Describing Assessments and Experiment Metadata with the Neuroimaging Data Model (NIDM)

Keator David, Helmer Karl, Satrajit Ghosh, Auer Tibor, Camille Maumet, Das Samir, Flandin Guillaume, Nichols Thomas, Krzysztof Gorgolewski, Jessica Turner, et al.

Describing Assessments and Experiment Metadata with the Neuroimaging Data Model (NIDM)

Keator D.B.[1], Helmer K.G.[2], Ghosh S.S.[3], Auer T.[4], Maumet C.[5], Das S.[6], Flandin G.[7], Nichols T.[5,8], Gorgolewski K.J.[9], Turner J.[10], Kennedy D.[11], Poline J.B.[12],Nichols B.N.[13,14].

1. Department of Psychiatry and Human Behavior, University of California, Irvine, CA, USA.
2. Massachusetts General Hospital, Boston, MA.
3. Massachusetts Institute of Technology, MA, USA.
4. MRC Cognition and Brain Sciences Unit, Methods Group, United Kingdom.
5. University of Warwick, Warwick Manufacturing Group, Coventry, United Kingdom.
6. Montreal Neurological Institute, Montreal, Canada
7. Wellcome Trust Centre for Neuroimaging, UCL Institute of Neurology, London, United Kingdom.
8. University of Warwick, Department of Statistics, Coventry, United Kingdom.
9. Psychology, Stanford University, Stanford, CA, USA
10. Psychology and Neuroscience, Georgia State University, Atlanta, GA, USA
11. University of Massachusetts Medical School, Department of Psychiatry, MA, USA
12. Helen Wills Neuroscience Institute, Brain Imaging Center, University of California, Berkeley, CA, USA
13. Neuroscience Program, SRI International, Menlo Park, CA, USA
14. Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, CA, USA

Introduction

A fundamental goal of organizing and annotating scientific data is to provide investigators with an effective mechanism to access and share data with the wider scientific community. To attain these goals, the data needs to have both adequate descriptions to make it useable [1] and be represented in a structured form, accessible to computational tools. The Neuroimaging Data Model (NIDM; http://nidm.nidash.org/) is an ongoing effort to represent, in a single data model, the different components of a research activity (e.g., participants, project information, derived data), their relations, and provenance [2]. NIDM has a modular design currently consisting of NIDM-Results, NIDM-Workflows, and NIDM-Experiment. NIDM-Workflows is focused on the description of data analysis pipelines and detailed software-specific variations. NIDM-Results is focused on the representation of mass-univariate neuroimaging results using a common descriptive standard across neuroimaging software packages. NIDM-Experiment is focused on the representation of the experiment design, the source data collected during the experiment and information on the participants, including generic assessments, demographics, and visit information. This abstract presents our prototyping work on representing assessments and metadata collected during the course of a typical neuroimaging experiment.

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 is an extension of the W3C recommended Provenance Data Model (PROV-DM ;www.w3.org/TR/prov-dm/) for derived data and experiment descriptions in neuroimaging related fields. NIDM Experiment captures details about an investigation (Figure 1). This pattern models metadata from neuroimaging data management systems at three levels: Investigation,

Session, and Series. The Investigation level captures administrative information, while the Session and Series levels model assessments and MRI sessions. The NIDM-Assessment object model (Figure 2) provides a generic model for describing the data dictionary and acquired assessment data. Using the NIDM-Assessment model, an assessment "DataStructure" is composed of individual "DataElements" corresponding to the questions in the assessment. When an assessment's question contains multiple response choices, it is coded as a "ValueSet" in NIDM-Assessment and documents the answer choices and their coded values (Figure 3). Data collected for a particular assessment are stored as acquisition objects of the assessment type described by the corresponding "DataStructure". These objects are associated with an acquisition activity found at the Session or Series levels of the NIDM-Experiment model, depending on when the assessment was administered. The NIDM-Assessment representations are serialized using the Terse RDF Triple Language (http://www.w3.org/TeamSubmission/turtle/) and queried using SPAQRL (https://www.w3.org/TR/rdf-sparql-query/).

Results

The NIDM-Experiment and NIDM-Assessment models are actively under development and currently being evaluated by mapping multiple datasets (e.g. FBIRN [3], OpenfMRI [4], NCANDA [5], etc.) into the structures. Our examples are made available through the NIDM GitHub respository: https://github.com/incf-nidash/nidm. The Conte Center on Brain Programming in Adolescent Vulnerabilities at the University of California, Irvine (http://contecenter.uci.edu) has made use of the early NIDM-Experiment related models in building its production informatics resource, which provides 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. Our initial experiments have shown the model to be flexible and applicable to a wide range of heterogenous data types.

