Open lightweight tools for safe and efficient data management, processing and validation

Michael Sonntag[1], Achilleas Koutsou[1], Jiří Vaněk[1], Christian Garbers[1], Christian Kellner[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

Maintaining reproducible data workflows while keeping data in sync, backed up, and easily accessible from within and outside the lab is a key challenge in research. To minimize time and effort invested in these tasks scientists have to spend on these tasks, we provide a suite of tools designed for comprehensive, reproducible and versioned management of scientific data.

Organize and Store Data and Metadata

odML: Manage all information about an experiment

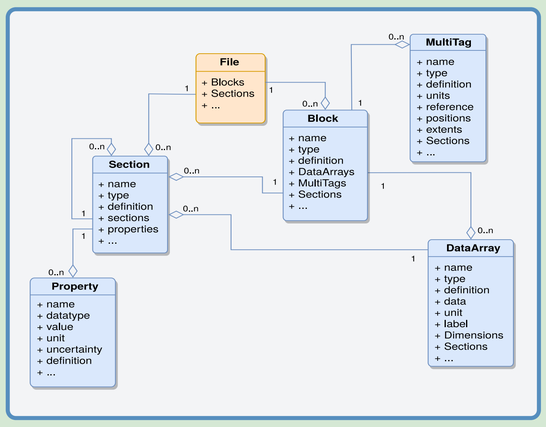
The odML format
- Open metadata format
- Flexible hierarchical key-value storage
- Template system for reusable metadata structures

meta.g-node.org
- export odML to RDF
- access diverse metadata datasets
- all datasets are publicly available
- searchable by SPARQL via API and web

Re-usable metadata concepts:

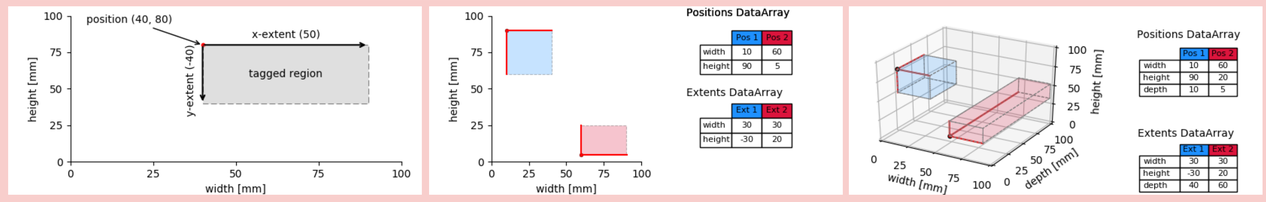
https://templates.g-node.org
Re-usable building blocks to construct metadata files.

https://terminologies.g-node.org
Importable definitions to link to metadata entities.



NIX: Manage data and metadata in one versatile format

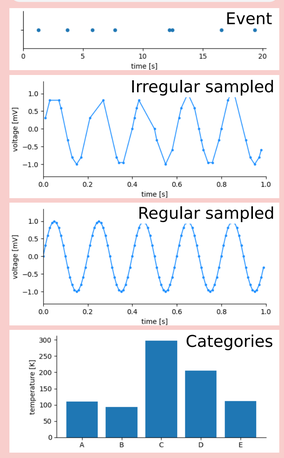
ROIs via MultiTags



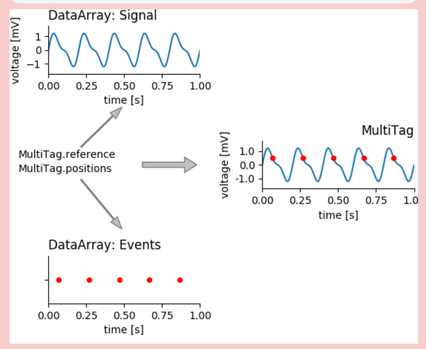
The NIX format
- Open data format
- Raw data, analysis results, and metadata in the same file
- Descriptive associations between data, analysis results, and metadata

