Handling neuroscientific data with GIN

Present and Future

Michael Sonntag
sonntag@bio.lmu.de

German Neuroinformatics Node
Department Biology II
Ludwig-Maximilians-Universität München

Slides
https://gin.g-node.org/G-Node/PublicResources

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

G-Node German Neuroinf

A brief introduction of the G-Node

http://www.g-node.org

- Funded by BMBF
- Hosted at the Ludwig-Maximilians Universitaet Muenchen
- Bernstein network of Computational Neuroscience
- Associated INCF Node

Thomas Wachtler

funded by
Federal Ministry of Education and Research

Development of Tools for Efficient Data Management

- Well-defined **data model** for neuroscience data that accounts for all types of recorded data
- Flexible methods for **data annotation** and metadata management that can be adapted to the requirements of the experiment and laboratory
- Format and tools for **integrated organization of data and metadata**, including interfaces for common tools and languages, to facilitate data access, data management, and data analysis
- Development and hosting of **data management infrastructure** for **collaboration** and **data sharing**.

Development of Tools for Efficient Data Management

odML (open metadata Markup Language)

http://www.g-node.org/odml

NIX (Neuroscience Exchange format)

http://www.g-node.org/nix

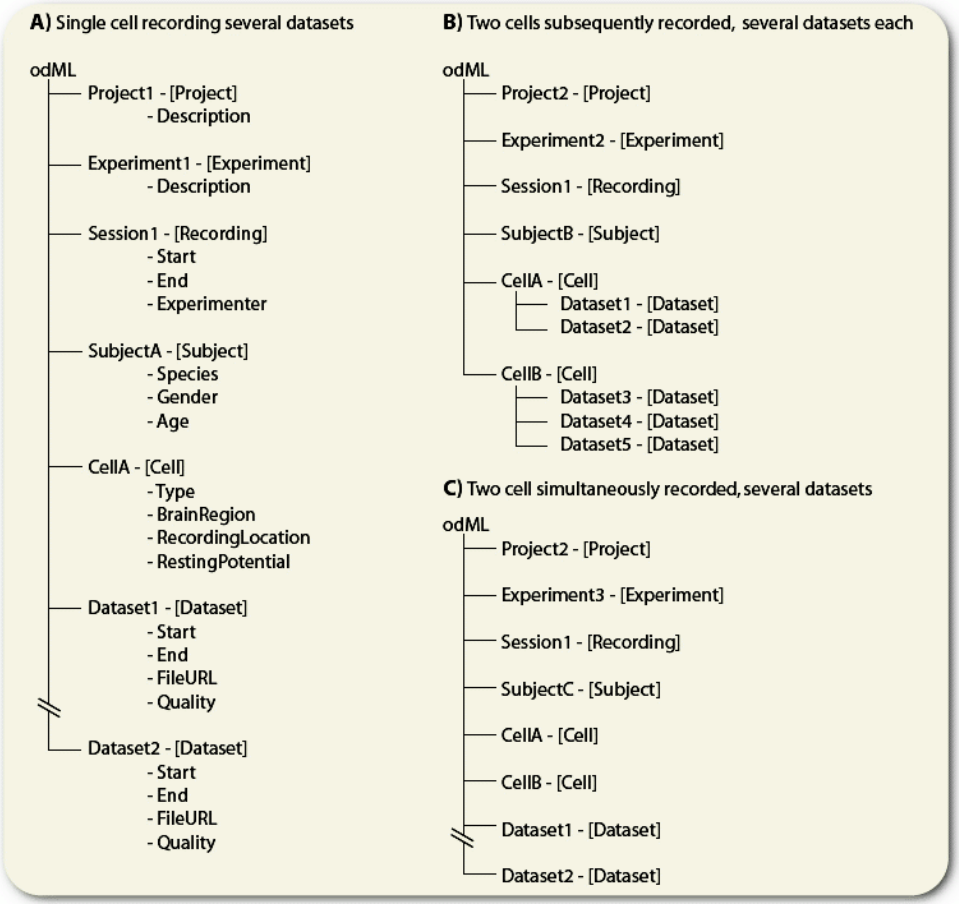
GIN (G-Node Infrastructure services)

https://gin.g-node.org

odML (open metadata Markup Language)

http://www.g-node.org/odml

- format: hierarchical structure of key-value pairs: simple, flexible, inherently extensible ➔ can be adapted to the specifics of the lab or experiment
- can carry **any metadata**
- machine write- and readable, **facilitates automated collection** of metadata in the laboratory
- community-driven standardization through shared terminologies [1]



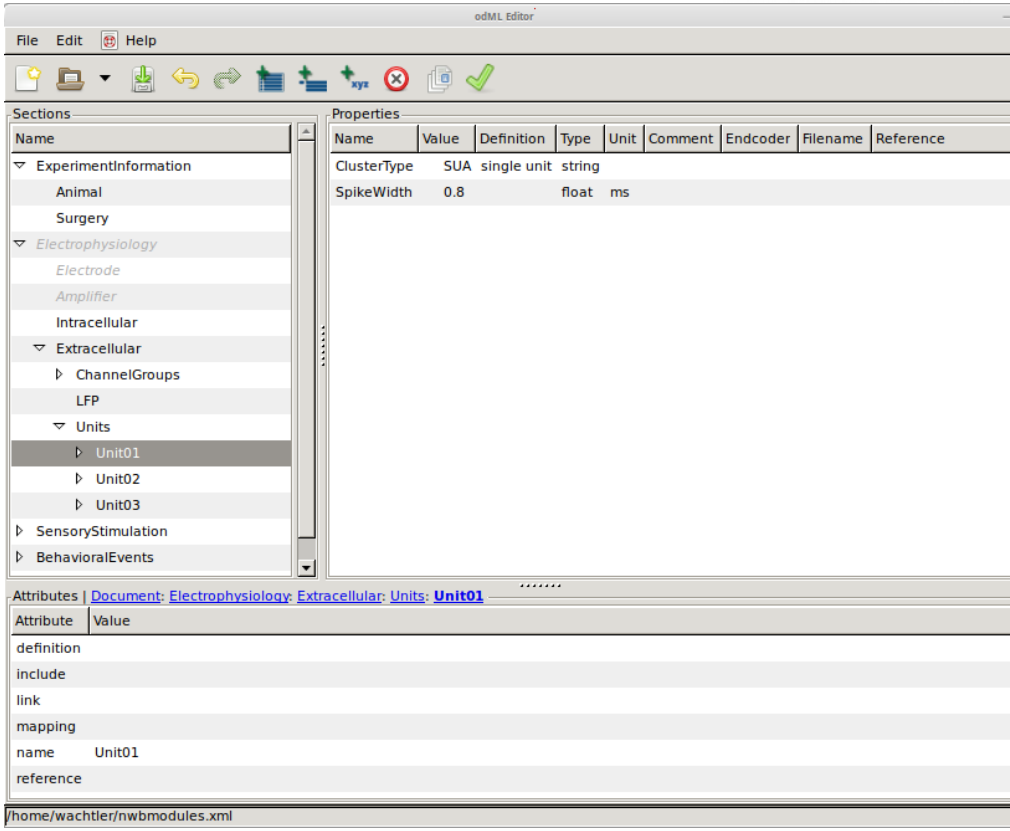
[1] https://github.com/G-Node/odml-terminologies

