

The best of both worlds: using semantic web with JSON-LD

An example with NIDM-Results & Datalad

Camille Maumet¹, Satrajit Ghosh², Yaroslav O. Halchenko³, Dorota Jarecka², Nolan Nichols⁴, Jean-Baptiste Poline⁵, Michael Hanke⁶

1. Inria, Univ Rennes, Inserm, CNRS, Rennes, France. 2. Massachusetts Institute of Technology, Cambridge, MA, USA. 3. Dartmouth College, NH, USA. 4. Bioinformatics and Computational Biology, Genentech Inc, CA, USA. 5. McGill University, Montreal, Canada, and UC Berkeley, CA, USA. 6. Institute of Neuroscience and Medicine, Brain & Behaviour (INM-7), Research Centre Jülich, Jülich, Germany and Institute of Systems Neuroscience, Medical Faculty, Heinrich Heine University Düsseldorf, Düsseldorf, Germany.

Introduction

NIDM-Results provides a harmonised representation for fMRI results reporting using Semantic Web technologies (Maumet et al. 2016). While those technologies are particularly well suited for aggregation across complex datasets, using them can be costly in terms of initial development time to generate and read the corresponding serialisations. The technology is machine accessible, but it can be difficult to comprehend by humans. This hinders adoption by scientific communities and by software developers used to more-lightweight data-exchange formats, such as JSON.

JSON-LD: a JSON representation for semantic graphs ("JSON-LD 1.1" n.d.) was created to address this limitation and recent extensions to the specification allow creating JSON-LD documents that are structured more similar to simple JSON. This representation is simultaneously readable by a large number of JSON-based applications and by Semantic Web tools.

Here we review our work on building a JSON-LD representation for NIDM-Results data and exposing it to Datalad (Halchenko et al. 2018), a data-management tool suitable for neuroimaging datasets with built-in support for metadata extraction and search.

Methods

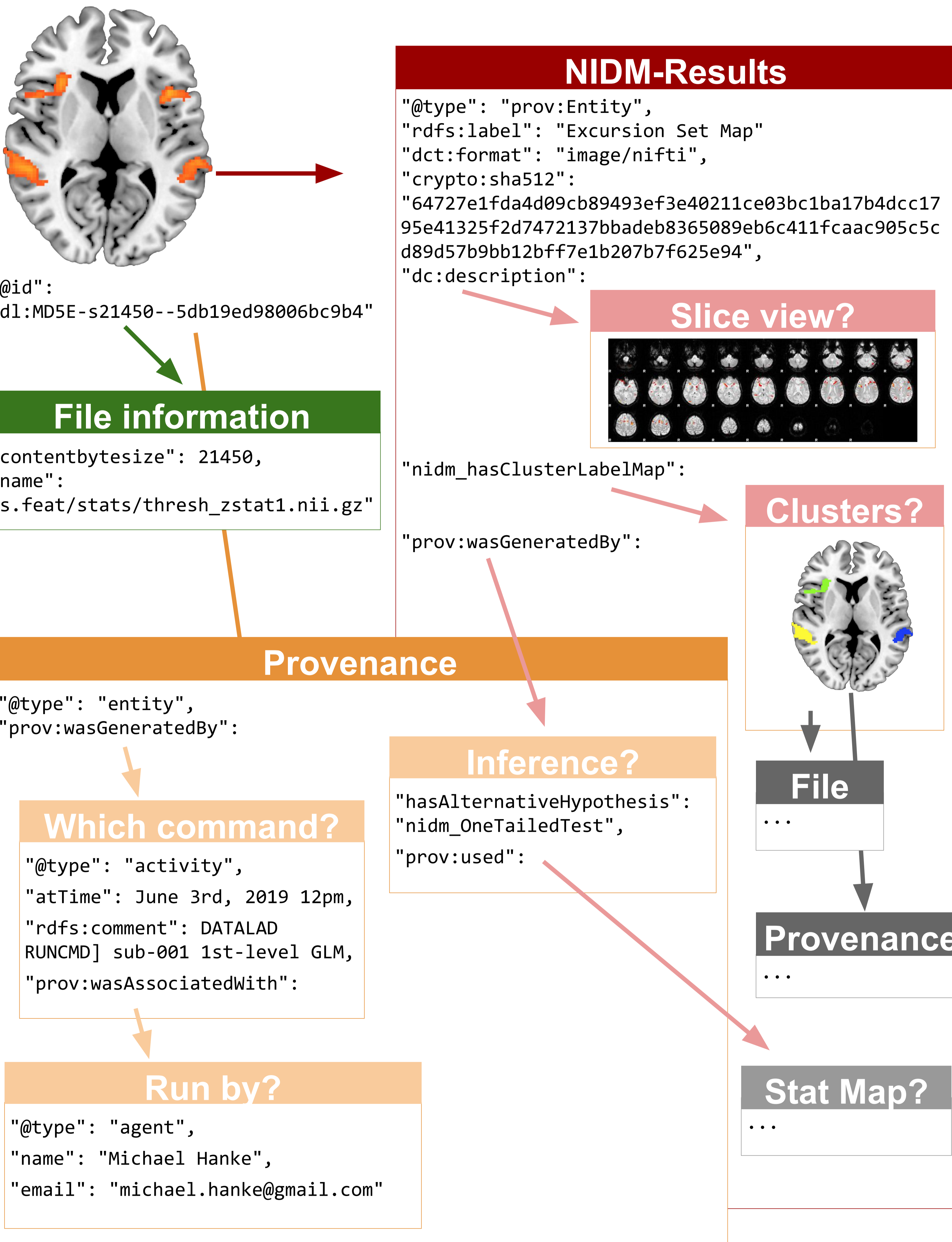
This work was initiated at two collaborative events: the "INCF Neuroinformatics Hackathon 2018" August 7-8 in Montreal, Canada and "DC Code Convergence" September 10-14 in Washington, DC, United States.