Integrates with: neo, elephant

Dimensions

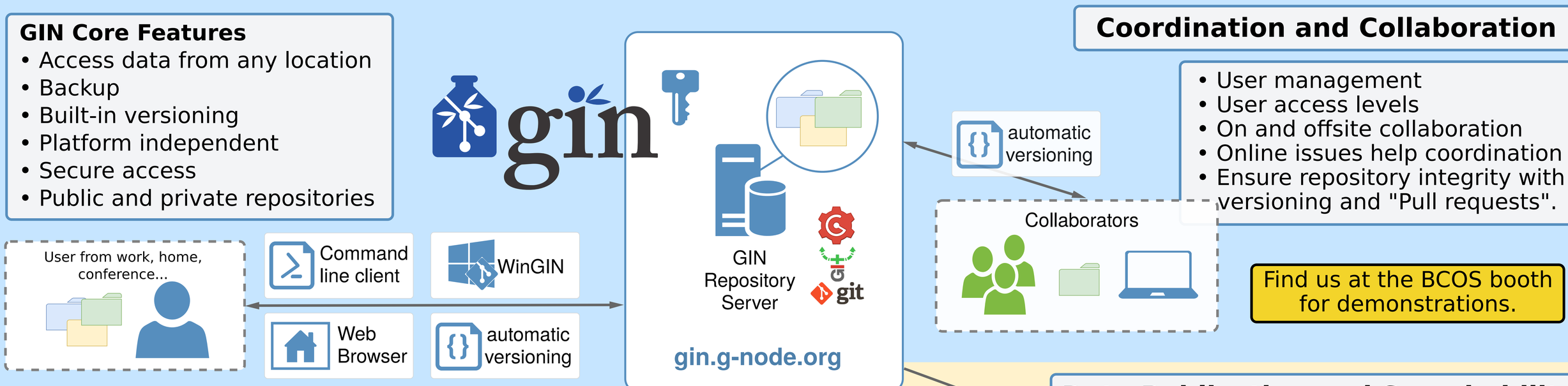


Putting it all together



Coordination and Collaboration

GIN Core Features
- Access data from any location
- Backup
- Built-in versioning
- Platform independent
- Secure access
- Public and private repositories



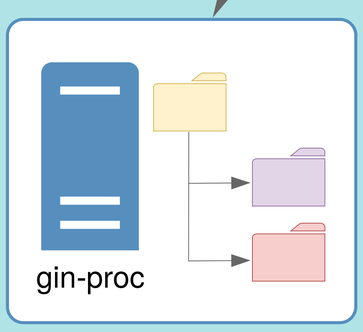
- User management
- User access levels
- On and offsite collaboration
- Online issues help coordination
- Ensure repository integrity with versioning and "Pull requests".

Find us at the BCOS booth for demonstrations.

Automation and Validation Tools

Data validation service
valid.gin.g-node.org

Automated Data Validation
- Automatically runs validation on selected repositories
- Supported validation formats:
  - BIDS - odML - NIX
- Easily extensible to more formats
- Format validation contributions are welcome



Data processing service
proc.gin.g-node.org

Automated Data Processing
- Automatically runs pre-defined processing pipelines
- Triggered on repository changes
- Automatically returns specified results
- Based on SnakeMake and DroneCI

Data Publication and Searchability

Data search service
gin.g-node.org/explore/data

Findable Data via GIN
- GIN provides automatic indexing of all text based files
- Online search for repository content
- Interactive rendering of
  - Markdown • YAML
  - JSON • XML

DOI service
doid.gin.g-node.org

SCIENTIFIC DATA — OFFICIAL RECOMMENDED REPOSITORY

PLOS

Persistent Identifiers
- Any public GIN repository can be registered
- Make your code and data citable
- DOIs for:
  - Data related to publications
  - Research software
  - Whole data sets

Resources and References

Grewe et al (2011), doi:10.3389/fninf.2011.00016
https://github.com/G-Node/python-odml
https://github.com/G-Node/odml-ui
https://github.com/INM-6/python-odmltables
https://github.com/G-Node/nix
https://github.com/G-Node/nixpy
https://github.com/G-Node/nix-mx
https://gin.g-node.org
https://github.com/G-Node/gin-cli
https://github.com/G-Node/wingin
https://github.com/G-Node/gogs
http://neuralensemble.org/neo
http://neuralensemble.org/elephant
http://bendalab.github.io/NixView

Contact:
dev@g-node.org

Supported by BMBF grants 01GQ1302, 01GQ1509

Federal Ministry of Education and Research