

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>

A brief introduction to the G-Node

<http://www.g-node.org>



Thomas Wachtler

- Funded by BMBF
- Hosted at the Ludwig-Maximilians Universität München
- Bernstein Network of Computational Neuroscience
- Associated Node of the INCF



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>



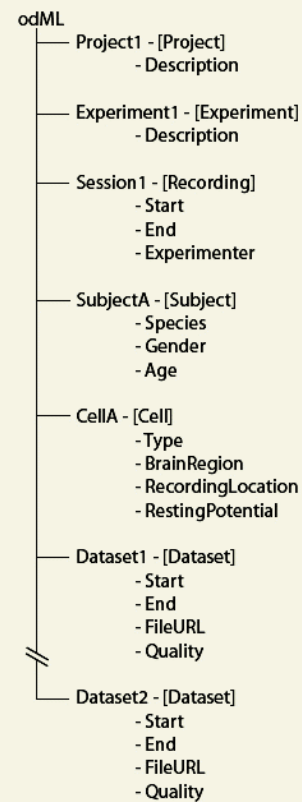
odML (open metadata Markup Language)

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

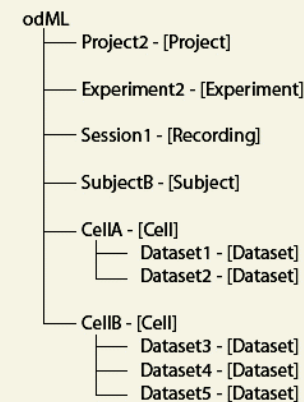


- format: hierarchical structure of key-value pairs: simple, flexible, inherently extensible → can be adapted to the specifics of the lab or experiment
- can carry **any metadata**
- machine write- and readable, **facilitates automated collection** of metadata in the laboratory
- community-driven standardization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

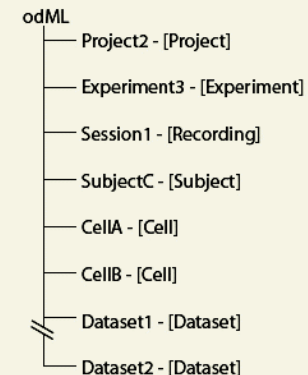
A) Single cell recording several datasets



