



HAL
open science

Tools for FAIR Neuroimaging Experiment Metadata Annotation with NIDM Experiment

David Keator, Karl Helmer, Camille Maumet, Smruti Padhy, Dorota Jarecka,
David N. Kennedy, Satrajit Ghosh, Jean-Baptiste Poline

► **To cite this version:**

David Keator, Karl Helmer, Camille Maumet, Smruti Padhy, Dorota Jarecka, et al.. Tools for FAIR Neuroimaging Experiment Metadata Annotation with NIDM Experiment. OHBM 2019 - 25th Annual Meeting of the Organization for Human Brain Mapping,, Jun 2019, Rome, Italy. pp.1-5. inserm-02379309

HAL Id: inserm-02379309

<https://inserm.hal.science/inserm-02379309v1>

Submitted on 25 Nov 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Tools for FAIR Neuroimaging Experiment Metadata Annotation with NIDM Experiment

Presented During: Posters Session

Wednesday, June 12, 2019: 12:45 PM - 02:45 PM

Poster No:

W601

Submission Type:

Abstract Submission

Authors:

David Keator¹, Karl Helmer², Camille Maumet³, Smruti Padhy⁴, Dorota Jarecka⁵, David Kennedy⁶, Satrajit Ghosh⁷, Jean-Baptiste Poline⁸

Institutions:

¹University of California, Irvine, Irvine, CA, ²Massachusetts General Hospital, Charlestown, MA, ³Inria, Univ Rennes, CNRS, Inserm, Rennes, France, ⁴Massachusetts Institute of Technology, Cambridge, MA, ⁵Massachusetts Institute of Technology, Boston, MA, ⁶University of Massachusetts Medical School, Worcester, MA, ⁷MIT/HMS, Cambridge, MA, ⁸McGill University, Montreal, Quebec

E-POSTER

Introduction:

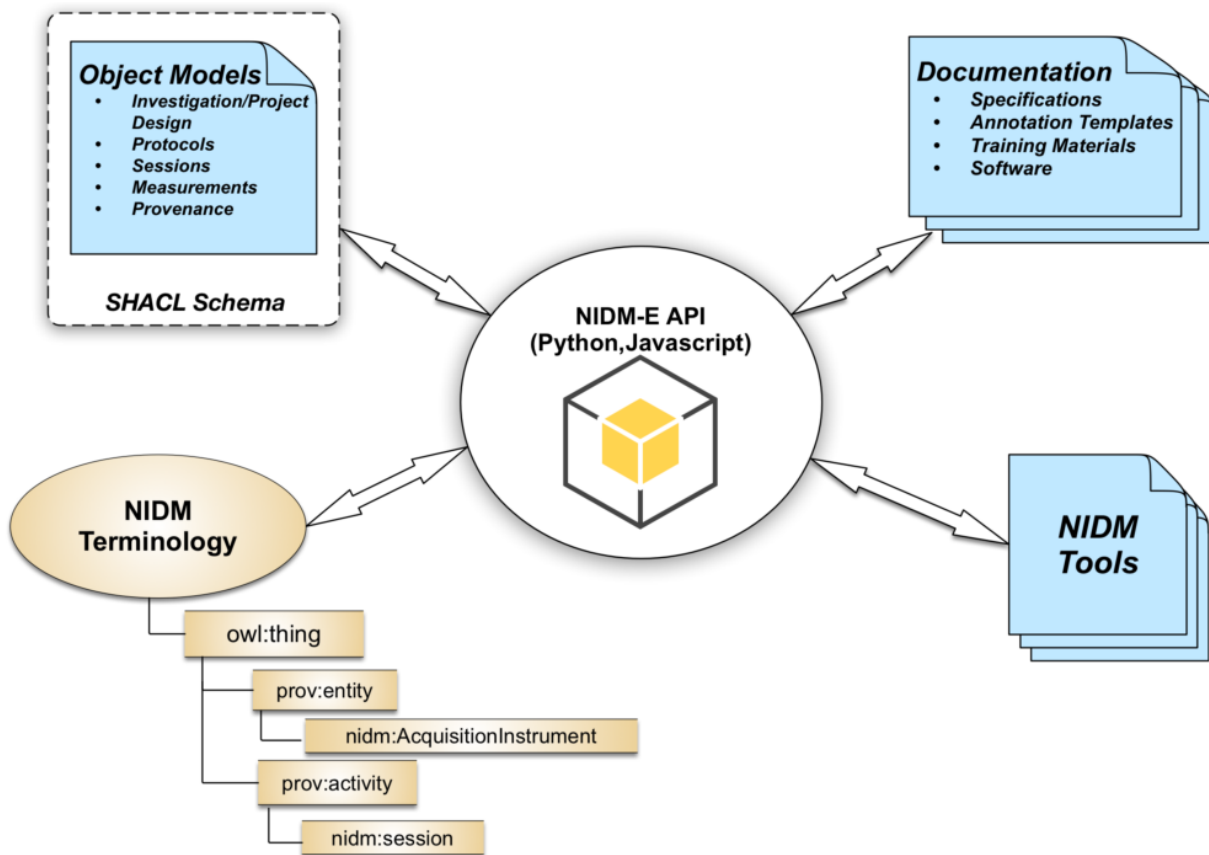
Acceleration of scientific discovery relies on our ability to effectively use data acquired by consortiums and/or across multiple domains to generate robust and replicable findings. Efficient use of existing data relies on metadata being FAIR¹ - Findable, Accessible, Interoperable and Reusable. Typically, data are shared using formats appropriate for the specific data types with little contextual information. Therefore, scientists looking to reuse data must contend with data originating from multiple sources, lacking complete acquisition information and often basic participant information (e.g. sex, age). What is required is a rich metadata standard that allows annotation of participant and data information throughout the experiment workflow, thereby allowing consumers easy discovery of suitable data.

