

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

LUDWIG-MAXIMILIANS-JNIVERSITÄT

C Garbers<sup>1</sup>, M Sonntag<sup>1</sup>, A. Koutsou<sup>1</sup>, C J Kellner<sup>1</sup>, J Grewe<sup>2</sup>, T Wachtler<sup>1</sup>

<sup>1</sup>Ludwig-Maximilians-Univität München, Germany; <sup>2</sup>Eberhard Karls 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 and versioned management of scientific data including convenient storage of data, analysis and metadata annotation for easy reproducability, data sharing and re-usability.



Open metadata format [1]

Template system for reusable

Flexible hierarchical key-value storage

Save to common structured formats:

**Main features** 

XML, JSON, YAML

metadata structures

#### Collect and manage all information about your experiment

- GUI editor [3]
- Available for macOS and Linux
- Cross-document drag-and-drop for metadata subtrees

py-odml

Storage

Data + Metadata

 $\equiv$ 

Metadata

圔

odML

- Export to RDF retaining your own terms and structure

Data

Metadata

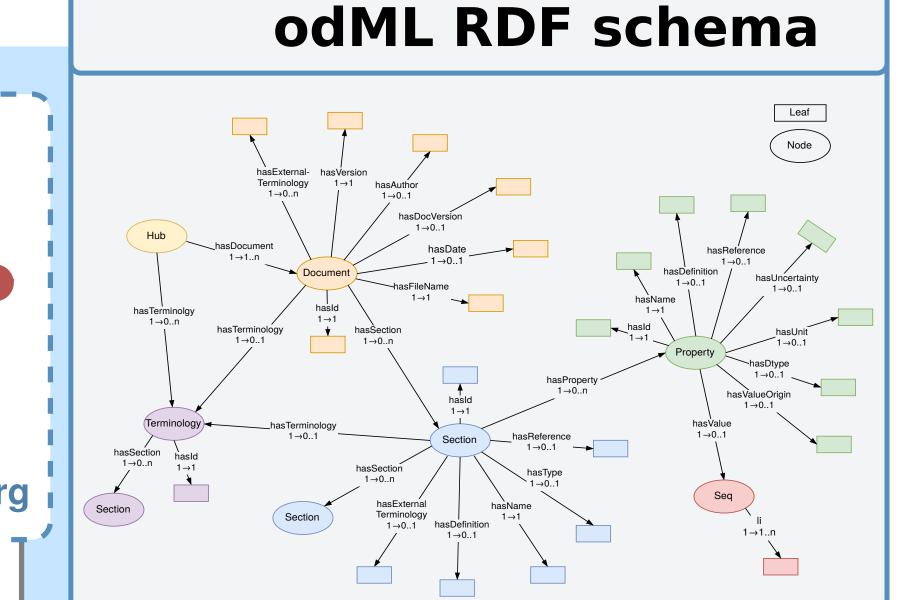
Annotation

manual

automated

py-odml

# The odML Metadata format meta.g-node.org Queries



# Secure data storage, easy collaboration and publication



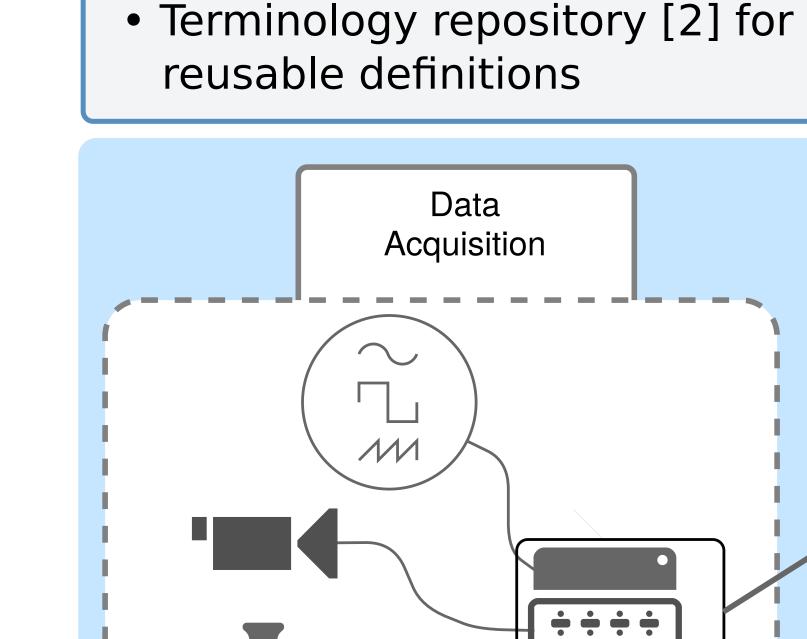
#### **Main features**

- Access data from any location
- Free storage for scientific data [11]
- Built in versioning (built on git [12])
- Platform independent
- Secure access

User from work, home, conference, ...