Grewe et al (2011) Front. Neuroinf. 5:16

odML: Tools

http://www.g-node.org/odml

- Python core library [1]
  enables integration into software tools
- odml-UI editor [2]
- odmltables plugin [3]
  (by INM-6 FZ Jülich)



[1] https://github.com/G-Node/python-odml
[2] https://github.com/G-Node/odml-ui
[3] https://github.com/INM-6/python-odmltables

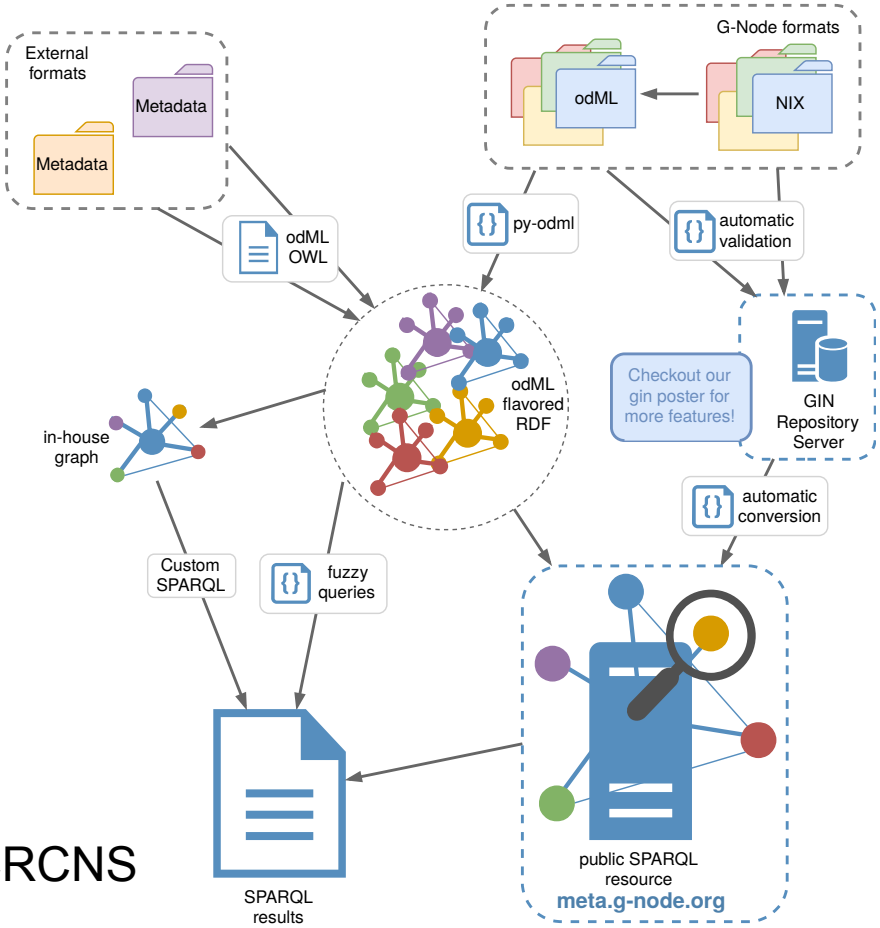
Latest odML format developments

http://www.g-node.org/odml

- streamlined data format (1.3 → 1.4)
- support for YAML and JSON
- export to RDF
- prototype odML flavoured Apache Jena Server
- “fuzzy queries”: experimental abstract SPARQL language

