Development of an Ontology for the INCF Neuroimaging Data Model (NIDM)

Karl Helmer, Keator David, Tibor Auer, Satrajit Ghosh, Camille Maumet, Thomas E. Nichols, Padhy Smruti, Jean-Baptiste Poline

Development of an Ontology for the INCF Neuroimaging Data Model (NIDM)

1. Karl G. Helmer, Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital; Dept. of Radiology, Harvard Medical School, Boston, MA.
2. David Keator, Department of Psychiatry and Human Behavior, University of California, Irvine CA.
3. Tibor Auer, Department of Psychology, Royal Holloway University of London, Egham, United Kingdom.
4. Satra Ghosh, McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge MA.
5. Camille Maumet, University of Rennes, Inria, CNRS, Inserm, IRISA, Rennes, France.
6. Thomas E. Nichols Oxford Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Nuffield Department of Population Health, University of Oxford, Oxford, UK.
7. Smruti Padhy, McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge MA.
8. Jean-Baptiste Poline, McGill University, Ludmer Centre, Montreal Neurological Institute, Montreal, Quebec, Canada

Introduction

The successful reuse of shared data relies on the existence of easily-available well-described metadata [1]. The metadata, as a rich description of the data, must capture information on how the data was acquired, processed and analyzed. The terms used to describe the data should be chosen with a logical, consistent framework in mind and include definitions to avoid ambiguity. In addition, a lexicon or ontology should reuse terms from existing efforts as much as possible [2].

We report here on recent progress in the construction of two ontologies used to describe experiments, originally focused on neuroimaging, but recently expanded to encompass other modalities such as electrophysiology experiments. The goal of this work is to provide a set of terms that can be used by the Neuroimaging Data Model (NIDM) [3] and by others. NIDM is comprised of three distinct models: 1) NIDM-Experiment, NIDM-Workflows, and NIDM-Results. These models are being developed under the aegis of the International Neuroimaging Coordinating Facility [4] (INCF). NIDM-Results was developed first and is now available for use [5]. The other two models are under active development.

Methods:

Currently, our efforts are focused on terms describing both a general experiment acquisition and workflow (e.g., “Project” and “Acquisition”), and those that are modality specific (e.g., “T2 Star Weighted Imaging” and “Intracellular Recording”). The NIDM ontology uses the PROV Ontology (PROV-O) [6] as its base ontology, which expresses the PROV Data Model (PROV-DM) [7] using the OWL2 Web Ontology Language (OWL2). The NIDM ontology also incorporates terms from several other ontologies (Fig. 1) that provide terms related to diverse categories such as assessment instruments. Blue entries in Fig. 1 denote products of this project, purple denotes the sources for imported terms.

We have also created a separate DICOM-tag ontology, which is available at https://github.com/incf-nidash/dicom-ontology. Included are terms for each of the DICOM tags, as well as terms that have been extracted from DICOM supplement documents. These terms will be used in the tagging of experimental parameters for data in the DICOM format.

To develop this ontology, we are using GitHub as an environment for openly discussing and defining terms and definitions. The agreed-upon definitions are then referenced by URI in Terse RDF Triple Language (Turtle) documents and expanded into a hierarchical Web Ontology Language (OWL) file. The final OWL file and files containing terms imported from other ontologies are also available on GitHub.

Results:

The NIDM-Experiment ontology is available at: https://github.com/incf-nidash/nidm/. We are currently adding definitions to the new NIDM namespace terms using a community based workflow. Potential contributors and users are encouraged to request terms by raising a new GitHub issue and participating in the definition process.

As an example, Figure 2 shows the basic element of the model for two activities: informed consent and the acquisition of neuropsychological testing data. The basic interaction in the Prov data model is that an Agent with a specific role is associated with an Activity that generates or uses an entity.

Conclusions:

We have created a set of defined terms for various aspects of the neuro-imaging enterprise for use in the INCF Neuroimaging Data Model as well as more broadly. In future work, we will be adding terms for other imaging modalities as well as other neuro-related domains such as electrophysiology.

