From metadata to the semantic web: services for data annotation and findable data

Michael Sonntag[1], Achilleas Koutsou[1], Jan Grewe[2], Thomas Wachtler[1]

[1]German Neuroinformatics Node, Department Biologie II, Ludwig-Maximilians-Universität München, Germany;
[2]Institut für Neurobiologie, Universität Tübingen, Germany

Overview

Annotation of research data with metadata is crucial to provide context for analysis and re-use. The odML[1] format (RRID:SCR_001376) offers a flexible and comprehensive solution for the scientist to collect and organize metadata in a structured form that is both human readable and machine actionable[2] for documentation and automated analysis. To further support the FAIR principles[3], we present tools to export metadata from odML to RDF[4], which opens metadata up to semantic web services. The G-Node SPARQL server[5] is aimed at providing searchable whole metadata sets for meta analyses and also providing links to the actual published scientific data set.

Scientists can upload their metadata to make their data findable and accessible even if it was a data publication or if it is an unpublished data set. Furthermore, the GIN[6] data hosting service (RRID:SCR_015864) provides an opt-in feature to automatically update the metadata service when changes to a dataset occur, to ensure the metadata is always up-to-date.
Finally with a metadata resource service the G-Node hosts a platform providing terminologies for metadata annotation and features a forum for general feedback, usage discussions and exchange of metadata templates with the scientific community.

Templates and Terminologies

odML features and hosts two concepts for consistent metadata management you can discover, re-use or share your own with your colleagues:

Templates - https://templates.g-node.org

Templates are pre-defined metadata building blocks that can be re-used in recurring experiments or data pipelines to automatically document equipment, software or subjects.

Terminologies - https://terminologies.g-node.org

odML facilitates and encourages standardization by providing and linking to terminologies. Entities in an odml-file can be to such a terminology. Terms can be imported from Interlex[7] or from G-Node hosted terminologies.

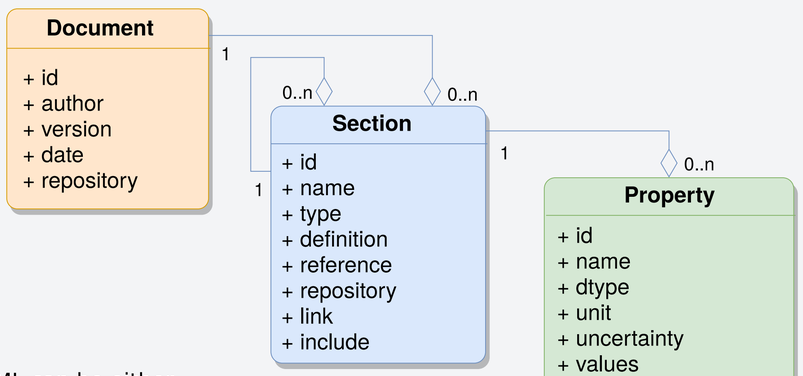
Get Involved!

If you have terminologies or templates e.g. for electrophysiology setup that you are using on a regular basis and want to share with and enable other scientists to use, consider contributing on github or contact us:

github.org/G-Node/terminologies
github.org/G-Node/templates

dev@g-node.org

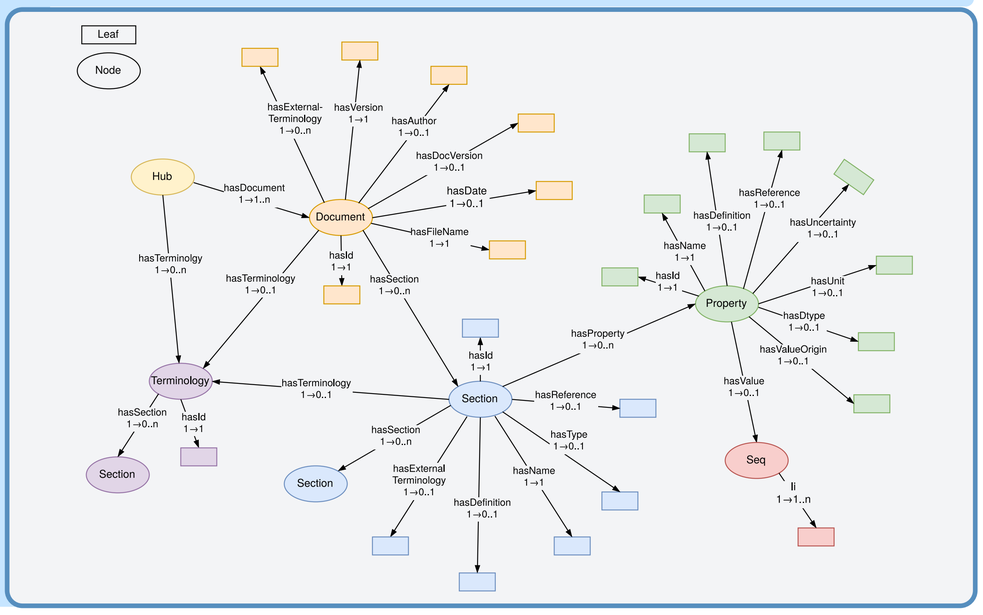
odML Data Model



odML can be either
- written directly to flatfiles using the python-odml implementation, or
- attached to the data and written to an integrated representation using nixpy.

Both format variants are interoperable.

RDF Representation of odML Data Model



Resources and References

References
[1] Grewe et al (2011); doi:10.3389/fninf.2011.00016
[2] Zehl et al (2016); doi:10.3389/fninf.2016.00026
[3] Wilkinson et al (2016); doi:10.1038/sdata.2016.18
[4] https://www.w3.org/RDF/
[5] https://meta.g-node.org
[6] https://gin.g-node.org
[7] https://neuinfo.org/interlex/dashboard

Find out more about G-Node projects at
https://g-node.github.io
https://templates.g-node.org/
https://terminologies.g-node.org/
https://meta.g-node.org
https://github.com/G-Node/python-odml
https://github.com/G-Node/nixpy

Poster presented at
INCF Neuroinformatics 2019
Warsaw, Poland

Federal Ministry of Education and Research

Supported by BMBF grants 01GQ1302 and 01GQ1509

Contact: dev@g-node.org