odML flavoured RDF workflow

http://www.g-node.org/odml



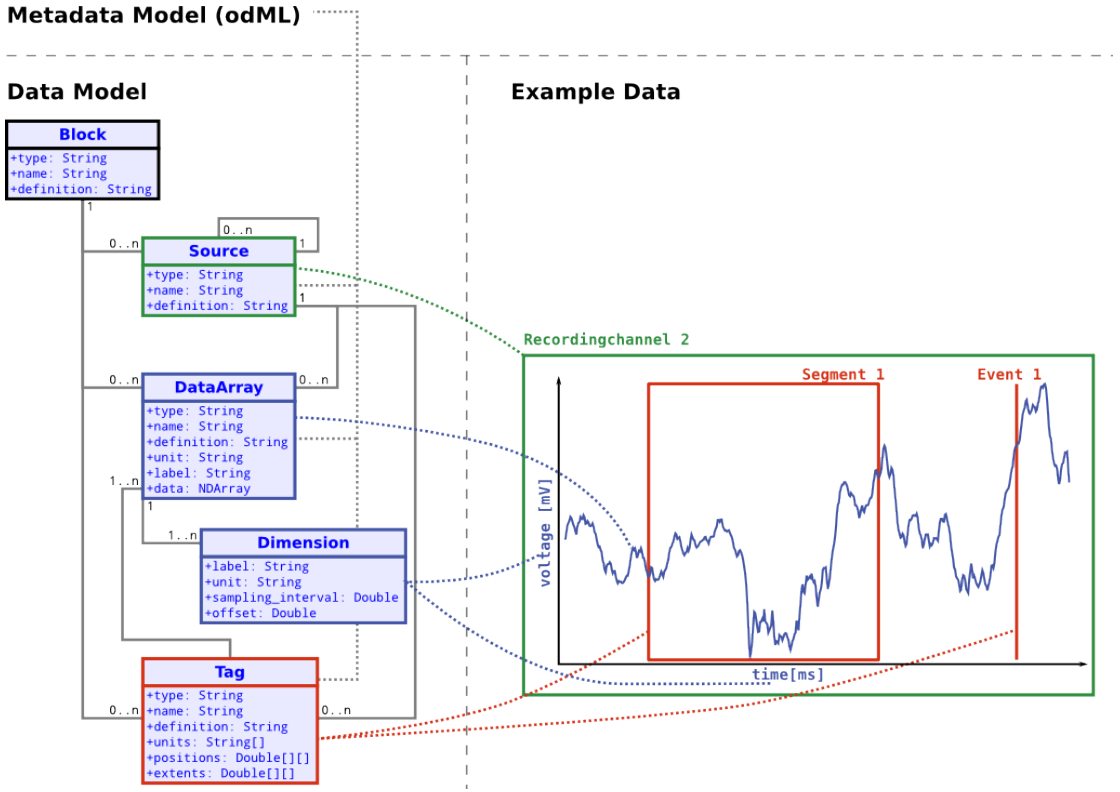
Joint project with the CRCNS

https://crcns.org/

NIX – Format for integration of data and metadata

http://www.g-node.org/nix

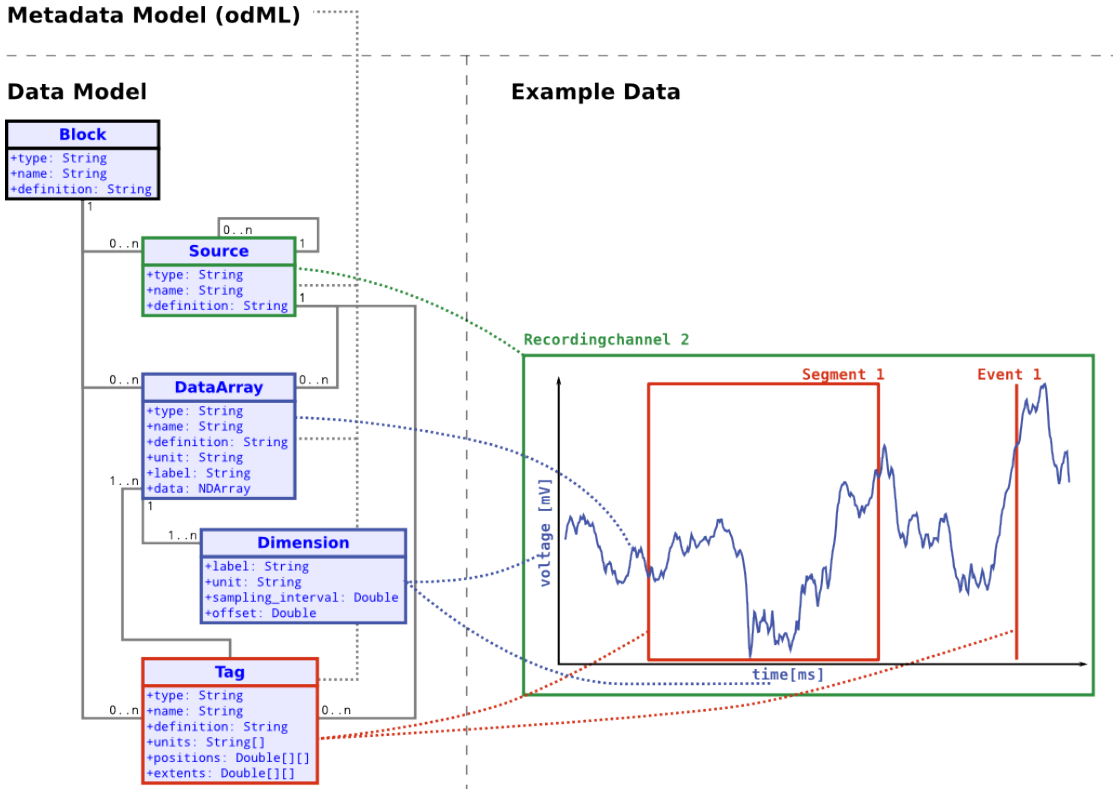
- general data model (derived from **Neo**) to represent recorded data, derived data, relations of data
- flexible data model for metadata (**odML**) for comprehensive annotation of data
- file backend: HDF5 file format
  - structure reflects data model, easy to understand
  - other storage backends possible
- libraries for different languages (C++, Python, Matlab, Java)



NIX – Format for integration of data and metadata

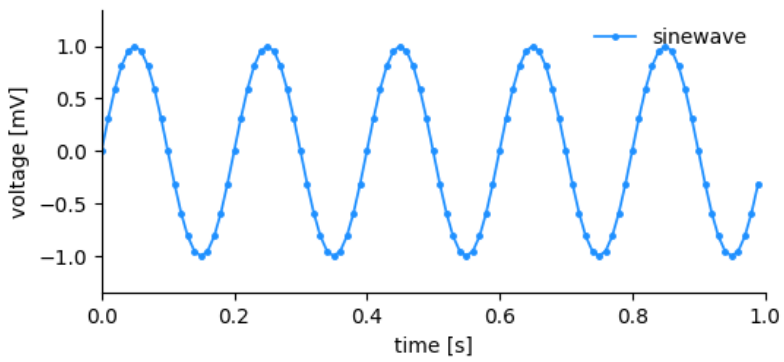
http://www.g-node.org/nix

- storage of n-dimensional data
- definition of regions of interest
- linkage of multiple steps of analysis
- storage of data annotation with data