B) Two cells subsequently recorded, several datasets each



C) Two cell simultaneously recorded, several datasets

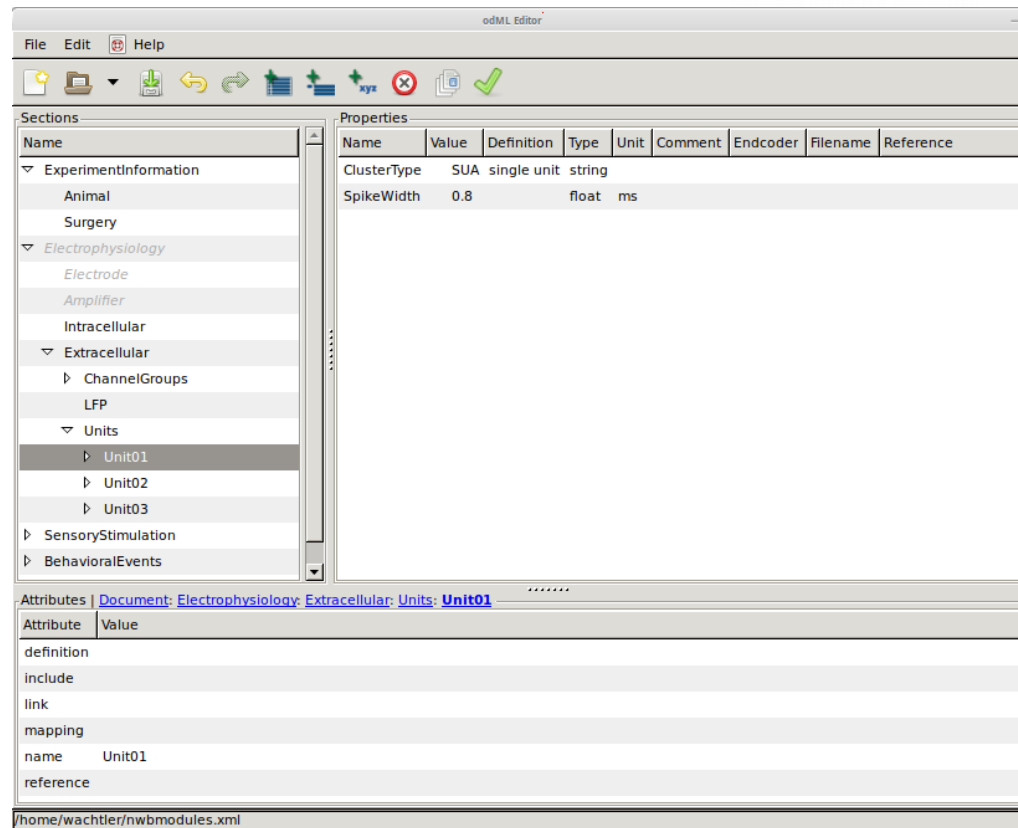


odML: Tools

<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

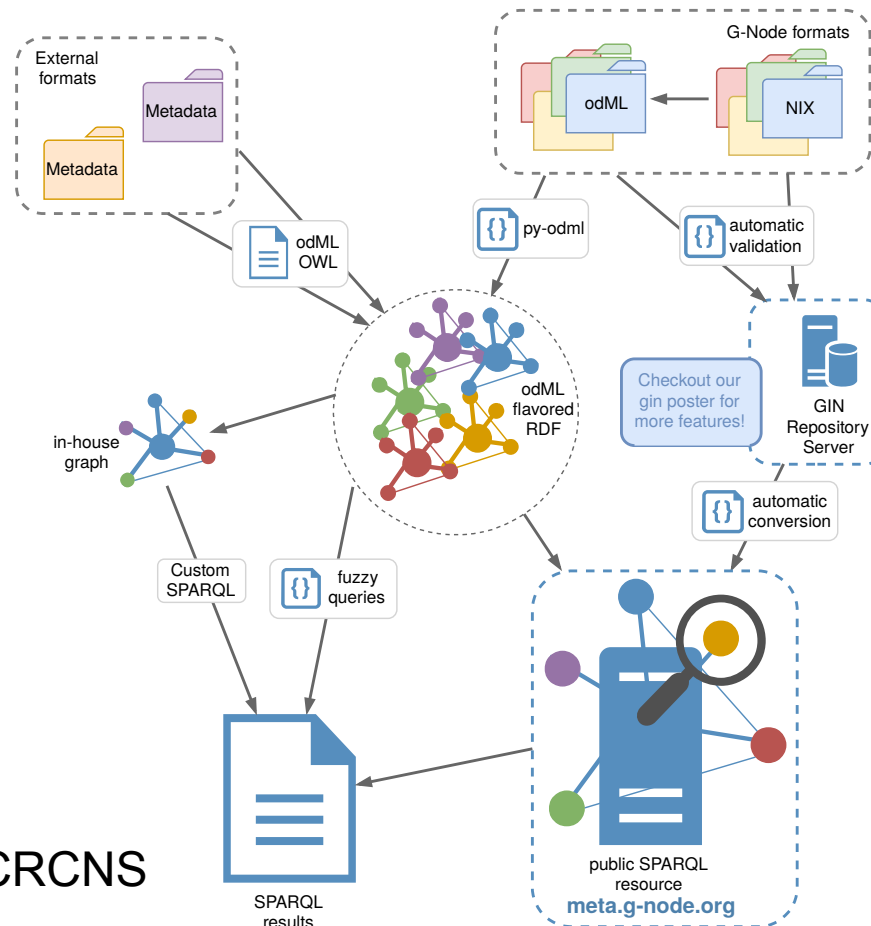
<http://www.g-node.org/odml>



- 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>



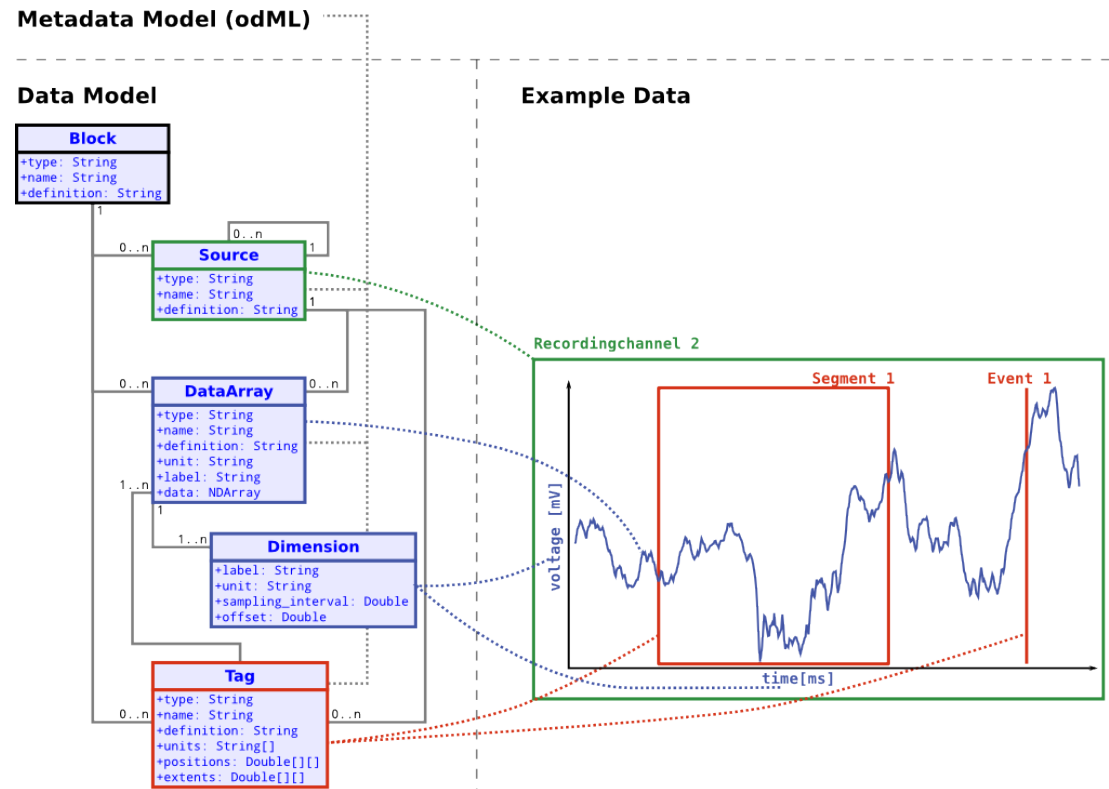
Joint project with the CRCNS

<https://crcns.org/>

NIX – Format for integration of data and metadata

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

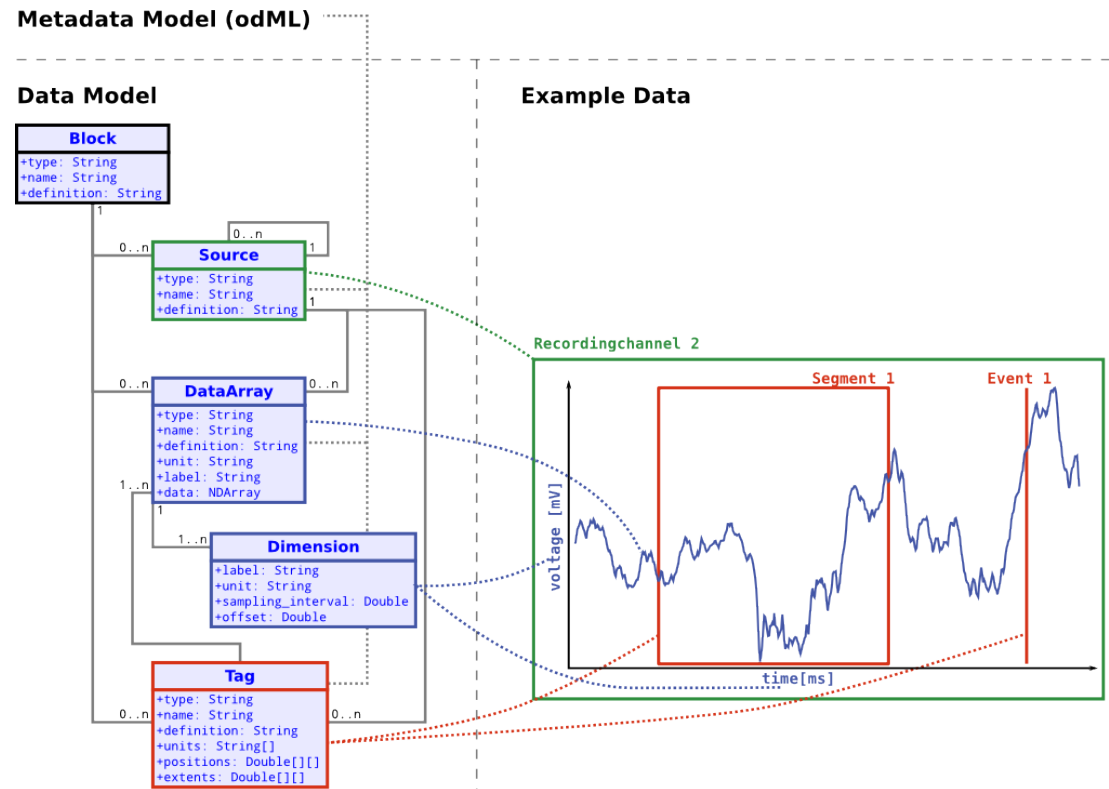
- general data model (derived from **Neo**) to represent recorded data, derived data, relations of data
- flexible data model for metadata (**odML**) for comprehensive annotation of data
- file backend: HDF5 file format
 - structure reflects data model, easy to understand
 - other storage backends possible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