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. Our preliminary results indicate a modeling approach based on the W3C PROV-DM and NIDM is appropriate for modeling the heterogeneous data often collected in the course of a neuroimaging-related investigation and demonstrates the benefit of using semantic-web methods for data annotation. Continued work will include finalizing the NIDM-Experiment model and creating applications to aid investigators in converting their existing data to the NIDM-Experiment representation, and using the representations in analysis tools and information management applications.

Acknowledgments

We acknowledge the work of all INCF task force members and many other colleagues who have helped in this effort. Further, we acknowledge support by the UCI Conte Center (1P50MH096889-01A1), the NCANDA Consortium (U01AA021697-01), and the ReproNim Center for Reproducible Neuroimaging Computation (1P41EB019936-01A1).

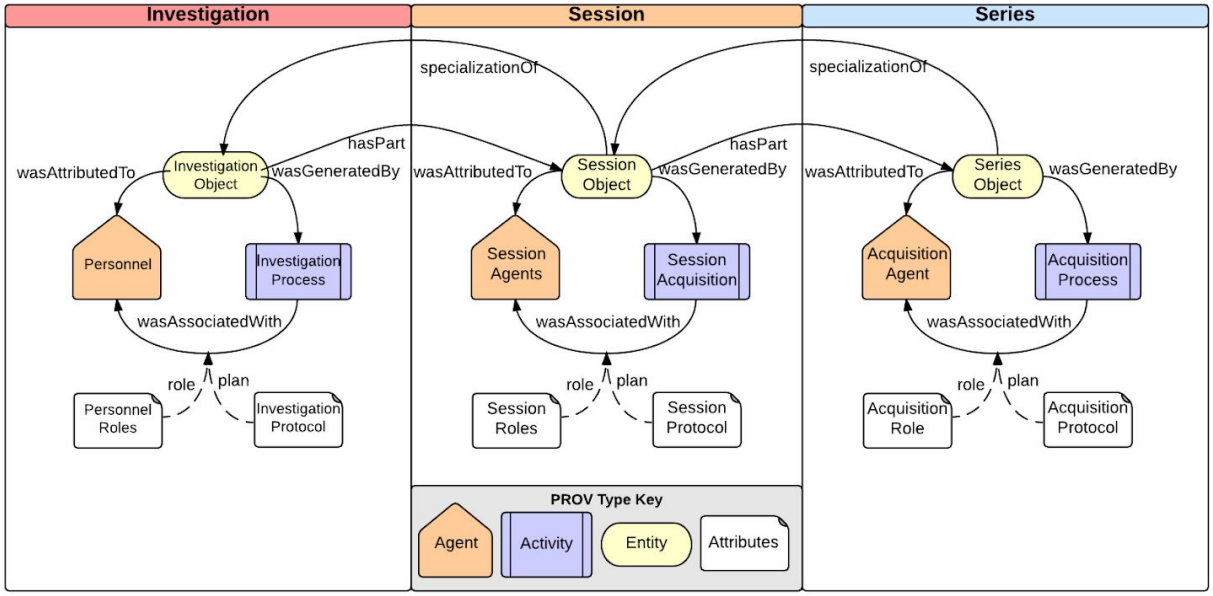

Figure 1: The NIDM-Experiment core structure. The Investigation level describes the overall project including investigators and project personnel, a description of the project, consent forms, etc. The Session level describes the data acquisition activities for a subject visits. The Series level describes the imaging series and behavorial/clinical assessments administered during a session.