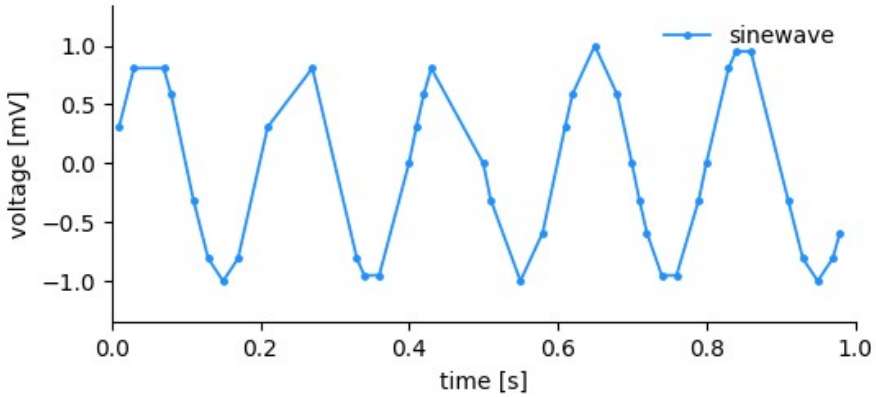


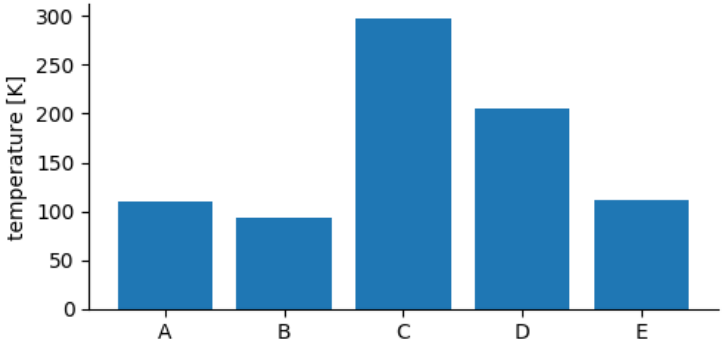
NIX features – Dimensions and tagging

Sampled dimension

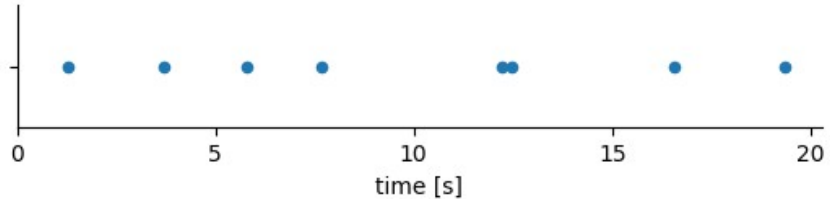


Range dimension





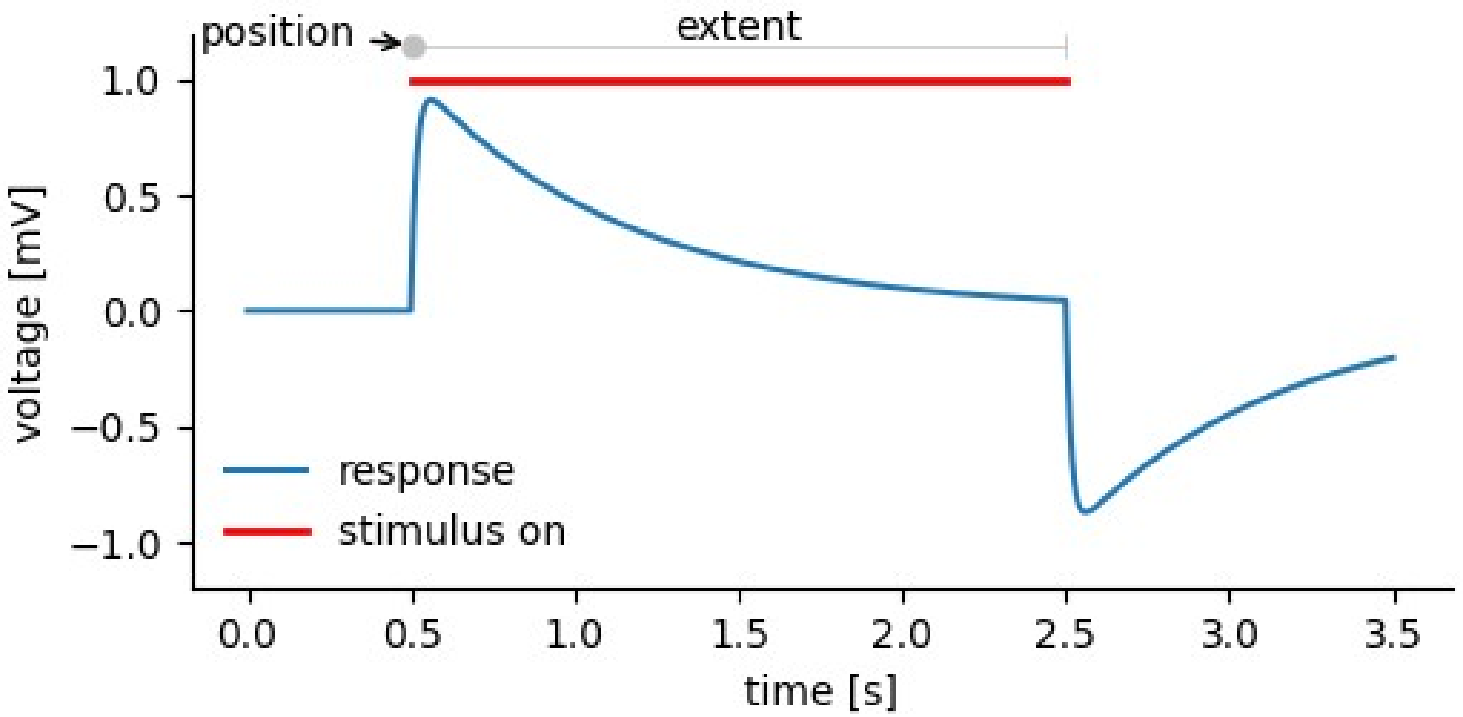
Set dimension



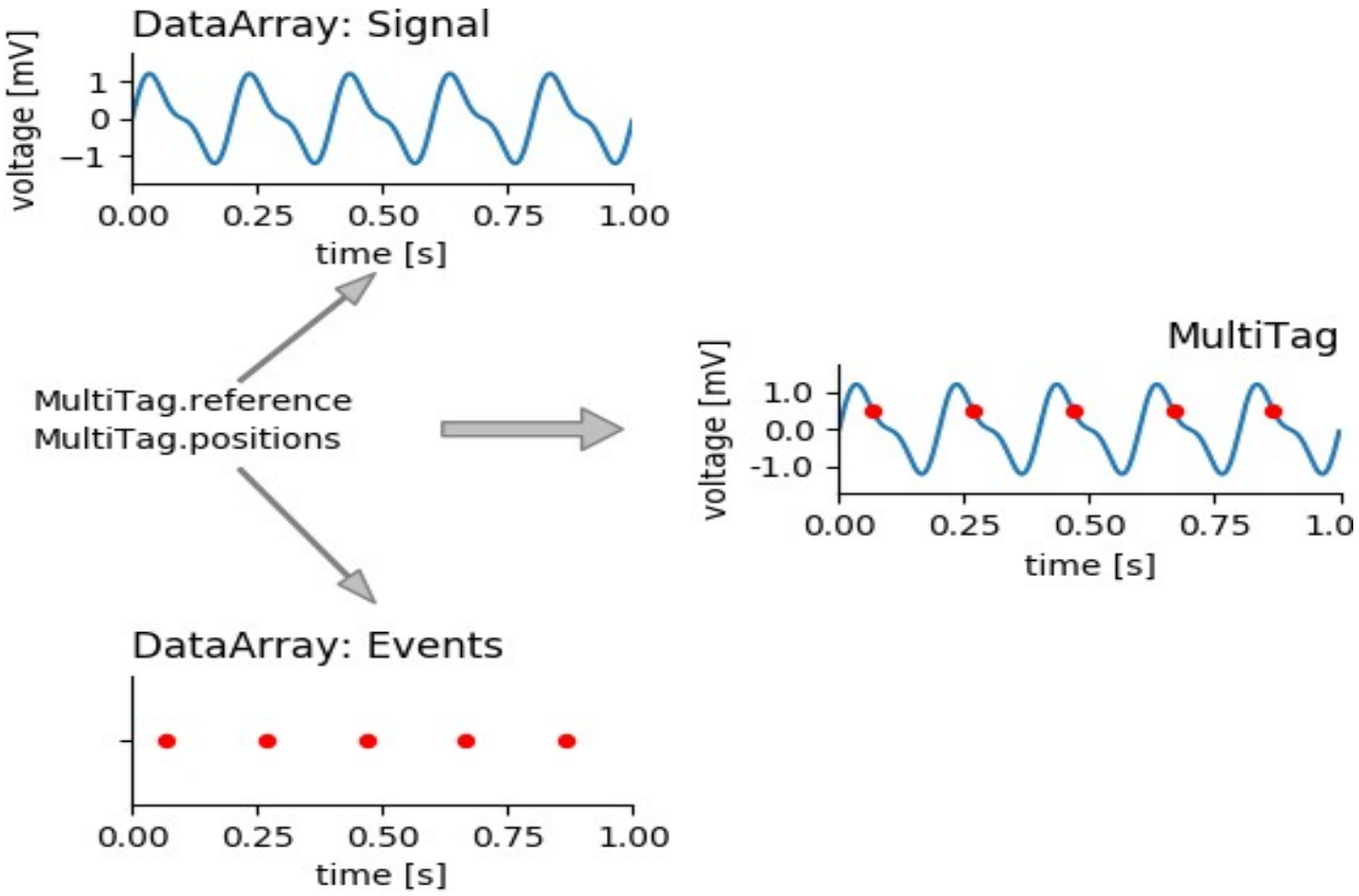
Alias Range dimension

Features automatic SI unit conversion

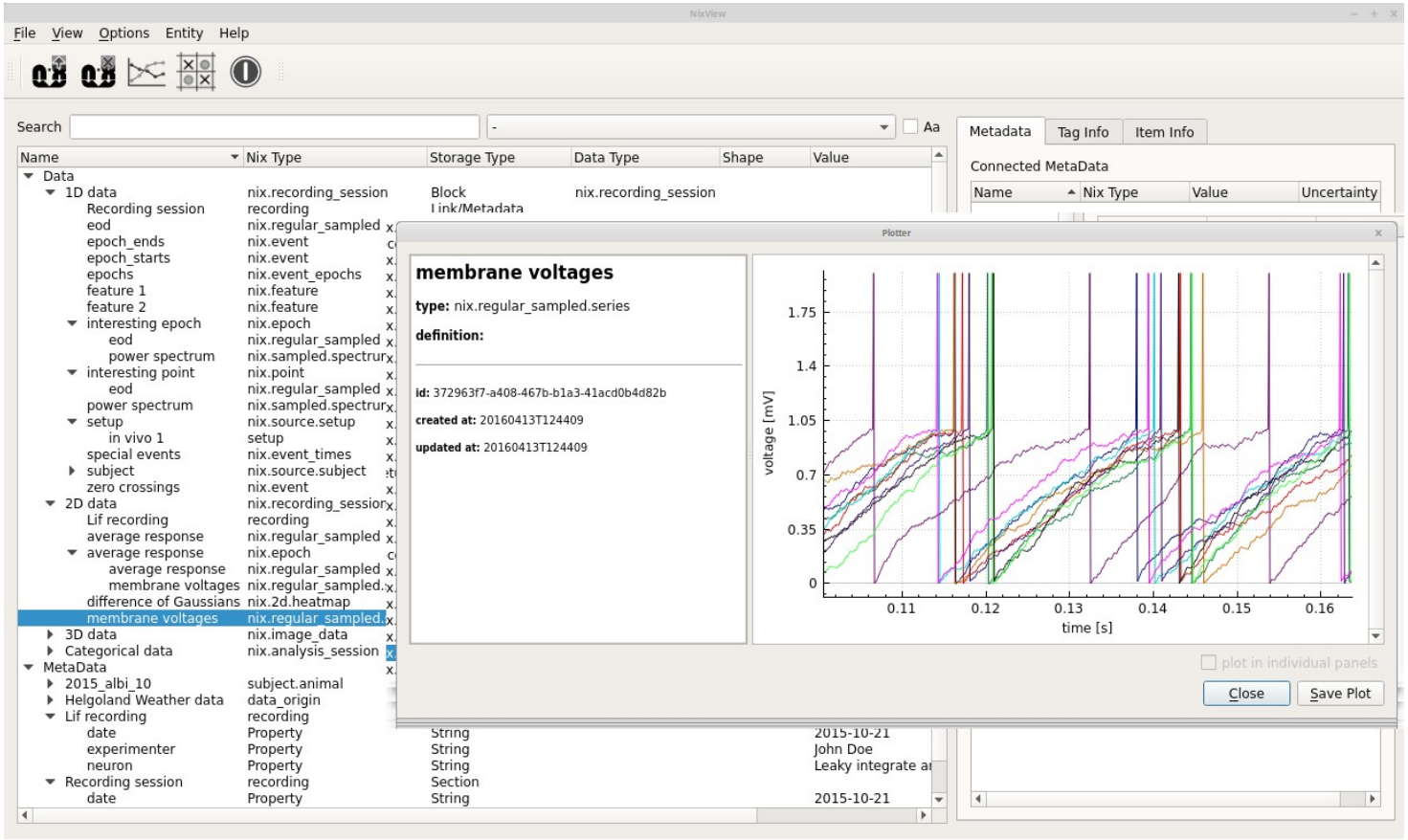
