Achieving reproducible data workflows: Lightweight tools for safe and efficient data management

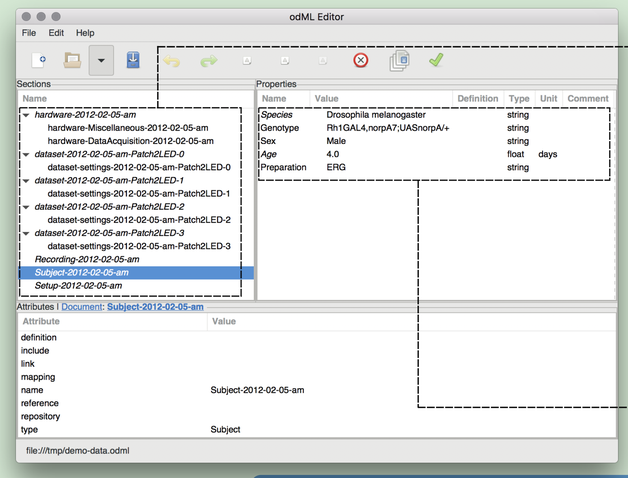
Achilleas Koutsou[1], Michael Sonntag[1], Christian Garbers[1], Christian Johannes Kellner[1], Jan Grewe[2], Thomas Wachtler[1]

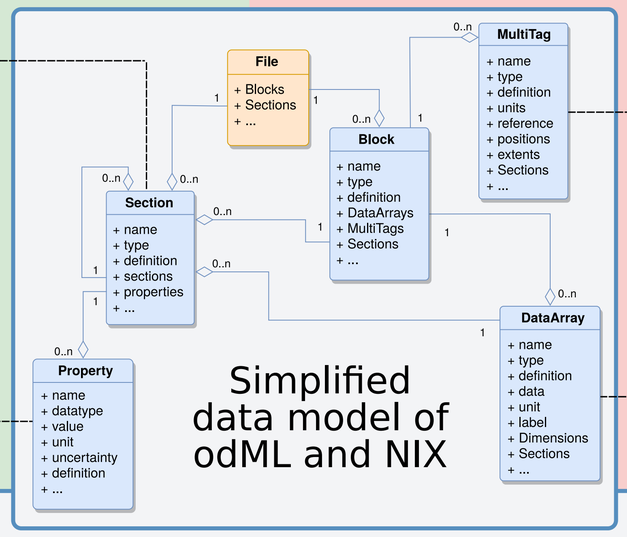
[1]German Neuroinformatics Node, Department Biologie II, Ludwig-Maximilians-Universität München, Germany;
[2]Institut für Neurobiologie, Universität Tübingen, Germany

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 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



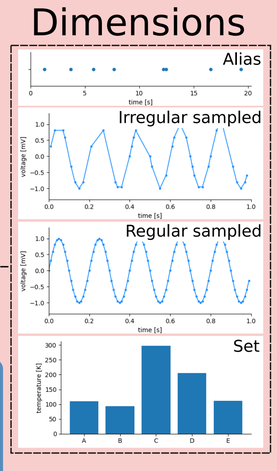


Simplified data model of odML and NIX

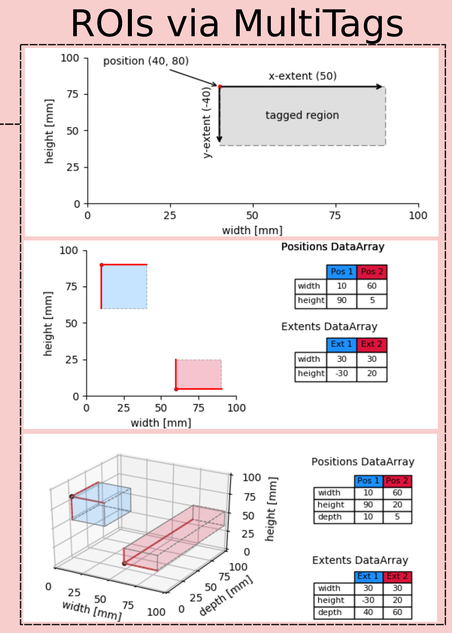
Graphical user interfaces

- odml-ui: browsing and template editing
- odmltables: filtering and tabular export

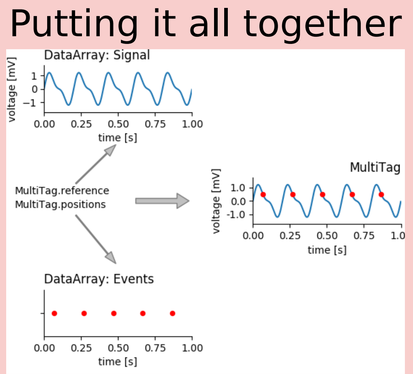
NIX: Manage data and metadata in one versatile format



