Application of the Neuroimaging Data Model to Represent and Exchange Primary and Derived Data

Nolan Nichols, David Keator, Satrajit Ghosh, Camille Maumet, Guillaume Flandin, Thomas Nichols, Krzysztof Gorgolewski, Y. O. Halchenko, Michael Hanke, Christian Haselgrove, et al.

Application of the Neuroimaging Data Model to Represent and Exchange Primary and Derived Data

B. N. Nichols[1,2], D. Keator[3], S. Ghosh[4], C. Maumet.[5], G. Flandin[6], T. Nichols[5,7], K. J. Gorgolewski[8], Y.O. Halchenko[9], M. Hanke[10], C. Haselgrove[11], K.G. Helmer[12], D. Marcus[13], R. Poldrack[8], J. Turner[14], D. Kennedy[11], J.B. Poline[15,], K. M. Pohl[1]

1. Neuroscience Program, SRI International, Menlo Park, CA, USA
2. Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, CA, USA
3. Department of Psychiatry and Human Behavior, Dept. of Computer Science, University of California, Irvine, CA, USA
4. Massachusetts Institute of Technology, MA, USA
5. University of Warwick, Warwick Manufacturing Group, Coventry, United Kingdom.
6. Wellcome Trust Centre for Neuroimaging, UCL Institute of Neurology, London, United Kingdom
7. University of Warwick, Department of Statistics, Coventry, United Kingdom.
8. Psychology, Stanford University, Stanford, CA, USA
9. Dartmouth College, NH USA
10. Otto-von-Guericke-University, Magdeburg, Germany
11. University of Massachusetts Medical School, Department of Psychiatry, MA, USA
12. Massachusetts General Hospital, Boston, MA
13. Neuroimaging Informatics and Analysis Center at Washington University, St Louis, USA
14. Psychology and Neuroscience, Georgia State University, Atlanta, GA, USA
15. Helen Wills Neuroscience Institute, Brain Imaging Center, University of California, Berkeley, CA, USA

Introduction

Access to the provenance of primary and derived data is increasingly recognized as an essential aspect of reproducibility in biomedical research [1]. While data sharing is the norm in some biomedical communities, neuroimaging has lagged in open data and provenance reporting. The overarching goal of this effort is to address neuroimaging data sharing barriers [2] and continue developing a fundamentally new, granular data exchange standard, called the Neuroimaging Data Model (NIDM), which incorporates provenance as a primitive to document cognitive neuroimaging workflow [3].

NIDM enhances W3C PROV [4] to provide a language that communicates provenance by representing primary data, workflow, and derived data as linked Agent, Activity, and Entity objects. Similar to the way a sentence conveys a standalone thought, NIDM enables provenance statements that express the way a given piece of data was produced.

This abstract presents our preliminary work to represent primary and derived experimental results with three uses cases that aim to 1) automate OpenfMRI dataset analyses, 2) integrate multi-modal measurements of the National Consortium on Alcohol & Neurodevelopment in Adolescence (NCANDA), and 3) capture derived data from a cross species Conte center project.

Methods

NIDM is developed using a community-driven process that engages stakeholders to participate in the identification of use-cases that drive development. In-person workshops and weekly video conferences are used to maintain communication while example NIDM documents, specifications, and software are developed by the INCF NIDASH task force members.

NIDM Experiment captures details about an investigation using the modeling pattern depicted in Figure 1. This pattern models metadata from neuroimaging data management systems at three levels from XCEDE [5]: Project, Study, and Acquisition. The Project level captures administrative information, while the Study and Acquisition levels model MRI sessions and neuropsychological (NP) tests.

