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

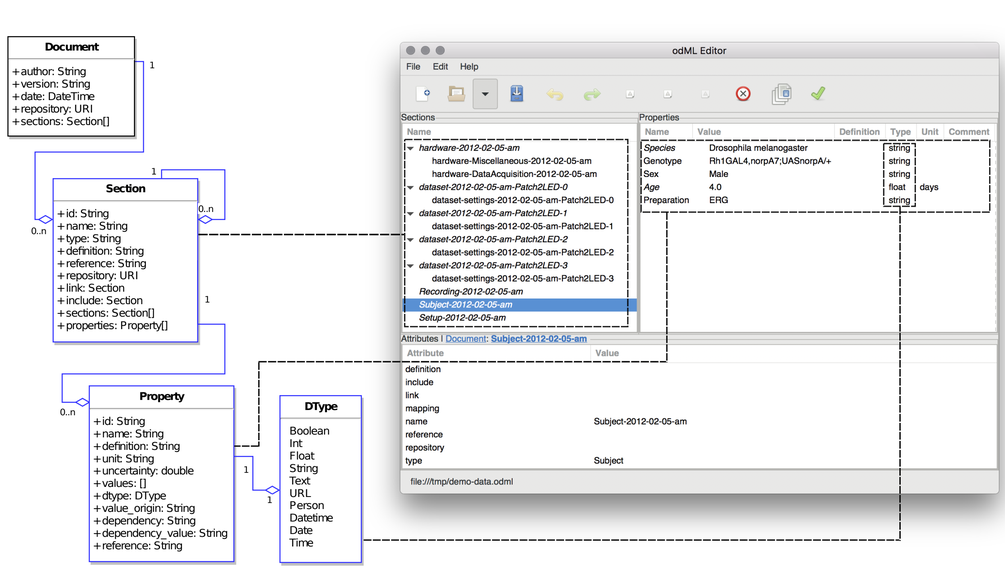
Christian Garbers[1], Michael Sonntag[1], Achilleas Koutsou[1], Christian Johannes Kellner[1], Jan Grewe[2], Thomas Wachtler[1]

[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.