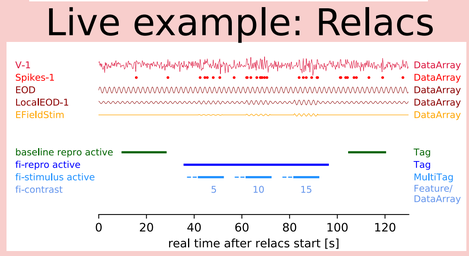
ROIs via MultiTags



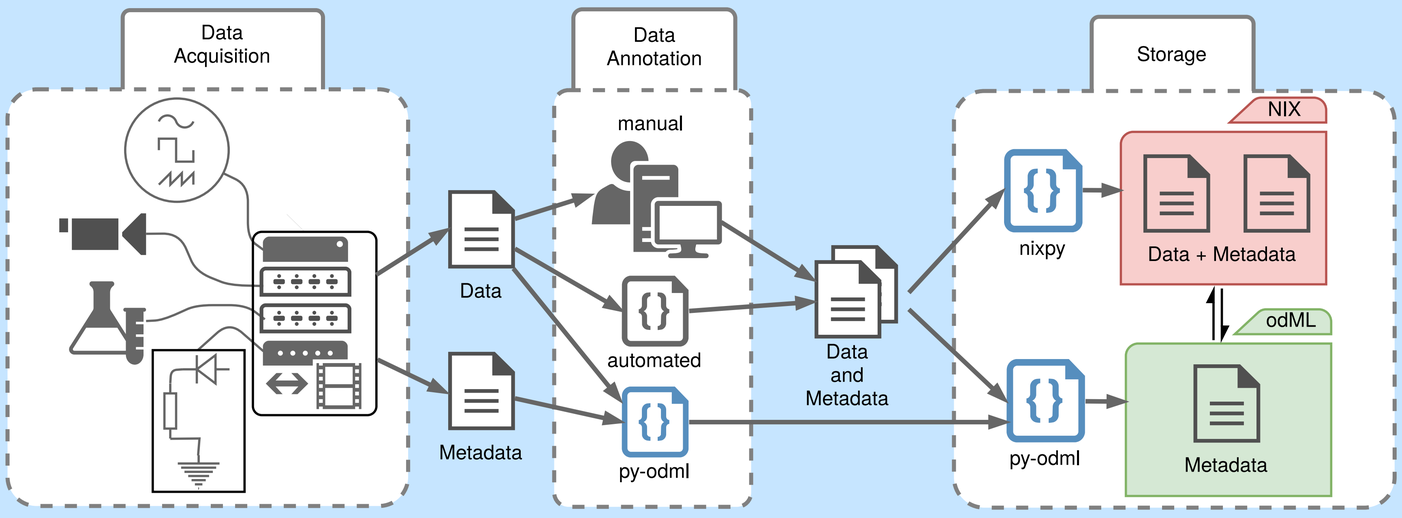
Putting it all together



Live example: Relacs



Data / Metadata acquisition workflow using odML and NIX



Changes to files can be tracked in GIN (see below)

The NIX format

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

The odML format

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

Libraries

Free open source libraries for:

elephant, neo, C++, python, MATLAB, Java

meta.g-node.org

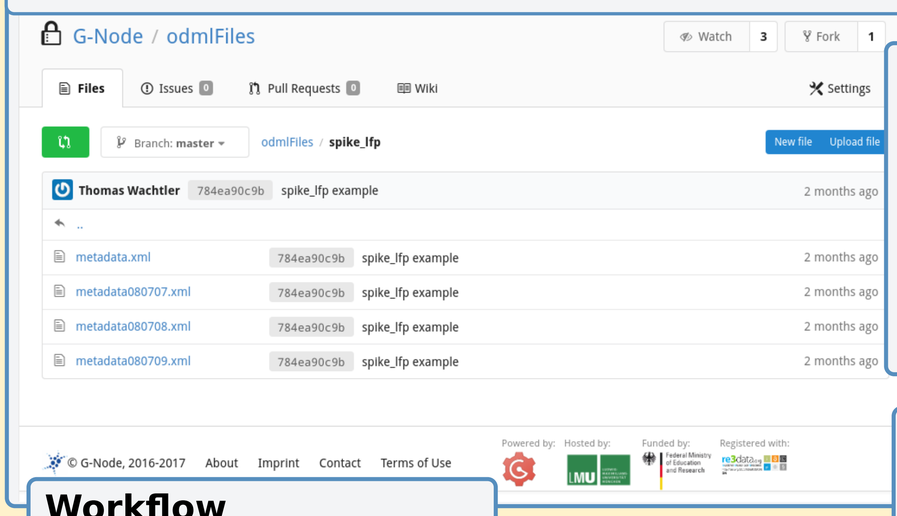
- export odML to RDF
- access diverse metadata datasets
- all datasets are publicly available
- searchable by SPARQL via API and web