Results

Use Case 1: OpenfMRI Database

- Task-based fMRI datasets with standardized organization and metadata
- NIDM Experiment models were developed using an example dataset [6] to capture demographics, primary data, and metadata (e.g., DICOM, Cognitive Atlas [7] terms)
- Models were designed to enable automated analyses

Use Case 2: NCANDA Data Integration

- A multi-site study investigating the effect of alcohol use on the structure and function of developing adolescent brains
- The system architecture developed by the data analysis core of the consortium integrates clinical, NP tests, and survey data [8]
- To enable reporting across these systems, NIDM models were developed to facilitate report generation within the consortium and enhance the metadata available during future data releases

Use Case 3: CONTE Center Cross-Species Modeling

- The Conte center at UC Irvine is studying the effect of early life fragmented maternal care on cognitive vulnerabilities in adolescence across species
- The overall informatics goal is to provide center investigators with inter-project data, including models for maternal and fetal heart rate, fetal movement, maternal blood oxygenations levels, derived DTI and fMRI measurements, and brain connectivity graphs
- A web application was developed to query and visualize data from NIDM object model graphs

Conclusions

NIDM Experiment supports a variety of use cases focused on representing primary and derived data organization with the intent of simplifying data exchange, integration, and sharing. These preliminary results indicate the flexibility of a community-driven data modeling approach and that provenance-based data exchange is being incorporated into a number of projects.

Acknowledgments

We acknowledge the work of all INCF task force members and many other colleagues who have helped in this effort. Further we acknowledge the BIRN coordinating center's (U24RR025736-01) long-standing support and the support by the NCANDA Consortium (U01AA021697-01).