NIX features – Dimensions and tagging



NIX features – Dimensions and tagging



NixView



https://github.com/bendalab/NixView

Latest NIX developments

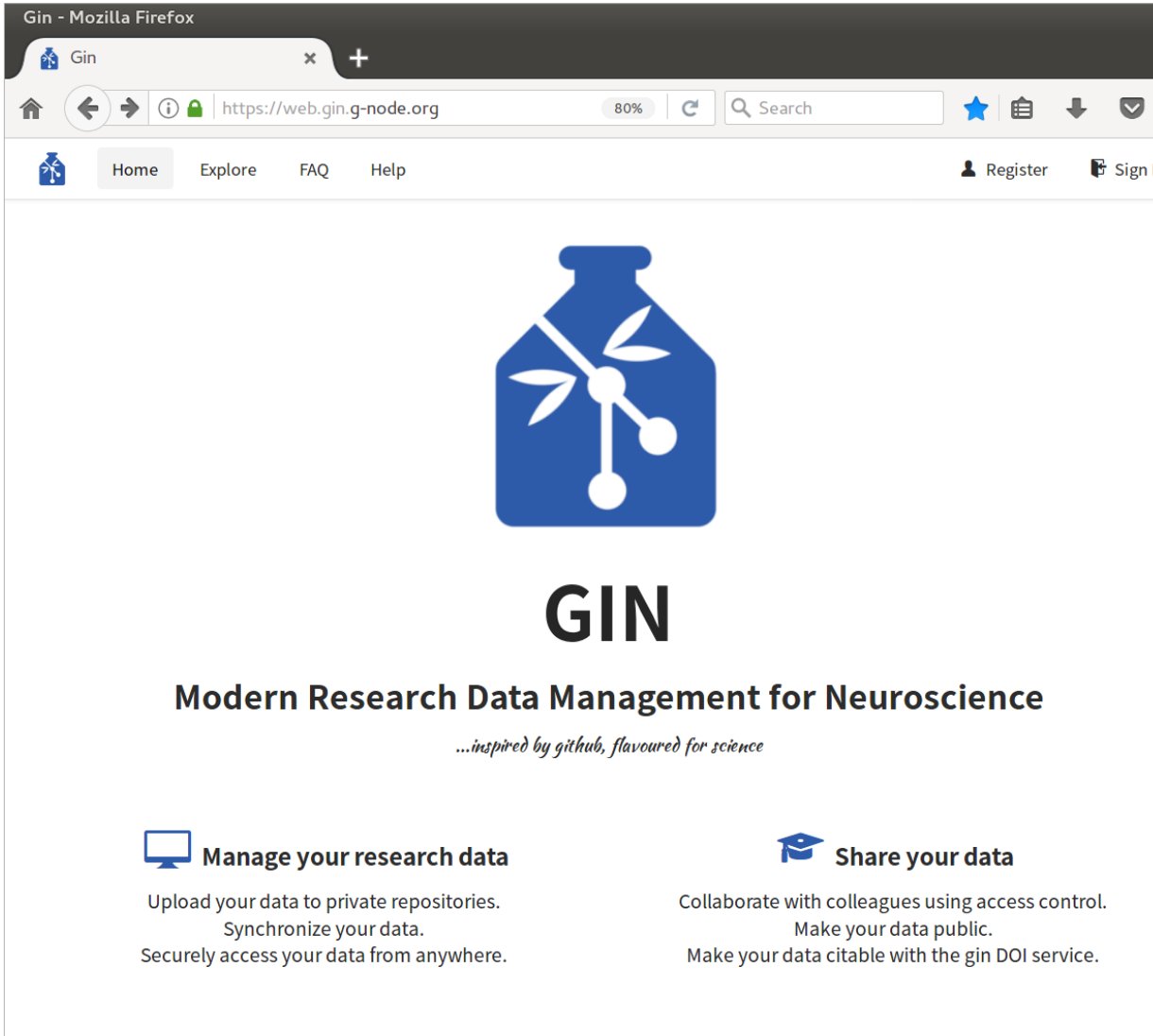
- Pure Python implementation
- Format consolidation NIX ↔ odML
- Suite of tools for conversion [1]
- Visualization tools easy integration into Jupyter notebooks [2]

[1] https://github.com/G-Node/nix-odml-converter
[2] https://github.com/G-Node/nixworks

GIN: Services for Data Organization and Sharing

https://gin.g-node.org

- Versioned management of data repositories (git)
- Distributed data management via file or web browser, or command line client (analysis script integration)
- Secure access and sharing
- Services for search, indexing, and publication (DOI)
- File type plugins
- In house installation
- Extensive usage documentation