Versioned Repositories

- Public and private repositories
- Citable data by DOIs



#### Query metadata using semantic web technologies Search cross document via SPARQL queries Make metadata publicly available on meta.g-node.org

#### {} py-odml all datasets are publicly available searchable by SPARQL via API and web

meta.g-node.org:

access diverse metadata datasets

### **GIN Repository Workflow**

- Browse, download, and upload data via web
- Download and upload large files via command line
- Automate workflows using command line client
- All changes are versioned automatically

#### odML files in gin:

**GIN Client** 

\$ gin login Login: achilleas Password: \*\*\*\*\*\*\*\*\*\*\*

Hello achilleas. You are now logged in

Initialising local storage... OK

metadata080709.xml

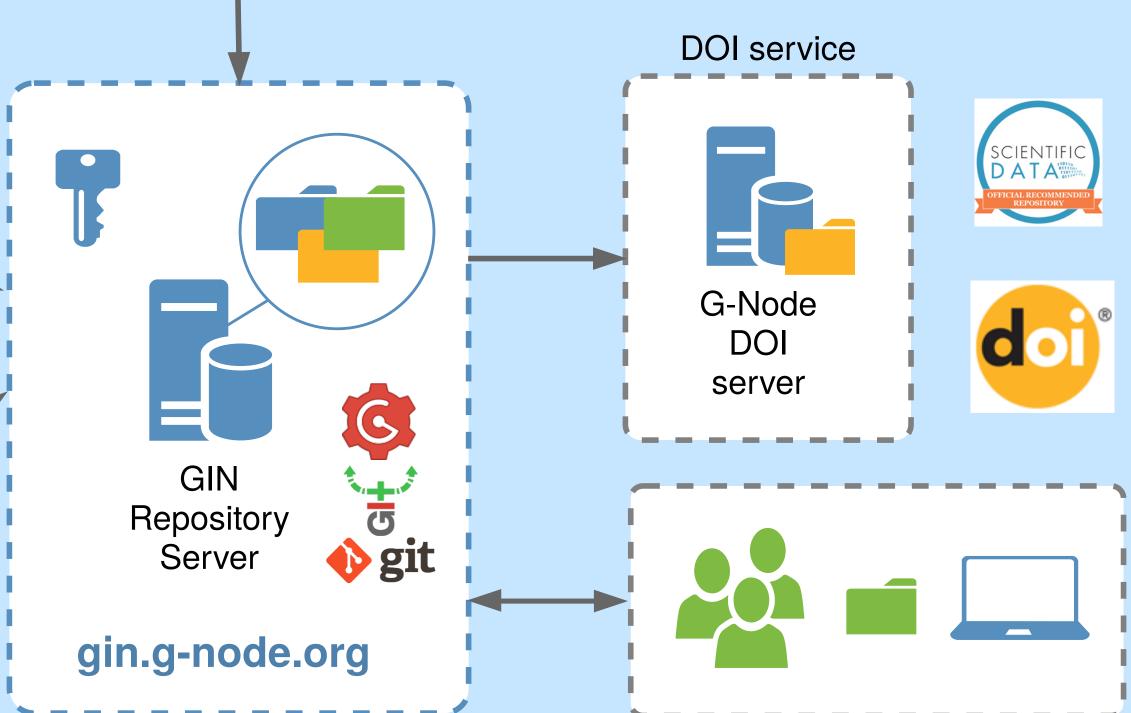
- indexed and searchable
- treeview rendering for convenient exploration

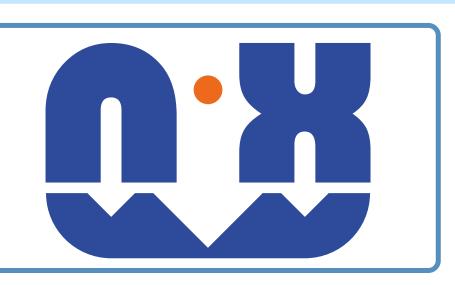
GIN Web

Powered by: Hosted by: Funded by: Registered with:

| Federal Ministry of Education and Research | Federal Ministry of Education | Federal Ministry of Educat

# Command line Browser





### Manage data and metadata together in an open, versatile format

#### Main features

- Open data format
- Store data, analysis results, and metadata conveniently in the same file

- Descriptive associations between data, analysis results, and metadata
- C++ P python MATLAB S Java Ne December 1