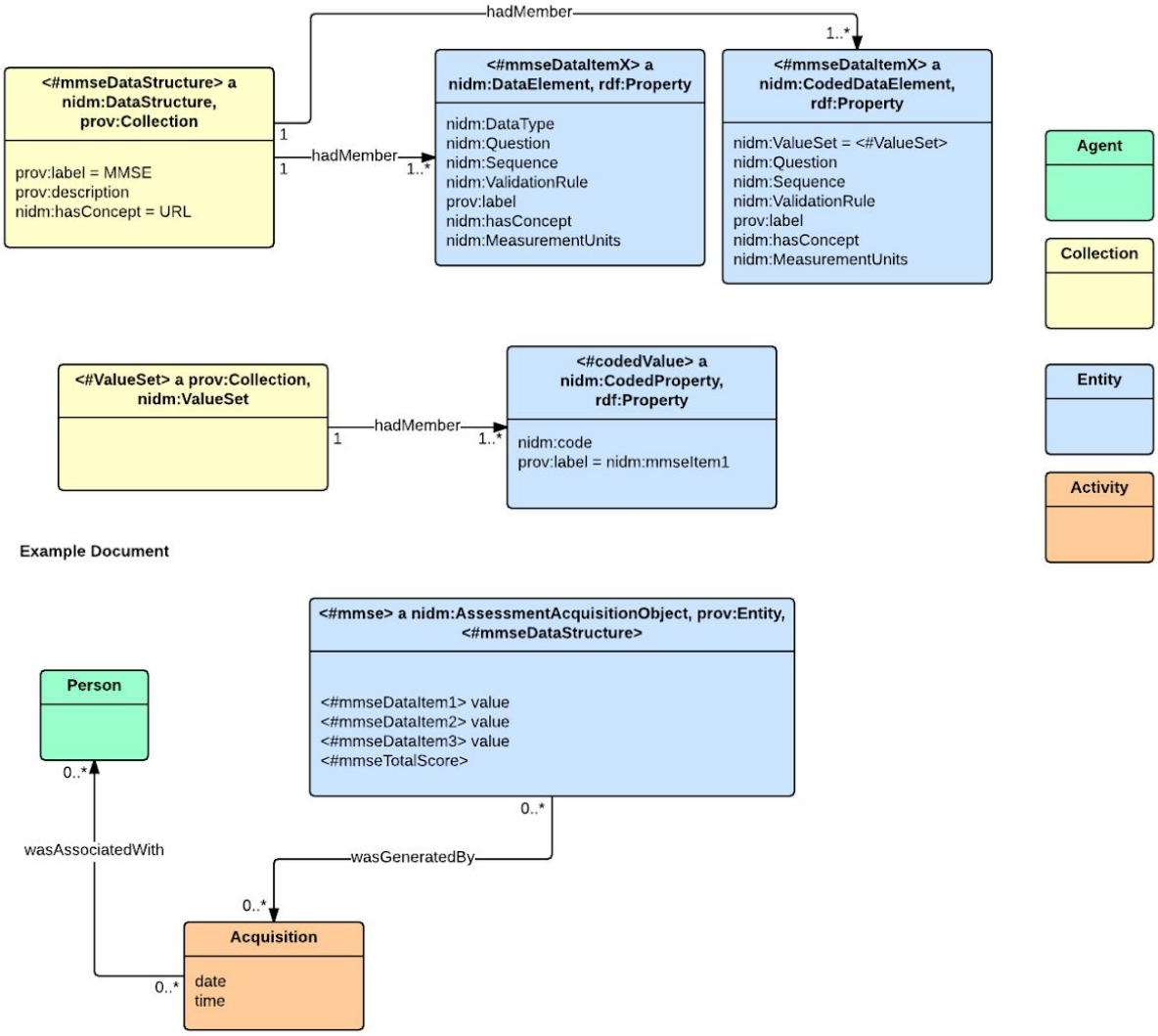

Figure 2: The NIDM-Assessment object model. An assessment consists of a DataStructure entity describing the overall assessment and one or more DataElement and/or CodedDataElements, describing the assessment questions. CodedDataElements are described using ValueSets.

```
hid:demographics_phaseII
        a                   nidm:DataStructure ;
        hid:assessmentID    "28984"^^xsd:int ;
        prov:hadMember      hid:Race, hid:MothersEducation , hid:LivingArrangement , hid:Age ,
hid:FathersEducation ,hid:NumberofChildren , hid:Ethnicity , hid:Handedness , hid:MaritalStatus ,
hid:FathersOccupation , hid:Education , hid:MothersOccupation , hid:Occupation , hid:Gender ;
        prov:label          "demographics(phaseII)" .

hid:Race  a             rdf:Property , nidm:DataElement ;
        nidm:DataType   "varchar" ;
        nidm:Question   "Race" ;
        nidm:ValueSet   hid:demographics_phaseII__Race_ValueSet ;
        hid:Sequence    "1"^^xsd:int .

hid:demographics_phaseII__Race_ValueSet
        a               prov:Collection , nidm:ValueSet ;
        prov:hadMember  hid:Race_1, hid:Race_2 ,hid:Race_3.

hid:Race_1  a       rdf:Property , nidm:CodedProperty ;
        nidm:code   "1"^^xsd:int ;
        prov:label  "American Indian/Alaska Native" .
hid:Race_2  a       rdf:Property , nidm:CodedProperty ;
        nidm:code   "2"^^xsd:int ;
        prov:label  "Asian" .
hid:Race_3  a       rdf:Property , nidm:CodedProperty ;
        nidm:code   "3"^^xsd:int ;
        prov:label  "Native Hawaiian or Other Pacific Islander" .
```

Figure 3: Partial NIDM representation of demographics assessment data dictionary.

References

1. 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
2. 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
3. Keator DB, van Erp TG, Turner JA, Glover GH, Mueller BA, Liu TT, Voyvodic JT, Rasmussen J, Calhoun VD, Lee HJ, Toga AW. The Function Biomedical Informatics Research Network Data Repository. NeuroImage. 2016 Jan 1;124:1074-9.
4. Poldrack RA, Barch DM, Mitchell JP, Wager TD, Wagner AD, Devlin JT, Cumba C, Koyejo O, Milham MP. Toward open sharing of task-based fMRI data: the OpenfMRI project.
5. 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.