Using JSON-LD 1.1, we developed a context file that can be used to transform the original semantic web file into a simple JSON representation while preserving the PROV data model. "Type indexing", a JSON-LD 1.1 feature, enables indexing objects by their types instead of their identifiers.

Results

The JSON-LD context is available at: <http://purl.org/nidash/context>. The metadata extractor for Datalad is available at: <https://github.com/datalad/datalad-n-euroimaging/pull/41>. Finally, a set of examples of JSON-LD files are included at: <https://github.com/incf-nidash/nidm-specs/pull/479>.

Connecting and interrogating fMRI results, their provenance and file information w/ NIDM & Datalad



Acknowledgments

The authors would like to thank the organizers and attendees of the INCF Neuroinformatics Hackathon, 2018, August 7-8 in Montreal, Canada and DC Code Convergence, 2018, September 10-14 in Washington, DC, United States. SG, DJ, JBP, YH were supported by ReproNim (NIH-NIBIB P41 EB019936). DataLad development was supported by the NSF (CRCNS 1429999), German federal state of Saxony-Anhalt and the European Regional Development Fund (ERDF), Project: Center for Behavioral Brain Sciences, Imaging Platform. JBP was partially funded by NIH-NIBIB P41 EB019936 (ReproNim) NIH-NIMH R01 MH083320 (CANDIShare) and NIH 5U24 DA039832 (NIF), as well as the Canada First Research Excellence Fund, awarded to McGill University for the Healthy Brains for Healthy Lives initiative.

References

Halchenko, Yaroslav, Michael Hanke, Benjamin Poldrack, Kyle Meyer, Debanjum, Gergana Alteva, Jason Gors, et al. 2018. DataLad/datalad 0.11.0. <https://doi.org/10.5281/zenodo.1470735>.
"JSON-LD 1.1." n.d. Accessed December 14, 2018. <https://json-ld.org/spec/latest/json-ld/>.
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). <https://doi.org/10.1038/sdata.2016.102>.

Results (continued)