NIX – Format for integration of data and metadata

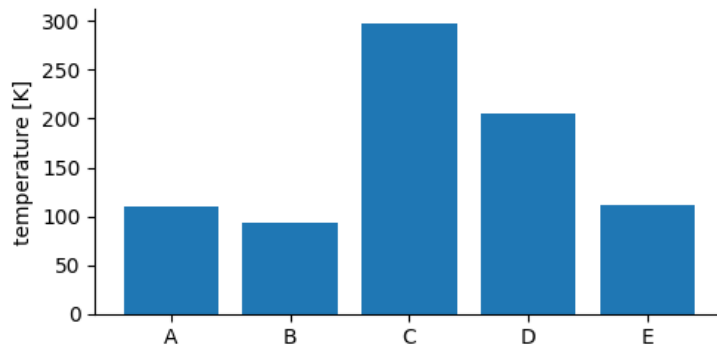
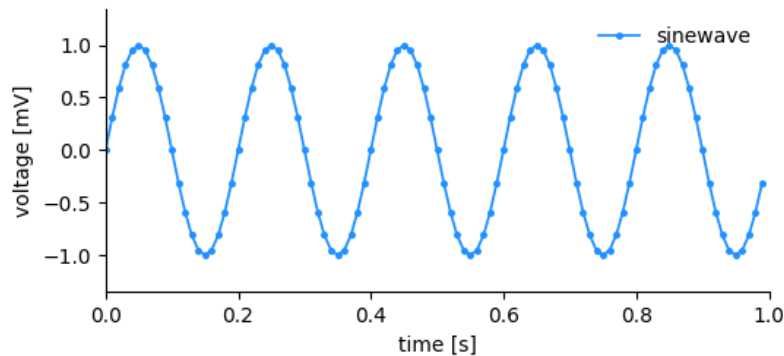
<http://www.g-node.org/nix>

- 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

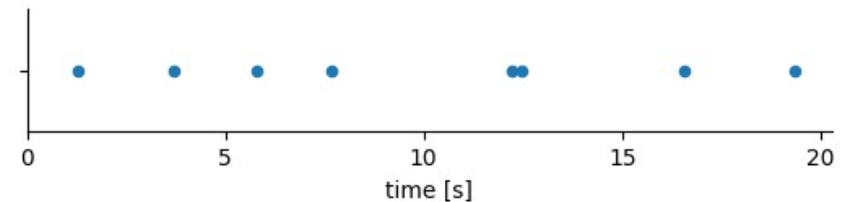
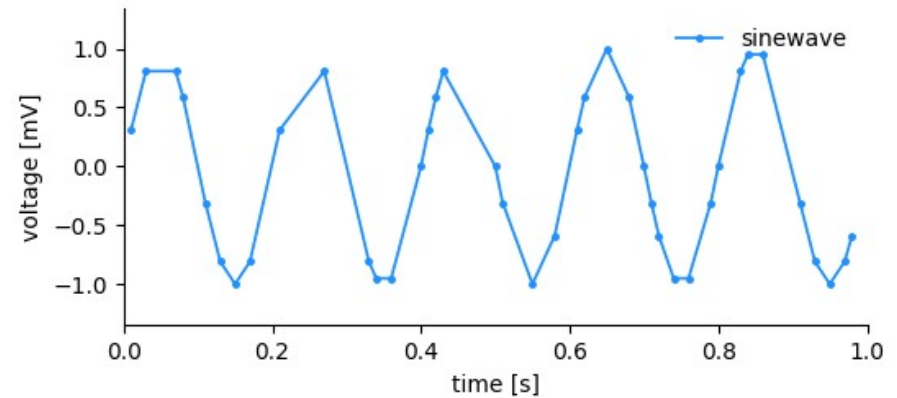
Sampled dimension