The Neuroimaging Data Model (NIDM)² is an ongoing effort to represent, in a single core technology, the different components of a research activity, their relations, and derived data provenance³. NIDM-Experiment (NIDM-E) is focused on experiment design, source data descriptions, and information on the participants and acquisition information. In this work we report on annotation tools developed as part of the PyNIDM⁴ application programming interface (API) and their application to annotating and extending

the BIDS⁵ versions of ADHD200⁶ and ABIDE⁷ datasets hosted in DataLad⁸.

Methods:

NIDM is built on the PROV family of documents^{9,10} consisting of three interconnected specifications: Experiment, Workflow, and Results. The organization of information in a NIDM-E document follows a simple structure. At each level, additional information is added in the form of tuples. Tuples are formed with uniform resource identifiers (URIs) which resolve to term definitions, providing precise annotation of the statements about the data and consistent with the Resource Description Framework (RDF)^{11,11}. We have created a basic python-based API (PyNIDM) which follows the simple organizational structure of NIDM-Exp documents with functions to create, query, export, import, and transform NIDM-Exp documents (Fig. 1). Using the API, we have built annotation tools which provide functionality to interactively annotate tabular data and BIDS-formatted MRI datasets, augment such datasets with additional metadata.

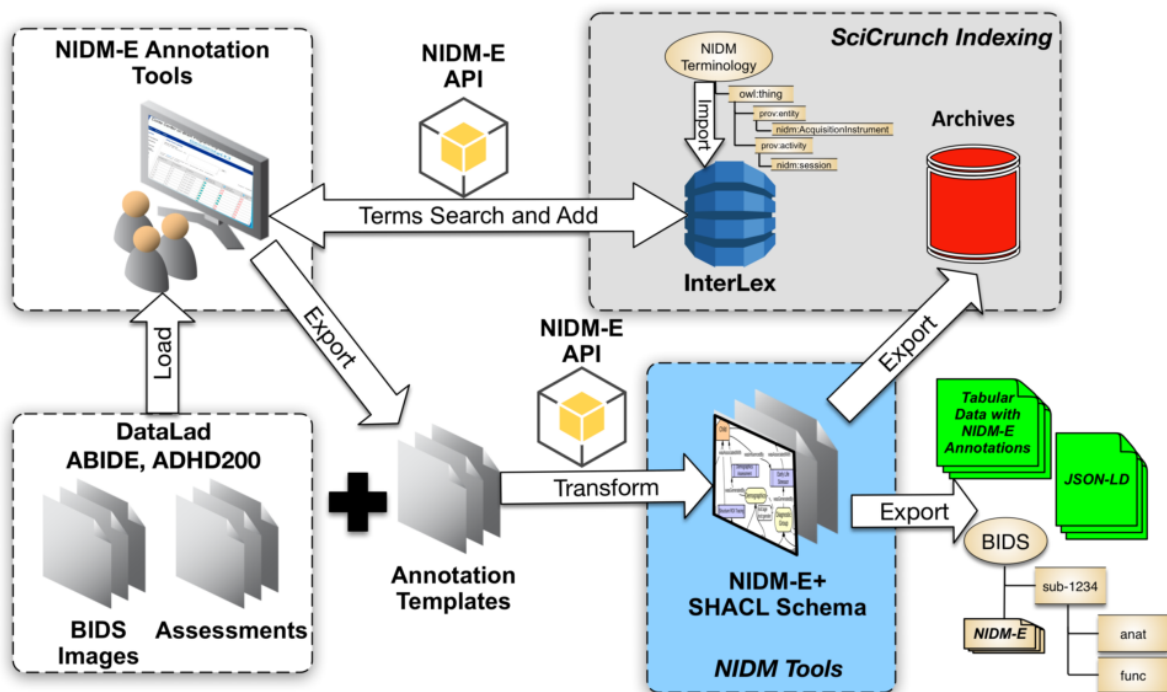


•PyNIDM include NIDM API, object models, SHACL schema for validation, NIDM terminologies, tools, and documentation. Each of the components are in continual development with varying levels of maturity