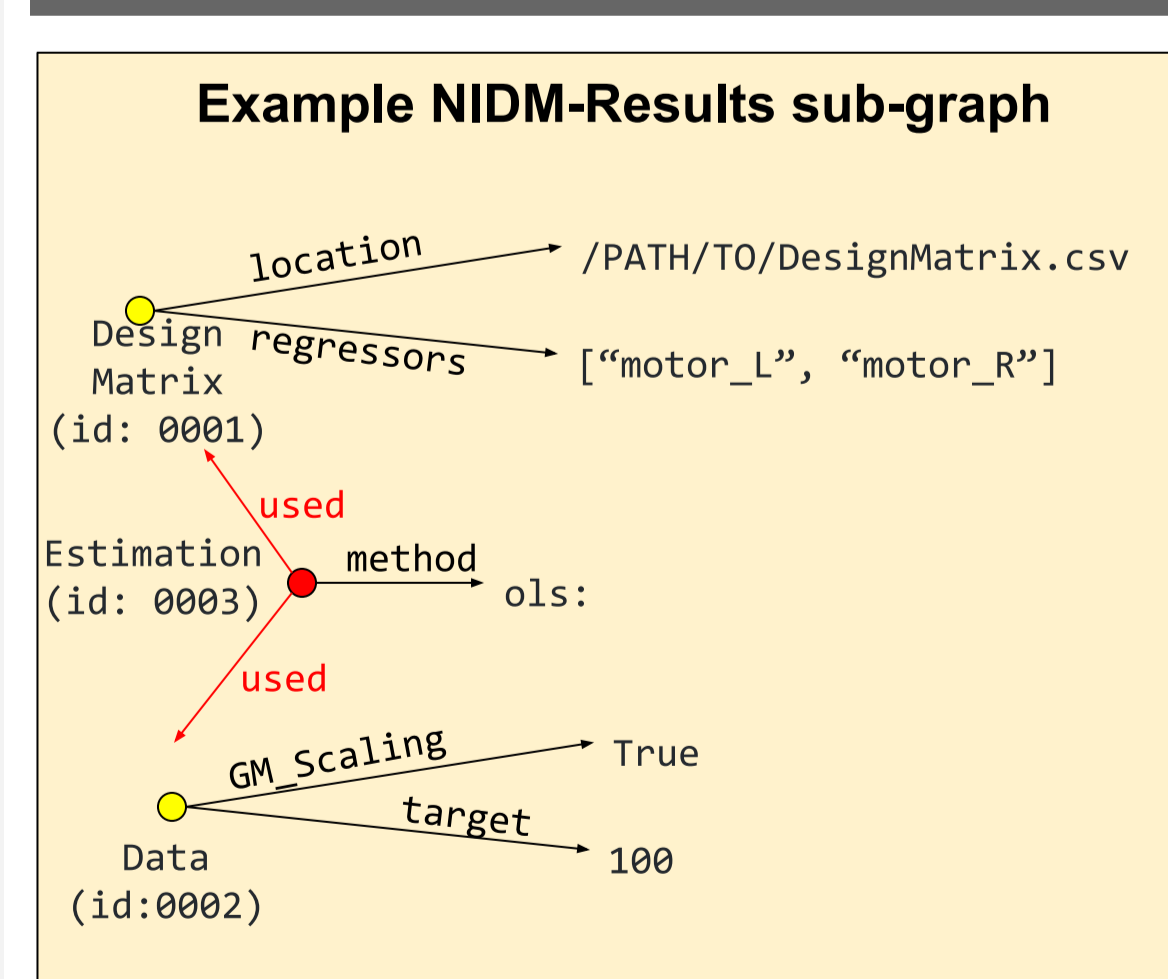
The central figure presents an exemplary invocation and result of a DataLad search, demonstrating that the proposed metadata representation is suitable for processing with simplest text matching procedures. Programmatic "understanding" of term semantics or constructing a true graph representation is not required.

Accessible metadata representations maximize utility across use cases. Here, DataLad automatically binds NIDM metadata to precise identity and version information, as well as the exact state of any other content of a dataset.

Conclusions

We have proposed a new serialisation for NIDM-Results using JSON-LD that is both easier to read by standard applications and equivalent to the original data model, i.e. preserving provenance, the graph structure, and merge/query capabilities. Key property of the developed metadata representation is that a large fraction of the information is programmatically accessible by direct access of a shallow dictionary-like key store. For any metadata use case where relevant term labels are known, this enables straightforward and complete utilization through generic applications like DataLad, without the otherwise unavoidable technical and conceptual overhead of semantic graphs.

Fig. 1. Example of JSON-LD file using type indexing on a simplified NIDM-Results graph made of 3 objects: "design matrix", "data" and "estimation".



Example JSON-LD

```
{  
  "@context": "http://purl.org/nidash/context"  
  "graph": {  
    "designMatrix": {  
      "@id": "niiri:0001",  
      "location": "/PATH/TO/DesignMatrix.csv",  
      "regressors": ["motor_L", "motor_R"]  
    },  
    "data": {  
      "@id": "niiri:0002",  
      "grandMeanScaling": "true",  
      "targetIntensity": "100"  
    },  
    "estimation": {  
      "@id": "niiri:0003",  
      "method": "ols",  
      "used": ["niiri:0001", "niiri:0002"],  
    }  
  }  
}
```