GIN: Services for Data Organization and Sharing

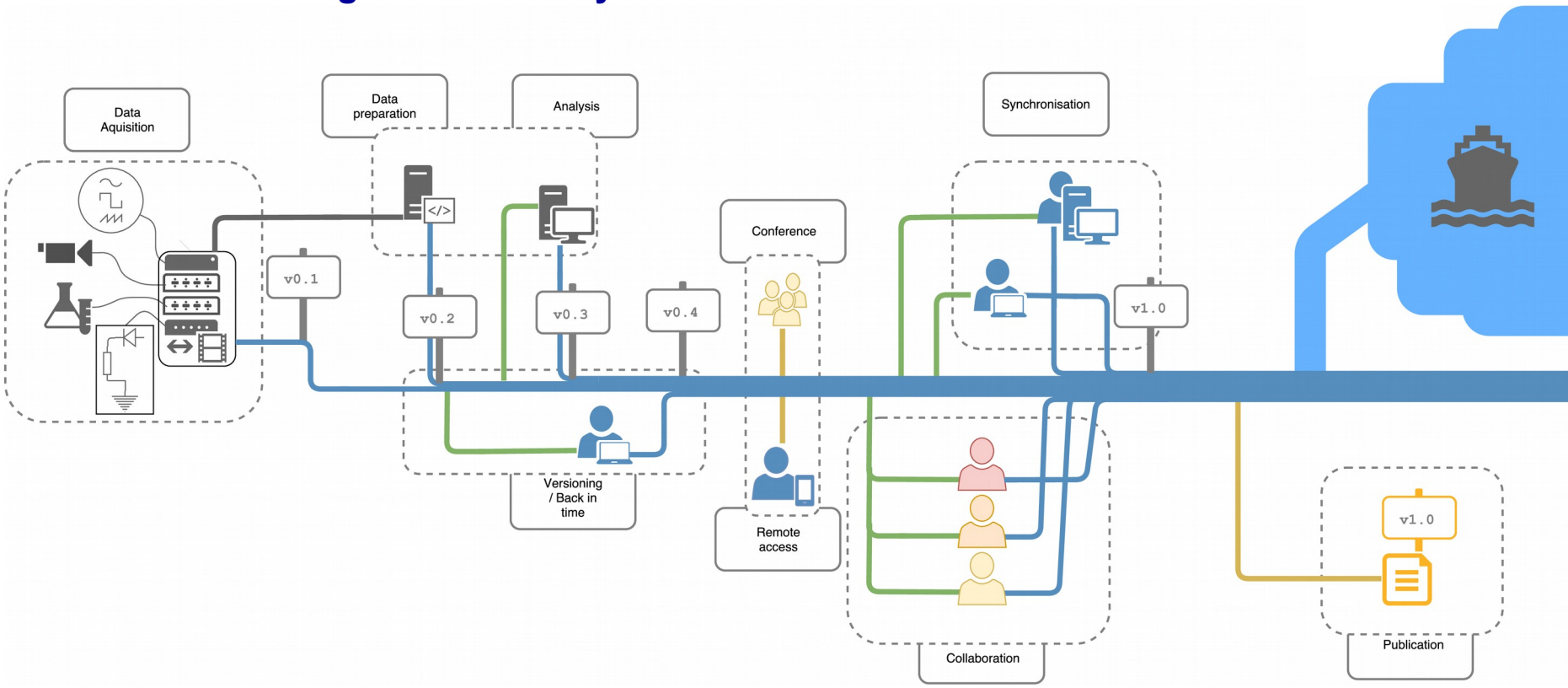
Supporting research data management through the entire data lifecycle

Data Acquisition and Processing

Data Analysis

Data Access and Collaboration

Data Publication

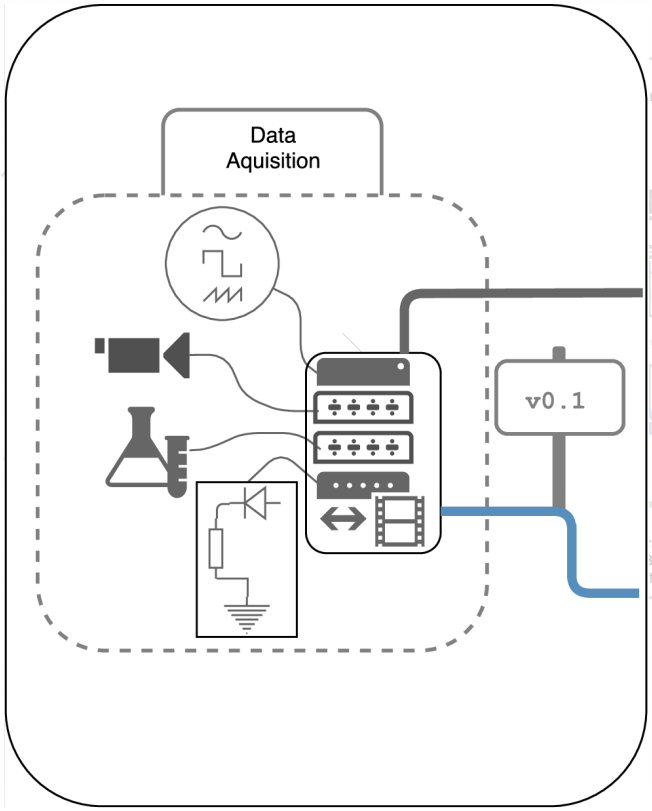


https://gin.g-node.org

GIN: Services for Data Organization and Sharing

Supporting research data management through the entire data lifecycle

Data Acquisition and Processing

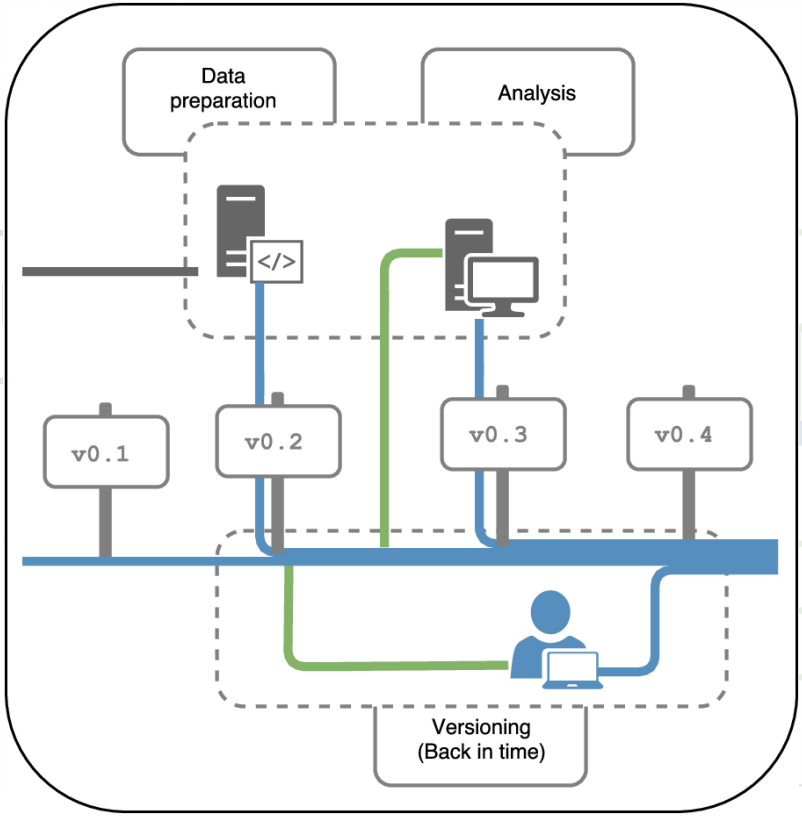


- Versioning of datasets from the beginning
- Keeping track of how dataset evolves
- Immediate access for other lab members
- Manage data with minimal data transfer