References:

[1] Poline, Jean-Baptiste, Janis L. Breeze, Satrajit Ghosh, Krzysztof Gorgolewski, Yaroslav O. Halchenko, Michael Hanke, et al. 2012 "Data sharing in neuroimaging research", Front. Neuroinform. 6: 9. doi:10.3389/fninf.2012.00009.
[2] Arp, Robert, Barry Smith and Andrew D. Spear 2015, "Building Ontologies with Basic Formal Ontology", MIT Press. https://doi.org/10.7551/mitpress/9780262527811.001.0001.
[3] http:// http://nidm.nidash.org/
[4] https://www.incf.org
[5] Maumet, Camille, Tibor Auer, Alexander Bowring, Gang Chen, Samir Das, Guillaume Flandin, Satrajit Ghosh, et al. 2016. "Sharing Brain Mapping Statistical Results with the Neuroimaging Data Model." Scientific Data 3 (160102). Nature Publishing Group. https://doi.org/10.1038/sdata.2016.102.
[6] https://www.w3.org/TR/prov-o/
[7] https://www.w3.org/TR/prov-dm/

Acknowledgements:

We would like to acknowledge the work of all the INCF task force members as well as of many other colleagues who have helped the task force. We are particularly indebted to M. Abrams and the INCF secretariat staff. We also acknowledge support of this project by the Wellcome Trust (CM, TEN). J.-B.P. was partially funded by NIH-NIBIB P41 EB019936 (ReproNim) NIH-NIMH R01 MH083320 (CANDIShare) and NIH 5U24 DA039832 (NIF).