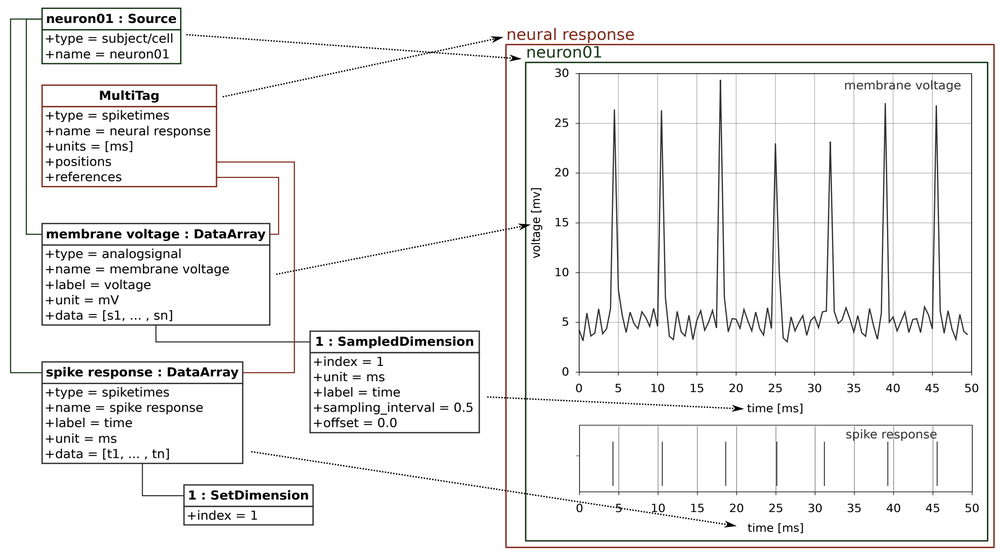
Organize and Store Data and Metadata

odML Metadata format

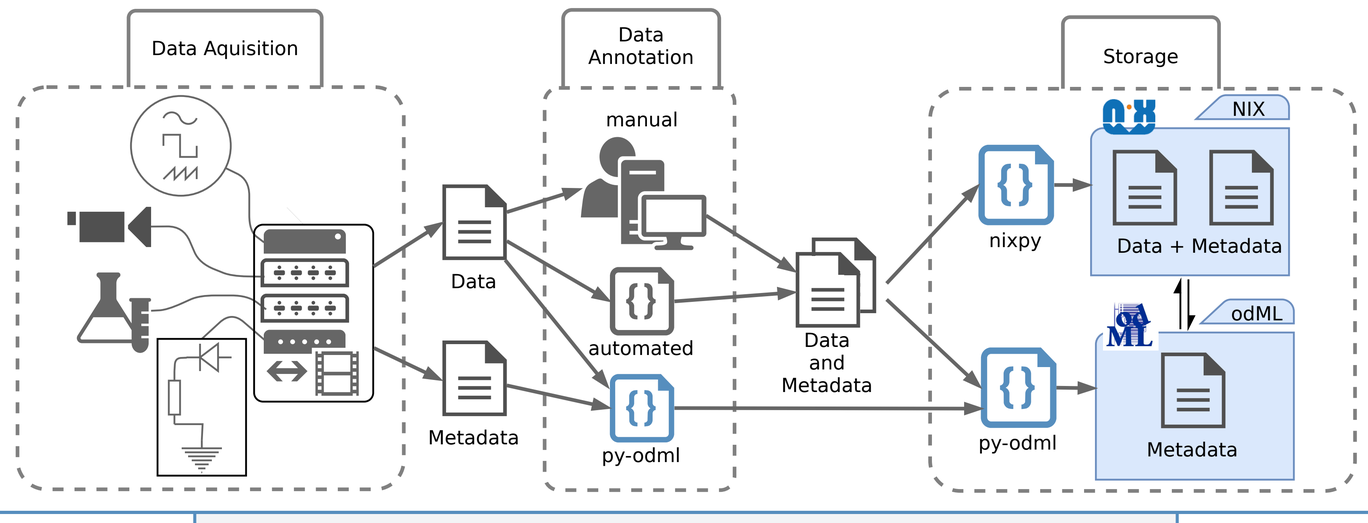


odML
- Metadata format used in NIX
- Read and write metadata using library or editor
- Export to RDF: Query using Semantic Web

The NIX data format to store scientific data



Data / Metadata acquisition workflow using odML and NIX



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

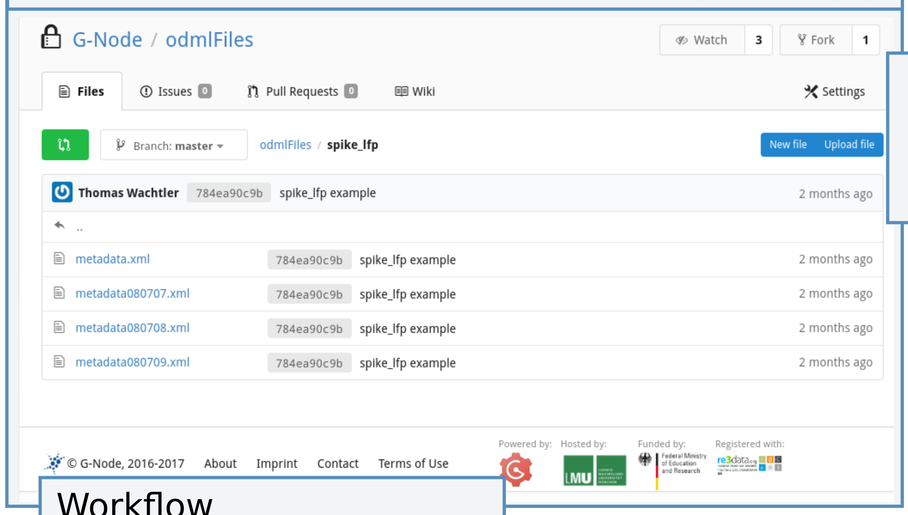
Libraries
Free open source libraries for:
C++ python MATLAB Java neo elephant