GIN: Services for Data Organization and Sharing

Supporting research data management through the entire data lifecycle

Data Analysis



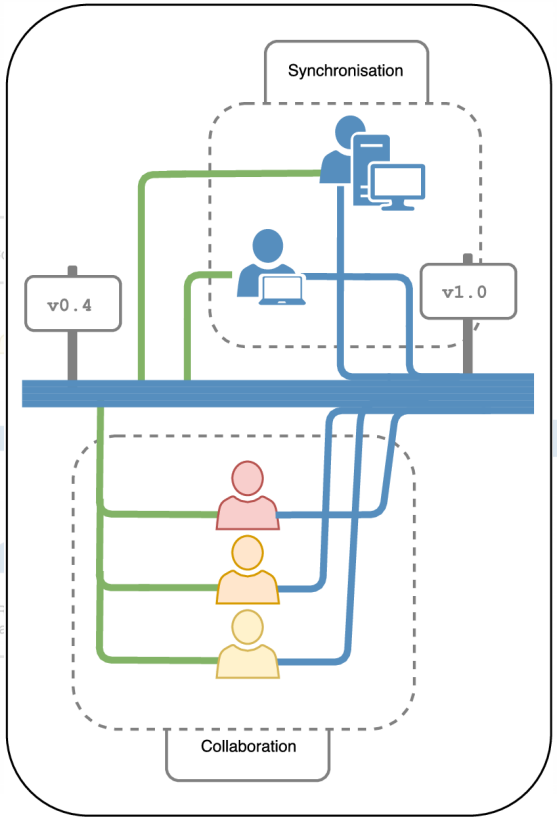
- Versioned analysis workflows
- Working together on data analysis
- Keeping track of who did what
- Reverting changes, Going back in time when necessary

GIN: Services for Data Organization and Sharing

Supporting research data management through the entire data lifecycle

- Accessing your data from outside the lab
- Keeping datasets in sync at diferent places with minimal data duplication
- Working on datasets remotely while transfering data only when needed
- Access for remote collaborators

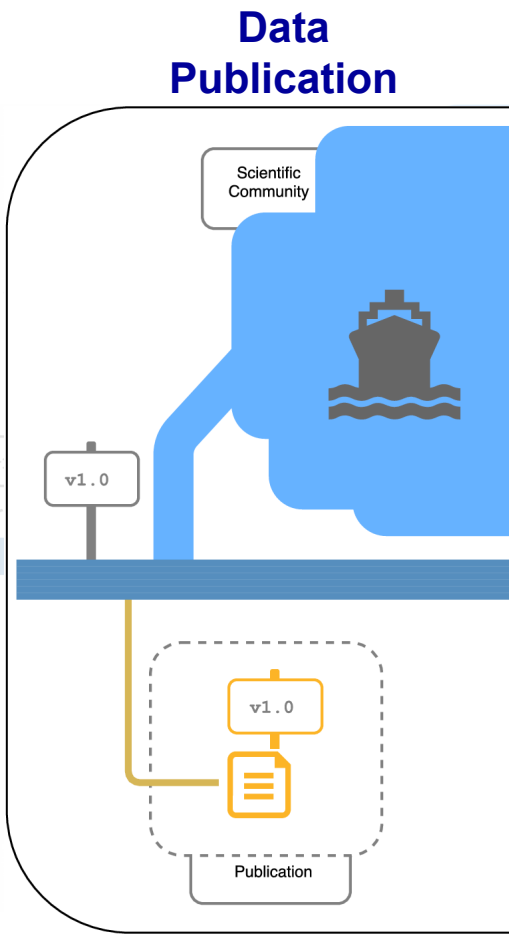
Data Access and Collaboration



GIN: Services for Data Organization and Sharing

Supporting research data management through the entire data lifecycle

- Indexing and data search (gin.g-node.org/explore)
- Making data public with a mouse click
- Dataset publication (supplemental datasets, data publications)
- Citable datasets through Digital Object Identifiers (DOI)



GIN: Services for Data Organization and Sharing

Getting started and Outlook

- Usage documentation at

  https://web.gin.g-node.org/G-Node/Info/wiki/

  https://web.gin.g-node.org/G-Node/Info/wiki/FaqTroubleshooting

- GIN-UI; graphical gin-cli wrapper on Windows
- Web GIN: odML integration
- Web GIN: NIX integration (upcoming)
- GIN microservice: Format validation service (prototype)
  - BIDS, odML, NIX, [your format here]
- GIN microservice: Continuous Integration service (upcoming)

Acknowledgments

G-Node Team

Achilleas Koutsou, Michael Sonntag, Jan Grewe,
Christian Kellner, Christian Garbers, Andrey Sobolev, Adrian Stoewer,
Andreas Herz, Willi Schiegel, Tiziano Zito

Collaborators, Contributors and Supporters

Hiroyuki Ai, Francesc Alted, Rembrandt Bakker, Jan Benda, Anubhav Chaturvedi,
Andrew Davison, **Michael Denker**, Markus Diesmann, Johannes Drever, Gaute Einevoll,
Felix Franke, Hagen Fritsch, Samuel Garcia, Daniel Gonzalez,
Sonja Grün, Michael Hanke, Hidetoshi Ikeno, Petr Jezek, Arvind Kumar,
Ajay Kumaraswamy, Yann Le Franc, Aljoscha Leonhardt, Philipp Meier, Balint Morvai,
Roman Moucek, Dipanjan Mukherjee, Matthias Munk, Martin Nawrot, Cristina Precup,
Robert Pröpper, Raphael Ritz, Jürgen Rybak, Michael Schmuker, Yaroslav Shalivskyy,
Julia Sprenger, Mehmet Süzen, Zbyszek Szmek, Christian Tatarau,
Alvaro Tejero Cantero, Kay Thurley, **Lyuba Zehl**

funded by
Federal Ministry
of Education
and Research

Get involved

Questions

Emails

dev@g-node.org

Chat

gnode.slack.com

Issues and Feature requests

github.com/G-Node

Data sharing and publication

gin.g-node.org

Slides

https://gin.g-node.org/G-Node/PublicResources