Set dimension

Features automatic SI unit conversion

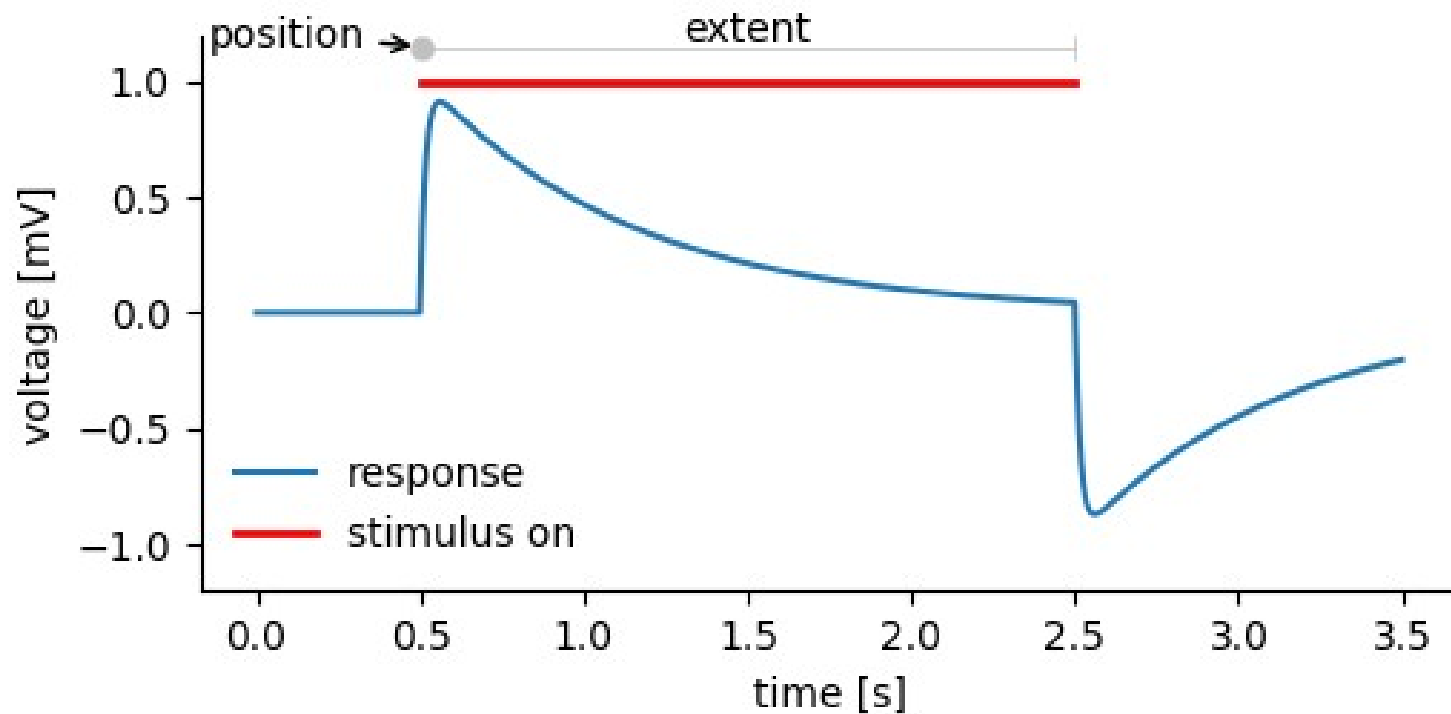
Range dimension



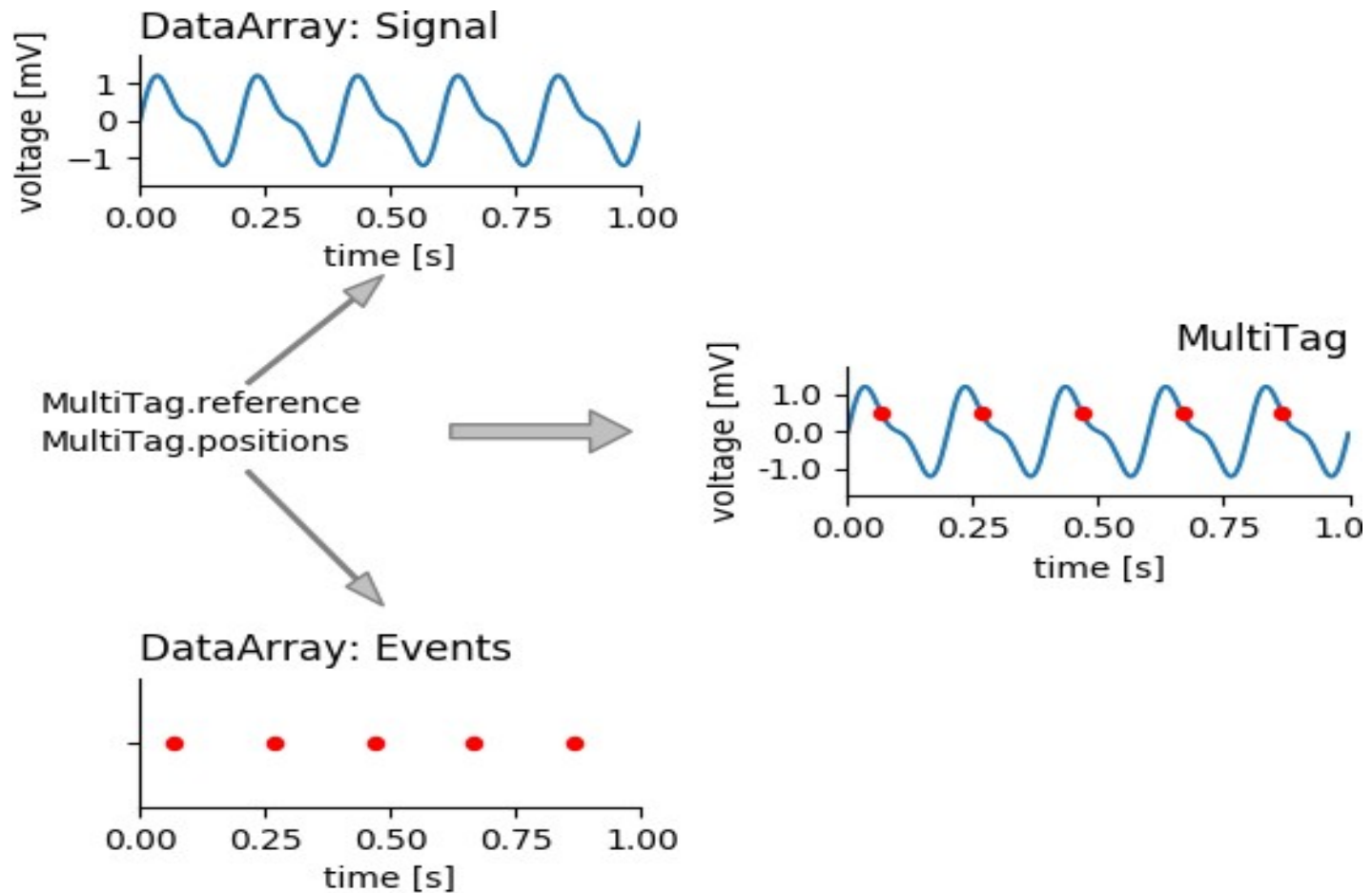
Alias Range dimension

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

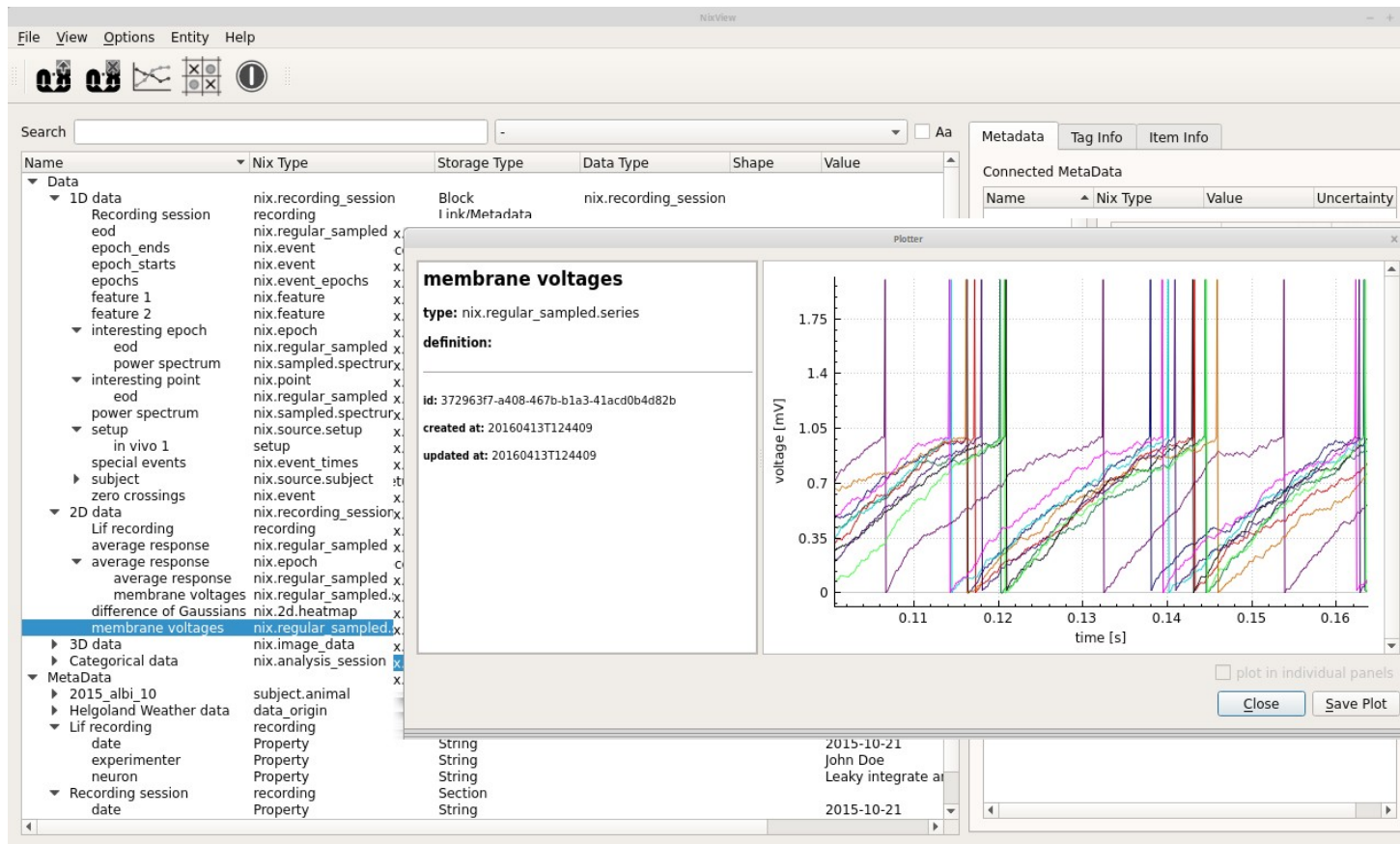
NIX features – Dimensions and tagging



NIX features – Dimensions and tagging



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 - Mozilla Firefox

Gin x +

https://web.gin.g-node.org 80% Search