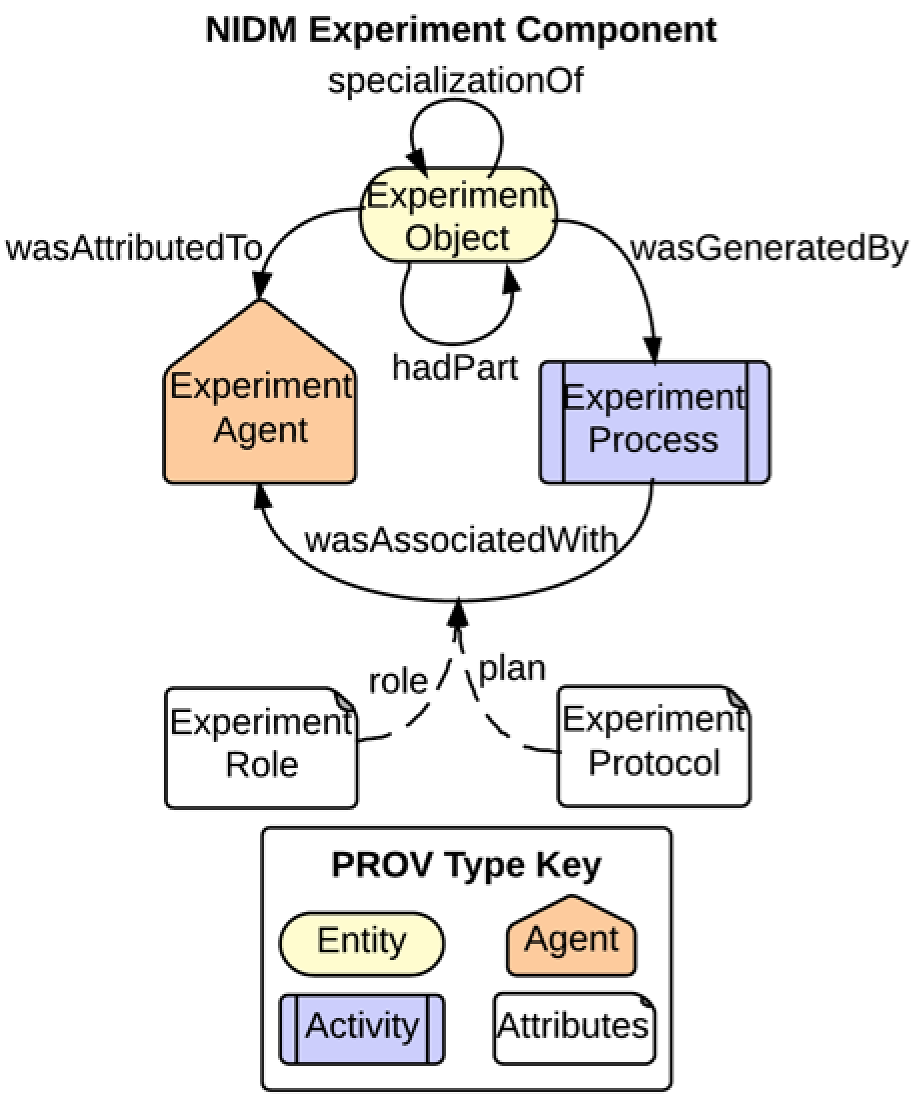

Figure 1. The NIDM Experiment core structure that is extended to model Project, Study, and Acquisition information.

```
# Aquisition Element for Questionnaire
:questionnaire-acquisition-object
    a prov:Entity, nidm:Acquisition, nidm:DemographicsQuestionnaire ;
    # ... repeated attributes
    :gender     "m" ;
    :ageMin     30 ;
    :ageMax     35 ;
    :handedness "r" ;
    # ... other items
    :listen_preference1_genre "Triphop" .
    dcat:downloadURL <http://openfmri.s3.amazonaws.com/ds113/demographics.csv>

# Aquisition Element for MRI

:anatomy-acquisition-object
    a prov:Entity, nidm:Acquisition, nidm:MRIAnatomicalT1 ;
    # ... repeated attributes
    prov:wasGeneratedBy   :mri-acquisition-process ;
    prov:wasAttributedTo  :sub001-study-agent ;
    prov:specializationOf :ds000113-study-object ;
    dcm:Manufacturer      "Philips Medical Systems" ;
    dcm:ModelName         "Achieva" ;
    dcm:EchoTime          5.797 ;
    dcm:SliceThickness    0.7 ;
    # ... additional DICOM attributes
    dcm:NumberOfVolumes   1 ;
    dcat:downloadURL <http://openfmri.../ds113/sub001/anatomy/highres001.nii.gz>
```

Figure 2. Example of an abbreviated NIDM Experiment Acquisition-level model to capture demographic information and a T1-weighted anatomical MRI scan.

References

1. Collins, F. S., & Tabak, L. A. (2014). Policy: NIH plans to enhance reproducibility. *Nature*, *505*(7485), 612–613.
2. Poline, J. B., Breeze, J. L., Ghosh, S., Gorgolewski, K., Halchenko, Y. O., Hanke, M., et al. (2012). Data sharing in neuroimaging research. *Frontiers in Neuroinformatics*, *6*, 9–9. doi:10.3389/fninf.2012.00009
3. Keator, D. B., Helmer, K., Steffener, J., Turner, J. A., Van Erp, T. G., Gadde, S., et al. (2013). Towards structured sharing of raw and derived neuroimaging data across existing resources. *NeuroImage*, *82*, 647–661. doi:10.1016/j.neuroimage.2013.05.094
4. Moreau, L., & Groth, P. (2013). Provenance: An Introduction to PROV. *Synthesis Lectures on the Semantic Web: Theory and Technology*. doi:10.2200/S00528ED1V01Y201308WBE007
5. Gadde, S., Aucoin, N., Grethe, J. S., Keator, D. B., Marcus, D. S., Pieper, S., FBIRN, MBIRN, BIRN-CC. (2012). XCEDE: an extensible schema for biomedical data. *Neuroinformatics*, *10*(1), 19–32. doi:10.1007/s12021-011-9119-9
6. Hanke, M., Baumgartner, F. J., Ibe, P., Kaule, F. R., Pollmann, S., Speck, O., et al. (2014). A high-resolution 7-Tesla fMRI dataset from complex natural stimulation with an audio movie. *Scientific Data*, *1*. doi:10.1038/sdata.2014.3
7. Poldrack, R. A., Kittur, A., Kalar, D., Miller, E., Seppa, C., Gil, Y., et al. (2011). The cognitive atlas: toward a knowledge foundation for cognitive neuroscience. *Frontiers in Neuroinformatics*, *5*, 17. doi:10.3389/fninf.2011.00017
8. Rohlfing, T., Cummins, K., Henthorn, T., Chu, W., & Nichols, B. N. (2013). N-CANDA data integration: anatomy of an asynchronous infrastructure for multi-site, multi-instrument longitudinal data capture. *Journal of the American Medical Informatics Association*, amiajnl–2013–002367. doi:10.1136/amiajnl-2013-002367