Fig. 1

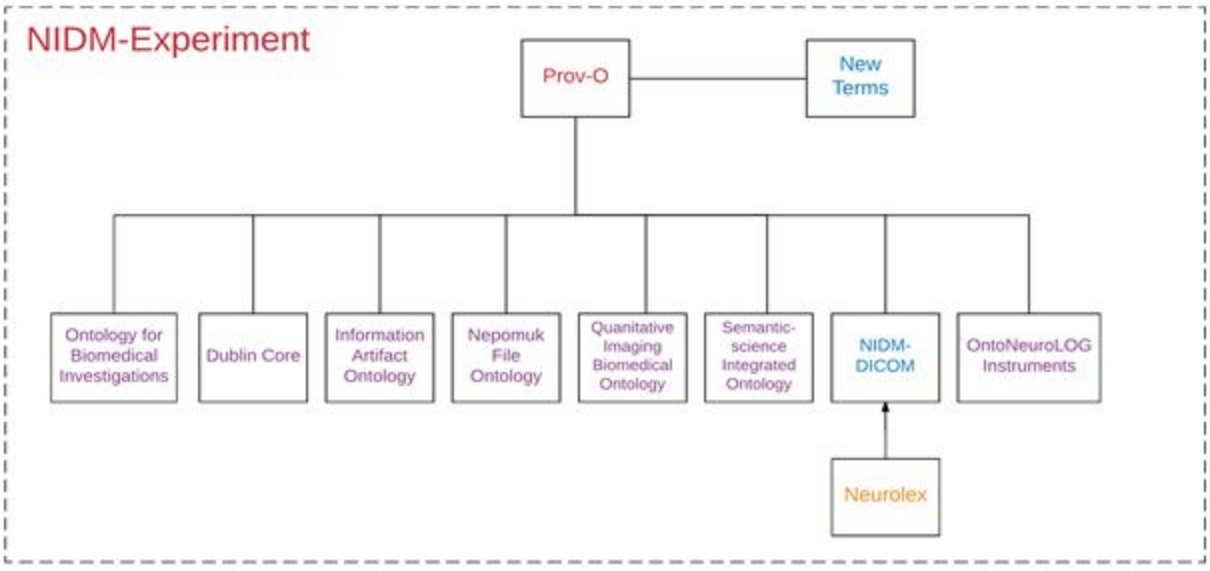
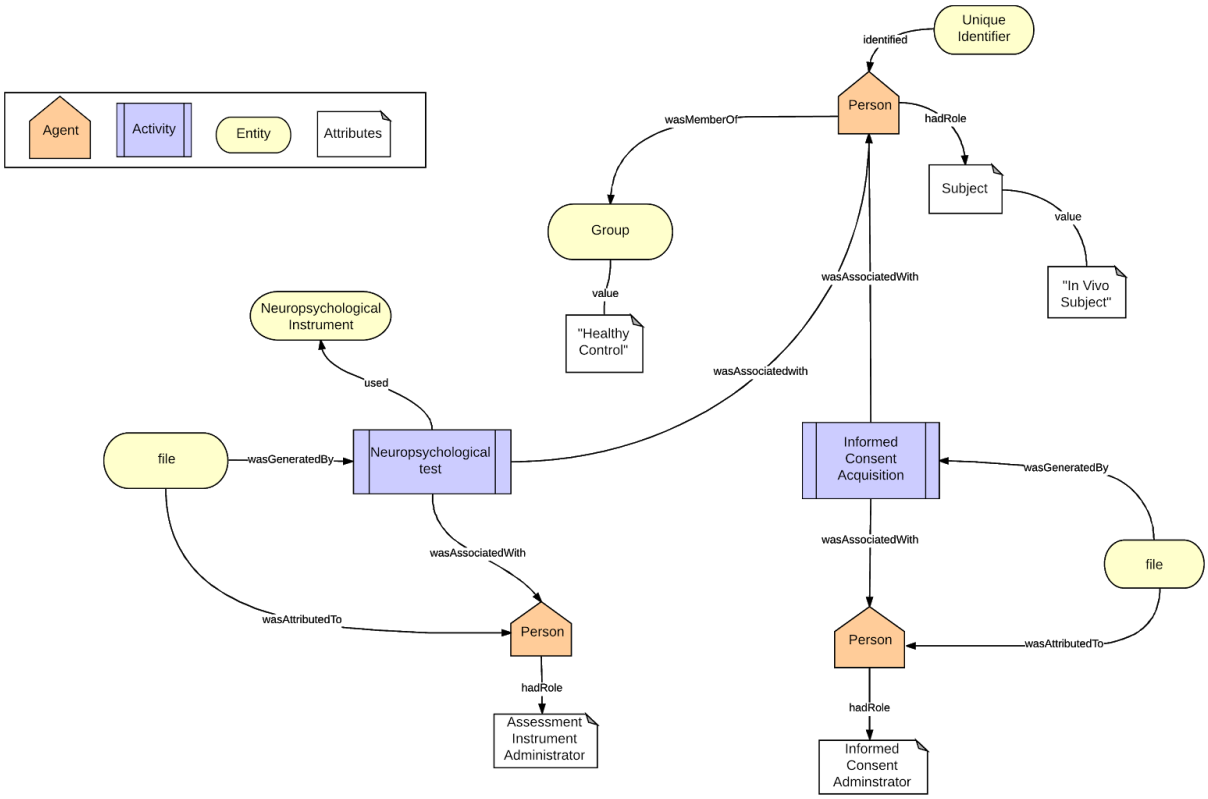

NIDM-Experiment
Prov-O
New Terms
Ontology for Biomedical Investigations
Dublin Core
Information Artifact Ontology
Nepomuk File Ontology
Quanitative Imaging Biomedical Ontology
Semantic-science Integrated Ontology
NIDM-DICOM
OntoNeuroLOG Instruments
Neurolex

Fig 2.


Agent
Activity
Entity
Attributes
Unique Identifier
identified
Person
wasMemberOf
hadRole
Subject
value
"In Vivo Subject"
Group
value
"Healthy Control"
wasAssociatedWith
Neuropsychological Instrument
used
wasAssociatedwith
file
wasGeneratedBy
Neuropsychological test
Informed Consent Acquisition
wasGeneratedBy
file
wasAssociatedWith
wasAssociatedWith
Person
wasAttributedTo
Person
wasAttributedTo
hadRole
Assessment Instrument Administrator
hadRole
Informed Consent Adminstrator