Home Explore FAQ Help Register Sign




GIN

Modern Research Data Management for Neuroscience

...inspired by github, flavoured for science

**Manage your research data**

Upload your data to private repositories.
Synchronize your data.
Securely access your data from anywhere.

**Share your data**

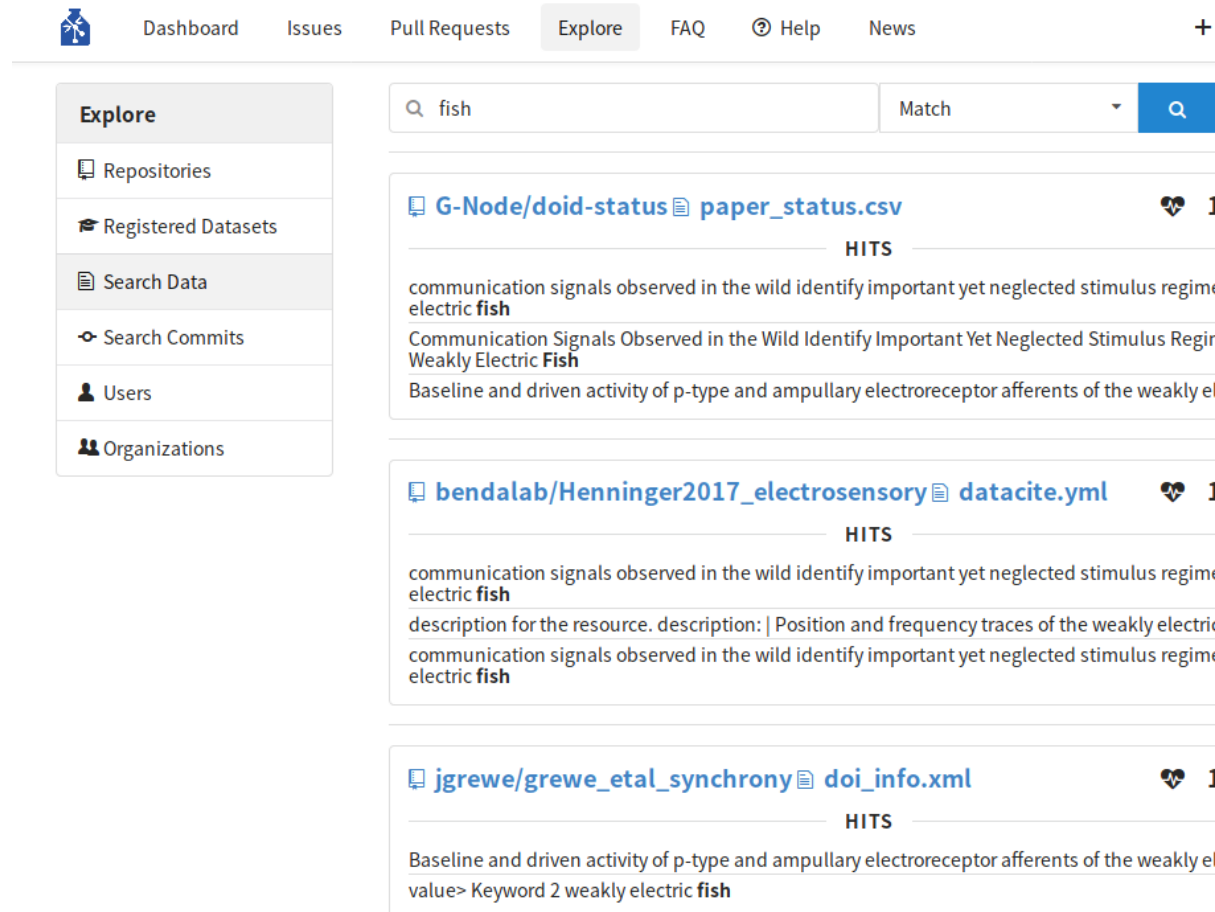
Collaborate with colleagues using access control.
Make your data public.
Make your data citable with the gin DOI service.