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

Keep Track of Your Data, Collaborate, Publish

GIN Web

https://gin.g-node.org



- Platform independent
- Secure access
- Public and private repositories
- Versioning of data

Main features
- Access data from any location
- Backup
- Built-in versioning

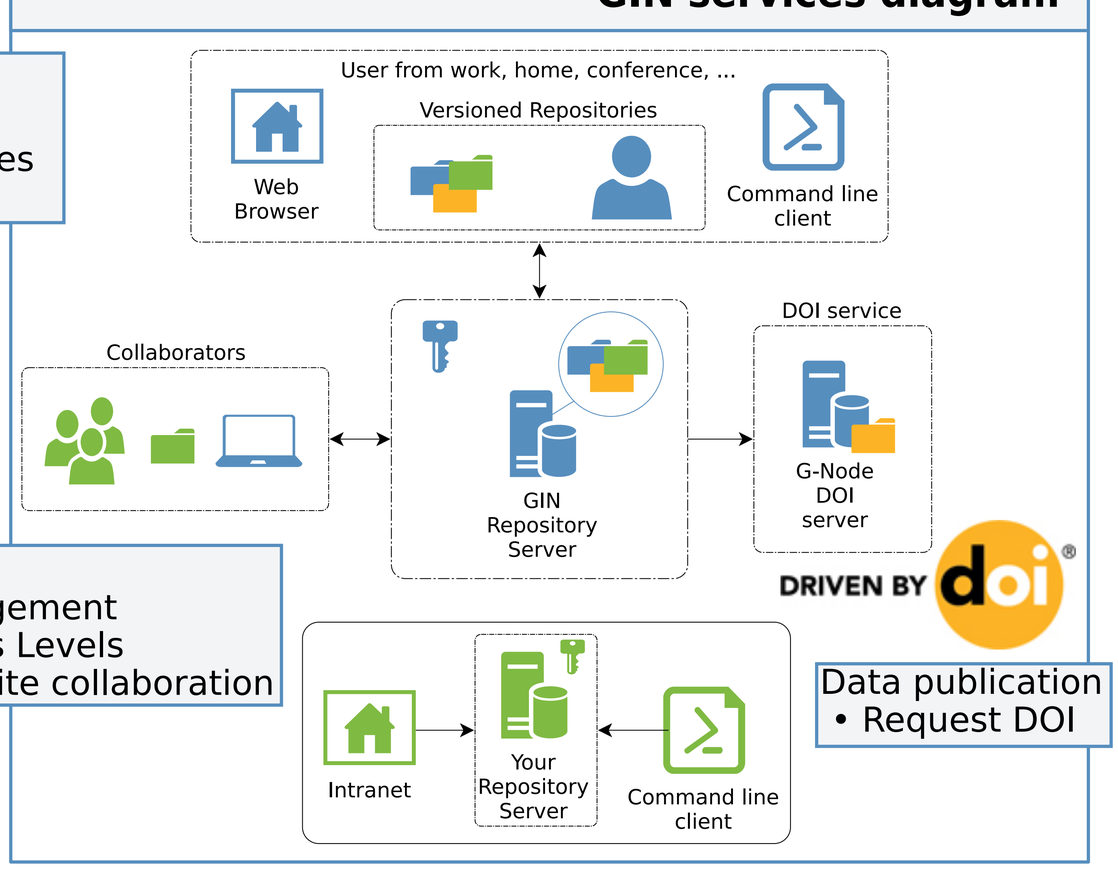
Workflow
- Manual access via web and command line
- Script automation

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

$ |
```

Collaboration
- User management
- User Access Levels
- On and offsite collaboration

GIN services diagram



Data publication
- Request DOI

Resources

Contact:
dev@g-node.org

Poster presented at the Bernstein Conference 2018 (Berlin)

doi: 10.12751/nncn.bc2018.0079

https://gin.g-node.org
https://github.com/G-Node/nix
https://github.com/G-Node/python-odml
https://github.com/G-Node/gin-cli
https://github.com/G-Node/gogs
http://neuralensemble.org/neo
http://neuralensemble.org/elephant

Supported by BMBF

Federal Ministry of Education and Research