Data

Metadata

- Matlab [6], Java [7] NIX IO for Neo [8] Enables interoperability with Neo compatible tools, e.g., the Elephant toolkit [9]

NIXView [10] Cross-platform GUI

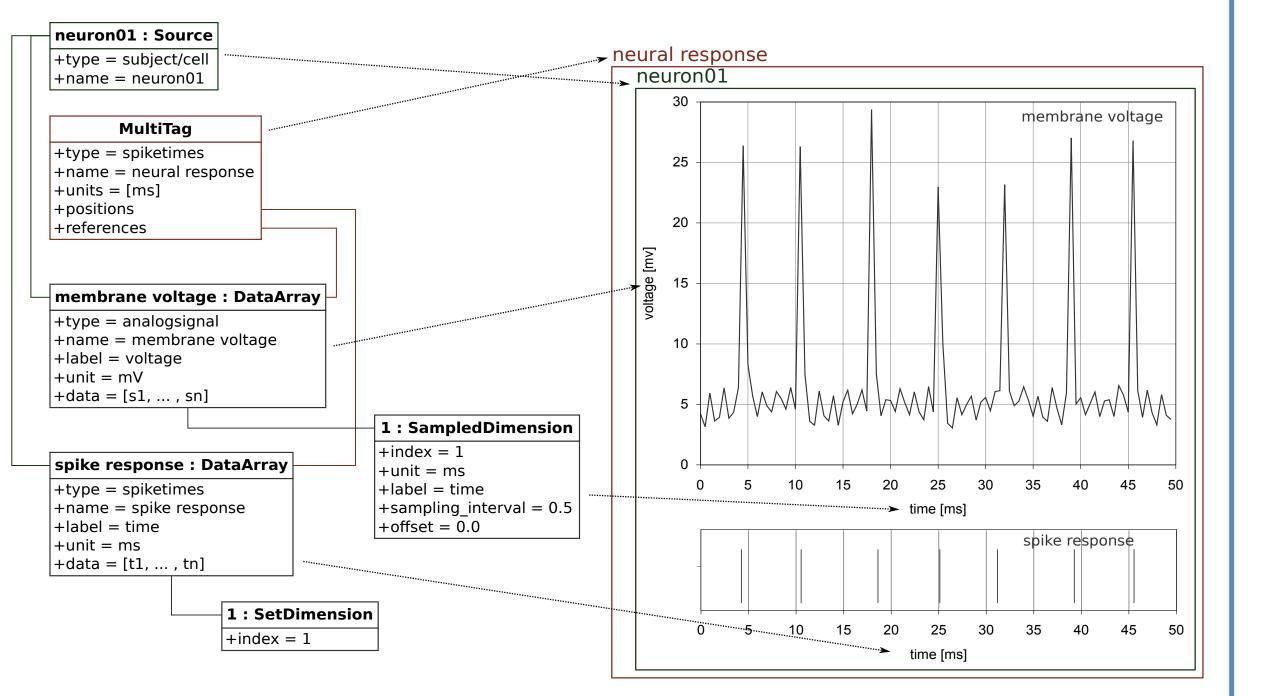
C++ [4], Python [5],

languages:

Free open source libraries

for multiple programming

### The NIX data model



#### **Upcoming features**

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

#### Collaboration

Collaborators

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

#### Resources

Contact: dev@g-node.org

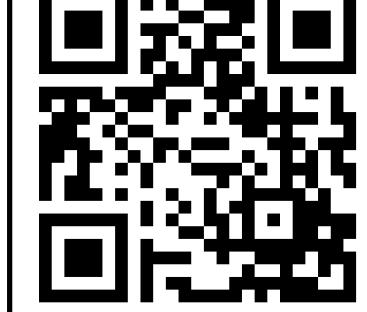
nttps://gin.g-node.org

ু Branch: master ▼ odmlFiles / spike\_lfp

Thomas Wachtler 784ea90c9b spike\_lfp example

metadata.xml

- [1] Grewe et al (2011), doi:10.3389/fninf.2011.00016
- [2] http://www.g-node.org/projects/odml/terminologies
- [3] https://github.com/G-Node/odml-ui
- [4] https://github.com/G-Node/nix
- [5] https://github.com/G-Node/nixpy
- [6] https://github.com/G-Node/nix-mx [7] https://github.com/G-Node/nix-java
- [8] http://neuralensemble.org/neo
- [9] http://neuralensemble.org/elephant
- [10] http://bendalab.github.io/NixView
- [11] https://gin.g-node.org [12] https://git-scm.com



Supported by BMBF grants 01GQ1302, 01GQ1509