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



The screenshot displays the GIN web interface. At the top, there is a navigation bar with links: Dashboard, Issues, Pull Requests, Explore (active), FAQ, Help, and News. Below the navigation bar, a search bar contains the text 'fish'. To the left of the search results is a sidebar with the 'Explore' section expanded, showing links to Repositories, Registered Datasets, Search Data (highlighted), Search Commits, Users, and Organizations. The main content area shows search results for 'fish'. The first result is from 'G-Node/doid-status' with the file 'paper_status.csv'. It has 1 hit and a description: 'communication signals observed in the wild identify important yet neglected stimulus regime electric fish'. The second result is from 'bendalab/Henninger2017_electrosensory' with the file 'datacite.yml'. It also has 1 hit and a description: 'communication signals observed in the wild identify important yet neglected stimulus regime electric fish'. The third result is from 'jgrewe/grewe_etal_synchrony' with the file 'doi_info.xml'. It has 1 hit and a description: 'Baseline and driven activity of p-type and ampullary electroreceptor afferents of the weakly electric fish'.

GIN: Services for Data Organization and Sharing

<https://gin.g-node.org>



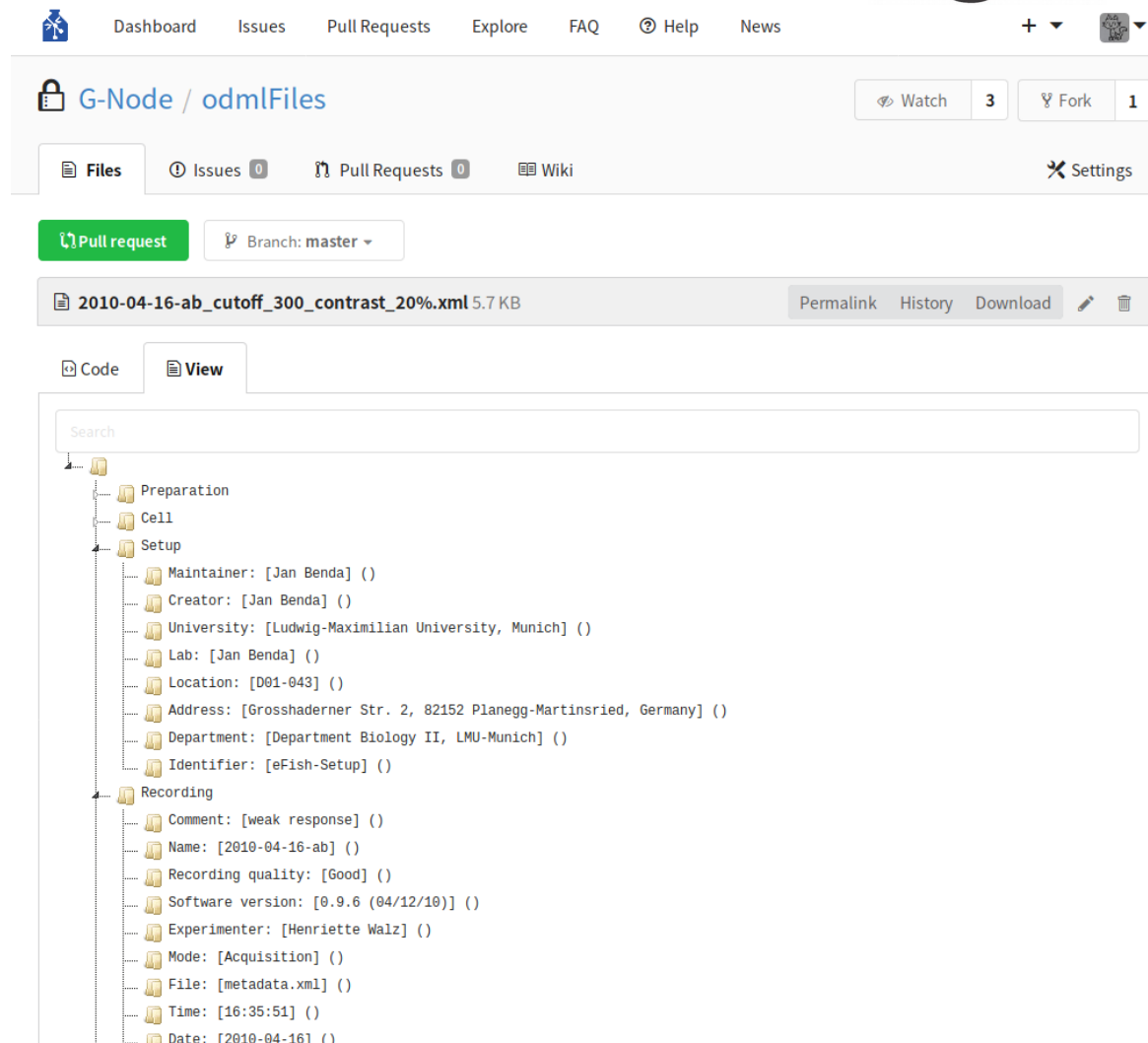
- Versioned management of data repositories (git)
- Distributed data management via file or web browser, or command line client (analysis script integration)
- Secure access and sharing
- Services for search, indexing, and publication (DOI)
- File type plugins
- In house installation
- Extensive usage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