Results:

Fig. 2 summarizes annotation through a user-centered illustration. From upper left, researchers interactively annotate their data through NIDM-E facilitated queries for terms in InterLex, the NIDM-E ontology, and other terminology sources. We have modeled data from the open-source ABIDE and ADHD200 datasets hosted on DataLad by extracting BIDS metadata into a NIDM-E document, annotating variables during the extraction process. For ADHD200, the BIDS representation didn't include assessment data. Using NIDM-E tools we annotated the assessment data and added it to the NIDM-E document created from the BIDS extraction. Mappings of dataset-specific variables to terms are created through the annotation process and applied to the original data creating NIDM-E annotated metadata and can be saved as JSON "sidecar" files in a BIDS dataset. Through SHACL schema documents, NIDM-E annotated

data can be validated against project-specific protocols. Further, the NIDM-E documents can then be imported into graph-based archives and/or stored and shared through file-based systems (e.g. DataLad, Amazon AWS, etc.). Using the NIDM-E documents and API, we currently provide simple project-specific and general SPARQL formatted queries. We anticipate the full query API to be completed in 2019.



·NIDM-E annotations and export, user-centric illustration.

Conclusions:

The PyNIDM is a simple python-based API for managing and creating NIDM-Exp documents. Future work will include additional functions and tools to support complex queries and integration with NIDM-Workflows and NIDM-Results.

Informatics:

Databasing and Data Sharing ²
Informatics Other ¹

Keywords:

Data Organization
Informatics
Other - NIDM

¹²Indicates the priority used for review

My abstract is being submitted as a Software Demonstration.

No

Please indicate below if your study was a "restingstate" or "task-activation" study.

Resting state
Task-activation

Healthy subjects only or patients (note that patient studies may also involve healthy subjects):

Healthy subjects

Was any human subjects research approved by the relevant Institutional Review Board or ethics panel? NOTE: Any human subjects studies without IRB approval will be automatically rejected.

Not applicable

Was any animal research approved by the relevant IACUC or other animal research panel? NOTE: Any animal studies without IACUC approval will be automatically rejected.

Not applicable

Please indicate which methods were used in your research:

Functional MRI
Structural MRI
Behavior
Neuropsychological testing

For human MRI, what field strength scanner do you use?

3.0T

Which processing packages did you use for your study?

Other, Please list - NIDM

Provide references using author date format

1. Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016).
2. Neuroimaging Data Model. Available at: <http://nidm.nidash.org/>.
3. Keator, D. B. et al. Towards structured sharing of raw and derived neuroimaging data across existing resources. *Neuroimage* 82, 647–661 (2013).
4. incf-nidash. incf-nidash/PyNIDM. GitHub Available at: <https://github.com/incf-nidash/PyNIDM>.
5. K. J. Gorgolewski, A. Tibor, V. D. Calhoun, R. Cameron Craddock, D. Samir, E. P. Duff, F. Guillaume, S. S. Ghosh, G. Tristan, Y. O. Halchenko, D. A. Handwerker, H. Michael, K. David, L. Xiangrui, M. Zachary, M. Camille, B. Nolan Nichols, T. E. Nichols, P. John, P. Jean-Baptiste, R. Ariel, S. Gunnar, S. Vanessa, T. William, J. A. Turner, V. Gaël, and R. A. Poldrack, “The Brain Imaging Data Structure: a standard for organizing and describing outputs of neuroimaging experiments,” 2015.
6. “ADHD200”. Available: http://fcon_1000.projects.nitrc.org/indi/adhd200/
7. “ABIDE”. Available: http://fcon_1000.projects.nitrc.org/indi/abide/.
8. DataLad Team, “DataLad.” [Online]. Available: <https://www.datalad.org/>.

9. PROV-Overview. Available at: <https://www.w3.org/TR/prov-overview/>.

10. Moreau, L. et al. Special Issue: The First Provenance Challenge. *Concurr. Comput.* 20, 409–418 (2008).

11. RDF 1.1 Turtle. Available at: <https://www.w3.org/TR/2014/REC-turtle-20140225/>.

12. RDF - Semantic Web Standards. Available at: <https://www.w3.org/RDF/>.