Find us at the BCOS booth for demonstrations and a data management Questionaire (with prizes)

Store data securely; publish and collaborate with ease

GIN Web

https://gin.g-node.org



Main features

- Access data from any location
- Backup
- Built-in versioning
- Platform independent
- Secure access
- Public and private repositories
- Citable data by DOIs

Collaboration

- User management
- User access levels
- On and offsite collaboration

Workflow

- Manual access via web and command line
- Script automation

GIN Client

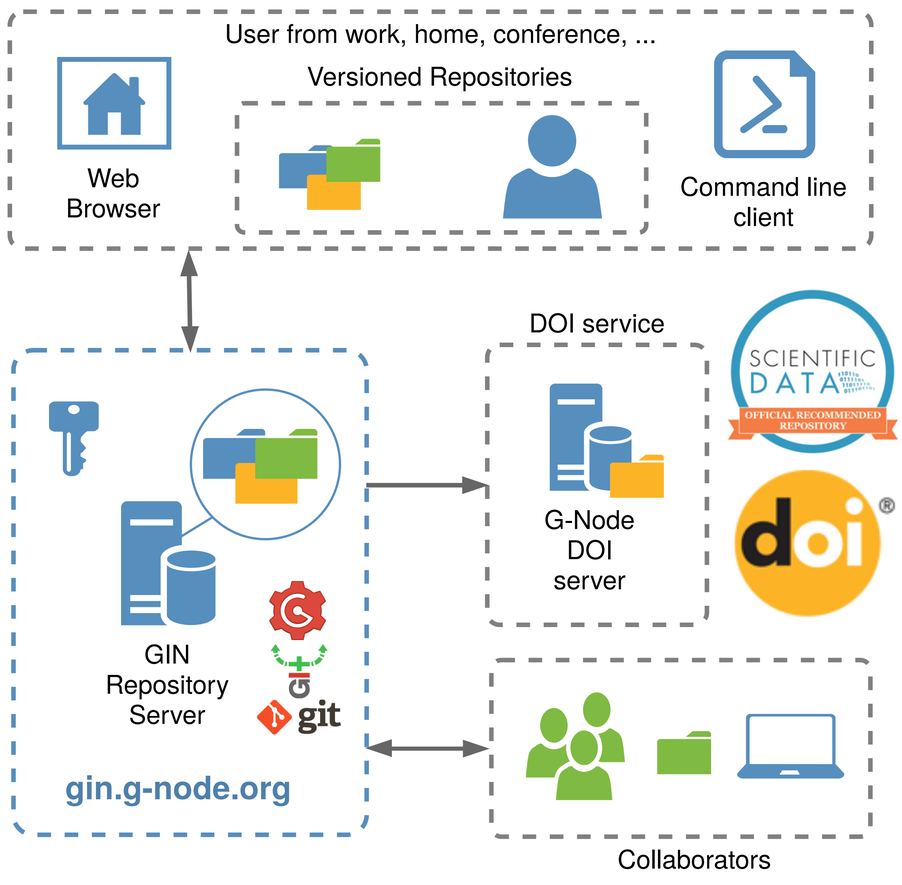
```
$ gin login
Login: achilleas
Password: **********
Hello achilleas. You are now logged in.

$ gin get G-Node/odmlFiles
Fetching repository 'G-Node/odmlFiles'... done.

$ gin ls
Synced:

        metadata.xml
        metadata080707.xml
        metadata080708.xml
        metadata080709.xml

$
```

gin

Upcoming features

- Format validation
  - BIDS, odML, NIX, custom formats
- CI for scientific data: automated tests for scripts and data integrity
- automated export of odML to RDF

GIN Services



Resources

Poster presented at the NWG Conference 2019 (Goettingen)

Contact:
dev@g-node.org

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/gogs
https://github.com/relacs/relacs
http://neuralensemble.org/neo
http://neuralensemble.org/elephant
http://bendalab.github.io/NixView

Supported by BMBF grants 01GQ1302, 01GQ1509

Federal Ministry of Education and Research