G-Node Open Data			
Registered Datasets			
Title	Date	DOI	
The ICLabel Dataset of electroencephalographic (EEG) independent component (IC) features Pion-Tonachini Luca, Kreutz-Delgado Ken, Makeig Scott	2019-01-31	10.12751/g-node.e3ddb5	
Dorsal Horn Gastrin-Releasing Peptide Expressing Neurons Transmit Spinal Itch But Not Pain Signals Albisetti Gioele, Pagani Martina, Platonova Evgenia, Höslä Ladina, Johannssen Helge C., Fritschy Jean-Marc, Windner Hendrik, Zeilhofer Hanns Ulrich	2019-01-11	10.12751/g-node.50baa6	
Similar Changes in Executive function after Moderate Resistance Training and Loadless Movement Vonk Matthew, Middleton Laura	2018-12-21	10.12751/g-node.91a992	
Who gets lost and why: A representative cross-sectional survey on sociodemographic and vestibular determinants of wayfinding strategies Ulrich Susanne, Grill Eva, Flanagan Virginia L.	2018-12-18	10.12751/g-node.298ce0	
Interhemispheric Integration for Complex Behaviors, Absent the Corpus Callosum in Normal Ontogeny: Supporting Data Layden Elliot, Schertz Kathryn, London Sarah, Berman Marc	2018-12-15	10.12751/g-node.6b6170	
Kuehn_and_Gollisch_RGC_spiketrains_for_moving_texture Kühn Norma Krystyna, Gollisch Tim	2018-12-12	10.12751/g-node.0300fd	
Liu_et_al_2017_RGC_spiketrains_for_STNMF Gollisch Tim, Liu Jian K.	2018-12-11	10.12751/g-node.62b65b	
Oscillatory infrasonic modulation of the cochlear amplifier by selective attention Dragicevic Constantino, Marcenaro Bruno, Navarrete Marcela, Robles Luis, Delano Paul	2018-12-03	10.12751/g-node.be7e46	

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

The screenshot displays the GIN web interface for a repository named 'odmlFiles' under the 'G-Node' organization. The top navigation bar includes links for Dashboard, Issues, Pull Requests, Explore, FAQ, Help, and News. The repository page shows a file named '2010-04-16-ab_cutoff_300_contrast_20%.xml' (5.7 KB) with options for Permalink, History, Download, and a trash icon. Below the file list, there are tabs for 'Code' and 'View'. The 'View' tab is active, showing a tree structure of the repository. The tree includes folders for 'Preparation', 'Cell', 'Setup', and 'Recording'. The 'Setup' folder contains metadata fields such as Maintainer, Creator, University, Lab, Location, Address, Department, and Identifier. The 'Recording' folder contains fields like Comment, Name, Recording quality, Software version, Experimenter, Mode, File, Time, and Date.

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

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

/FaqTroubleshooting
/UsageTutorial
/Webinterface
/DOIfile
/GinCliRecipes
/InHouse

GIN: Services for Data Organization and Sharing



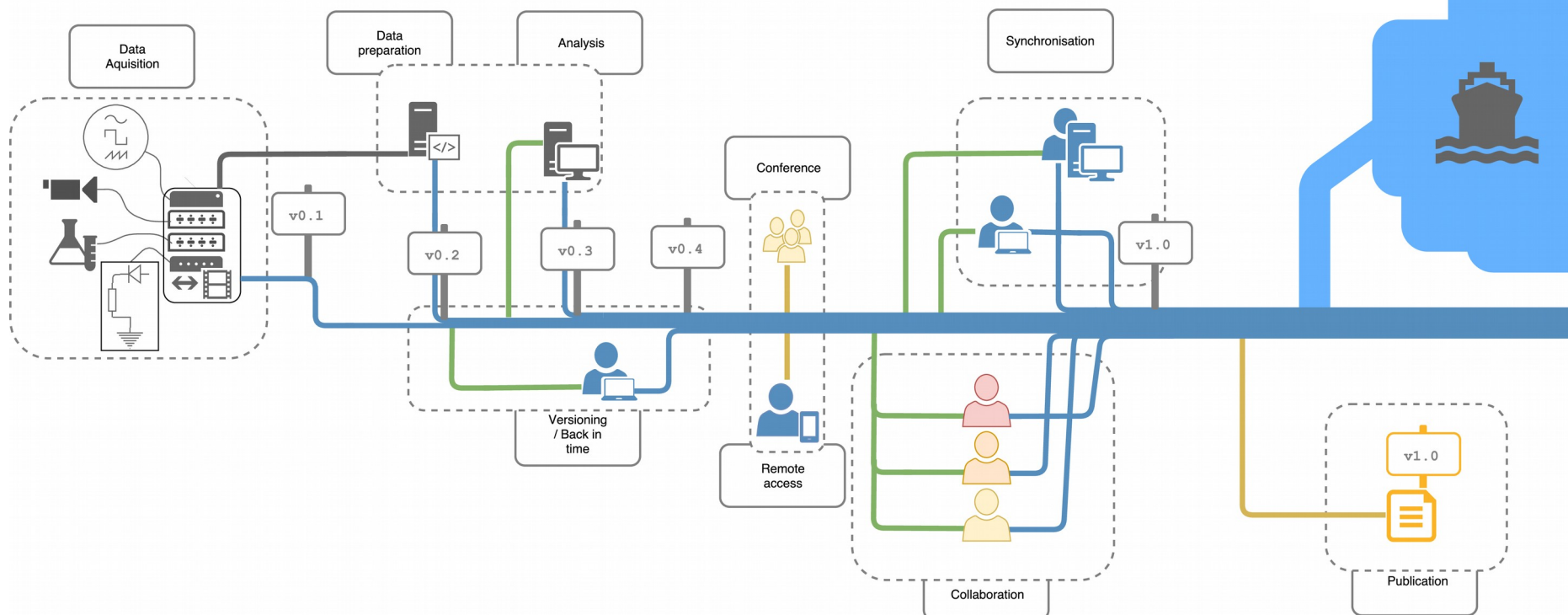
Supporting research data management through the entire data lifecycle

Data Acquisition and Processing

Data Analysis

Data Access and Collaboration

Data Publication



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 command line client 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 Schmücker, Yaroslav Shalivsky,
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>

