6290-553X (A. 5); 0000-0002-9794-749X (A. 6); 0000-0002-2465-2797 (A. 7); 0000-0001-5281-5576 (A. 8)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

terms, webpages for term URL resolution, and a framework for community involvement.

2. Methods

NIDM-E began by focusing on the processing and analysis of neuroimaging experiment data, but has since expanded to encompass other neuroscientific modalities as well as more general terms that describe experiments. NIDM-E was built by annotating of several large real-world multi-modality neuroscientific data sets.

NIDM-E reuses terms from other ontologies such as the Semanticscience Integrated Ontology (SIO) [5], Information Artifact Ontology (IAO) [6] and Prov-O [7]. These more general ontologies provide the scaffold onto which domain-specific terms can be added. Terms that are created for NIDM-E include formal definitions in the “X is a Y that Z” format [8]. NIDM-E also includes a wide range of datatype, object, and annotation properties.

Because NIDM-E began in support of a project to annotate neuroimaging data, it has particularly strong coverage in that domain. It contains two unique properties: “hadImageContrastType” and “hadImageUsageType” that are used to distinguish between the physical mechanism for the contrast in an image volume (e.g., “T1-weighted”) and the eventual application for that image (e.g., “Anatomical”). These are particularly important for the discovery of data in and across repositories, where datasets with different image contrasts may be annotated by usage. For example, T1-weighted, T2-weighted, and diffusion-weighted images all may be stored as “Anatomical” data. To further support the annotation of neuroimaging data, NIDM-E also includes terms from two widely used standards: DICOM [9] and the BIDS standards, which are ubiquitous in the neuroimaging domain over multiple imaging modalities. We have created a set of datatype properties, each representing a specific DICOM tag, which can be used to associate acquisition parameters with an acquisition object. We have also included BIDS terms so that datasets that are organized according to the BIDS standard can be annotated using BIDS-approved terms.

We show in Fig. 1 a simple example of how NIDM-E can annotate an acquisition object, T1.nii, with an image contrast type of “T1-weighted” and an image usage type of “Anatomical”, and showing the scan session activity it was acquired at (“Session:Visit_2”), the protocol that was used (“MyProtocol.pdf”), and the study participant (“ID:a8d4j3”) from which it was acquired and who had the role of “In-Vivo Participant”.

NIDM-E is currently used as a term source by the PyNIDM [10] package, a set of python-based open-source data annotation tools that is both customizable and extensible. Recently, PyNIDM has been used to augment the data dictionary terms for the OpenNeuro data repository [11], the publicly-accessible repository for NIH’s BRAIN project MRI data and other related datasets.

NIDM-E is accessible by cloning the repository from GitHub [3]. Web-accessible infrastructure has been built so that the neuroscientific community can suggest terms to NIDM-E vocabulary, terms have resolvable URI’s, and the ontology can be browsed to aid in term discovery. We use GitHub issue templates to allow users to suggest new terms or edits to existing ones. This allows us to have a record of discussions regarding a particular term and its resolution. To discover terms, we have provided a “Schema Browser” webpage [12] that allows users to view the entire graph of NIDM-E terms including all of the terms imported from other ontologies. For semantic web applications, we also have created a “Terms Resolution” page in which each term has a unique URL [13] so that terms by applications have a unique reference location.

3. Summary

NIDM-E is a flexible collection of classes and properties that can be used to annotate a wide range of neuroscientific data, with a strong current focus on neuroimaging. It provides tools to help users discover terms and provides a resolvable URL for each term. It also provides a workflow for the community to suggest new terms and edits to existing ones.

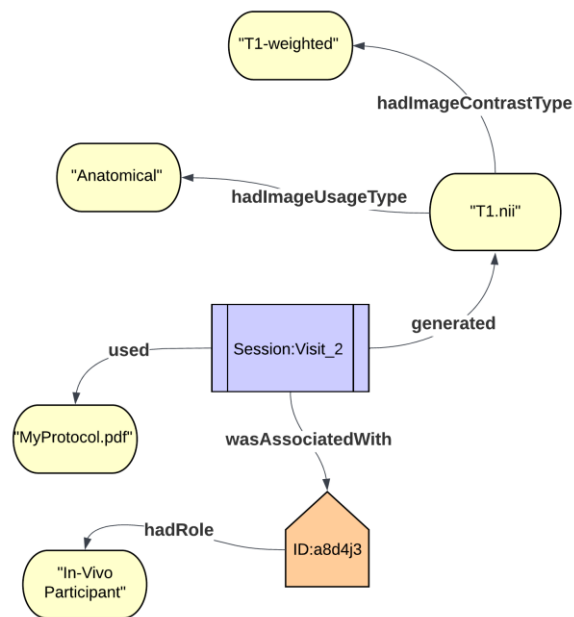


Figure 1: Simple example showing how NIDM-Experiment can annotate an acquisition object, showing the scan session it was acquired at and from what study participant.

4. Acknowledgements

We acknowledge support from NIH grant 1RF1MH120021-01 and ongoing support from the International Neuroinformatics Coordinating Facility.

5. References

[1] J-B Poline, J.L. Breeze, S. Ghosh, K. Gorgolewski, Y. O. Halchenko, M. Hanke, C. Haselgrove, K. G. Helmer, D. B. Keator, D. S. Marcus, R. A. Poldrack, Y. Schwartz, J. Ashburner, D. N. Kennedy. "Data sharing in neuroimaging research." *Front Neuroinform.* (2012) 6: 9. doi:10.3389/fninf.2012.00009..

[2] R. Arp, B. Smith, A. D. Spear. *Building Ontologies with Basic Formal Ontology*, MIT Press, 2015. <https://doi.org/10.7551/mitpress/9780262527811.001.0001>.

[3] incf-nidash/nidm-experiment: This repository contains the terms used in nidm-experiment and the code to create the term-resolution and term-schema webpages. URL: <https://github.com/incf-nidash/nidm-experiment>.

[4] Brain Imaging Data Structure. URL: <https://bids.neuroimaging.io/>

[5] M. Dumontier, C. J.O. Baker, J. Baran, A. Callahan, L. Chepelev, J. Cruz-Toledo, N. R Del Rio, G. Duck, L. I. Furlong, N. Keath, D. Klassen, J. P. McCusker, N. Queralt-Rosinach, M. Samwald, N. Villanueva-Rosales, M. D. Wilkinson, t Hoehndorf. The Semanticscience Integrated Ontology (SIO) for biomedical research and knowledge discovery. *J. Biomed. Semantics* 5: 14 (2014).

[6] W. Ceusters. An information artifact ontology perspective on data collections and associated representational artifacts. *Stud Health Technol Inform.* 180:68-72 (2012).

[7] PROV-O: The PROV Ontology. URL: <https://www.w3.org/TR/prov-o/>

[8] S. Seppalla, A Ruttenberg, B. Smith. Guidelines for writing definitions in ontologies. *Ciência da Informação* 46 (1): 73-88 (2017).

[9] DICOM. URL: <https://www.dicomstandard.org/>

[10] incf-nidash/PyNIDM. URL: <https://github.com/incf-nidash/PyNIDM>.

[11] OpenNeuro. URL: <https://openneuro.org/>

[12] NIDM-Experiment Schema Browser. URL: https://incf-nidash.github.io/nidm-experiment/schema_menu.html

[13] NIDM Experiment Term Resolution Page. URL: <https://incf-nidash.github.io/nidm-experiment/>