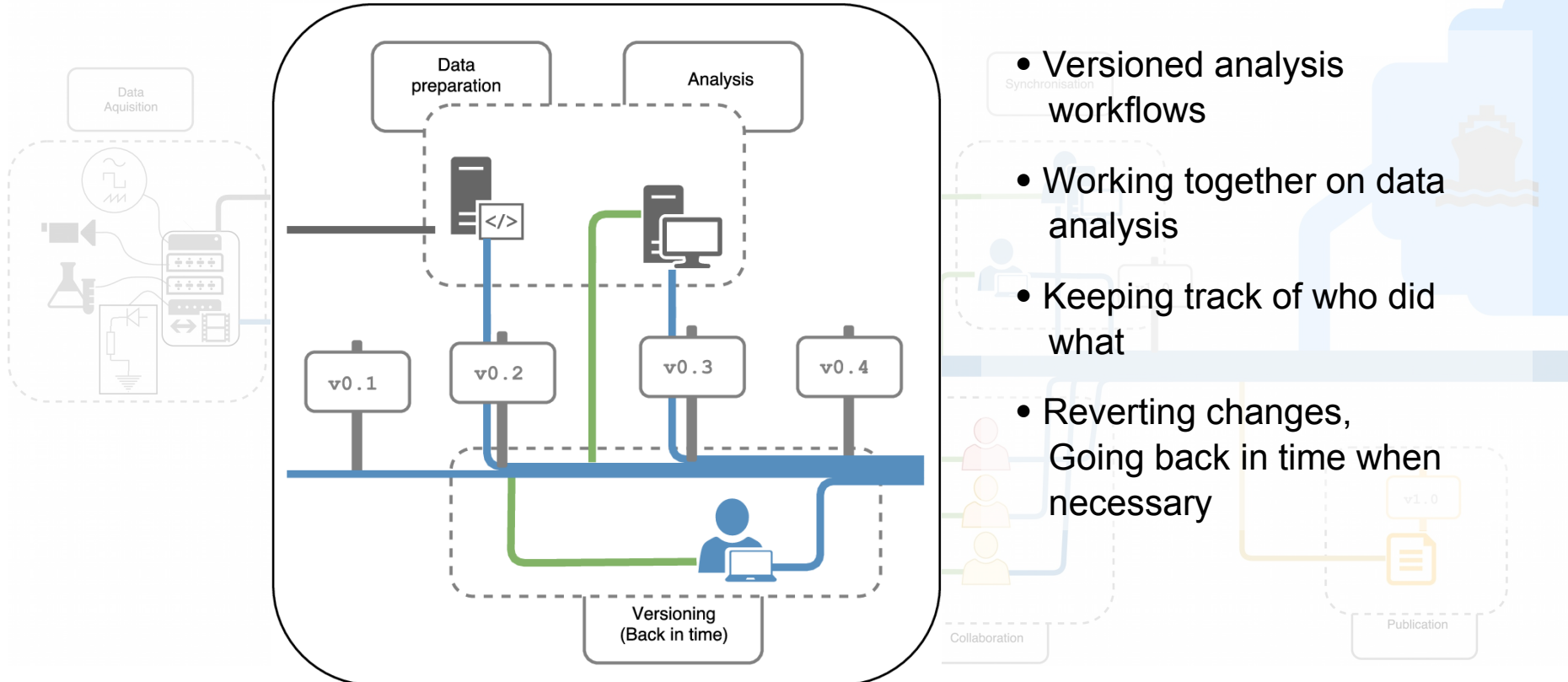
-
- The diagram illustrates a data management system architecture. On the left, a dashed box labeled "Data Acquisition" contains icons for a waveform, a camera, a flask, and a battery. These are connected to a central server icon. A line labeled "v0.1" connects the server to a larger box on the right. This box contains a "Remote access" section with a person icon, a "Collaboration" section with three person icons, and a "Publication" section with a document icon labeled "v1.0". A large blue cloud icon with a ship is on the far right. The background features faint icons for "Analysis", "Synchronisation", "Conference", "Versioning back in time", and "v0.3".
- Versioning of datasets from the beginning
 - Workflow automation
 - 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



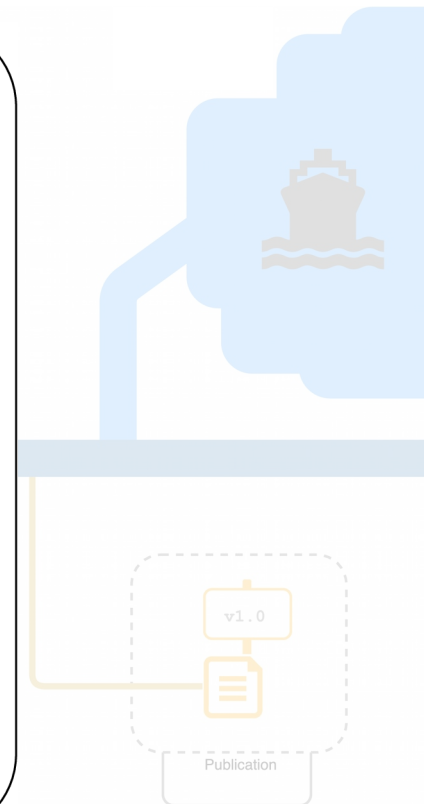
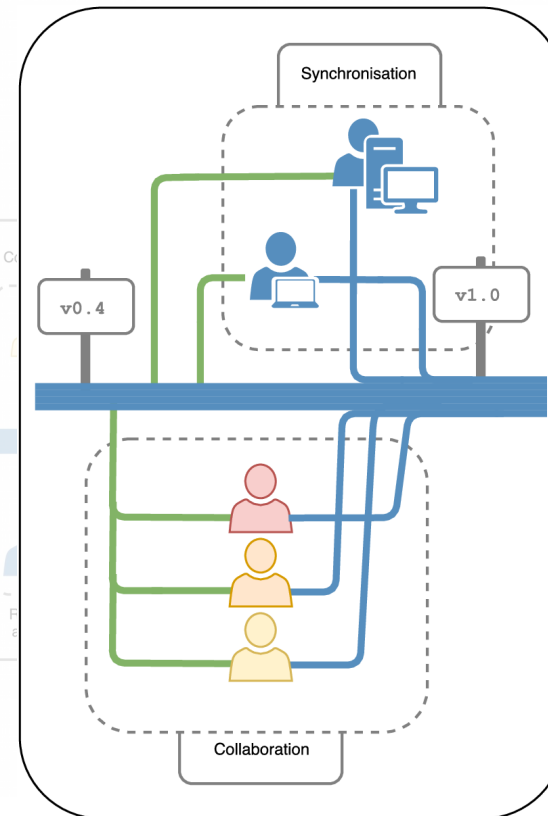
GIN: Services for Data Organization and Sharing



Supporting research data management through the entire data lifecycle

Data Access and Collaboration

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



GIN: Services for Data Organization and Sharing

Supporting research data management through the entire data lifecycle



Data Publication

