



# The Neurodata Without Borders Ecosystem for Neurophysiology Data Standardization

Driving Collaboration in Neuroscience

Oliver Rübel

Computational Biosciences Group  
Scientific Data Division  
Lawrence Berkeley National Laboratory



**BERKELEY LAB**  
Bringing Science Solutions to the World

# Overview

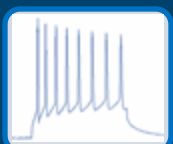
- Introduction to NWB
- NWB for neuroscience researchers
- NWB for neuroscience tools and technologies
- Summary: Current State of NWB

# Multidisciplinary team science at work

## Applications



Extracellular electrophysiology



Intracellular electrophysiology



Optical physiology




Behavior




Simulations

## Technology Teams


**BERKELEY LAB**




O. Rübel



S. Prince



R. Ly



M. Avaylon

**CATALYST NEURO**




B. Dichter



C. Baker



A. Weigl




E. Hennestad

**ALLEN INSTITUTE FOR BRAIN SCIENCE**



S. De Vries

**DataJoint**



A. Buccino

**DARTMOUTH**




K. Bakshi




Y.O. Halchenko


## NWB Executive Board




K. Bouchard (LBNL)




B.W. Brunton (UW)



E. Buffalo (UW)



A. Churchland (UCLA)



L. M. Frank (UCSF)



S. Ghosh (MIT)



A. Kepecs (WUSTL)



M. Murthy (Princeton)



U. Rutishauser (Cedars Sinai)

**Alumni:** L. Ng, C. Koch, F. Sommer, K. Svoboda, M. Meister, K. Amunds, H.D. Mansvelder

## Technical Advisory Board

Saskia de Vries, Anna (Szonja) Weigl, Alessio Buccino, Yaroslav O. Halchenko, Ryan Ly, Oliver Ruebel, Lawrence Niu

## Sponsors



## Industry Engagement

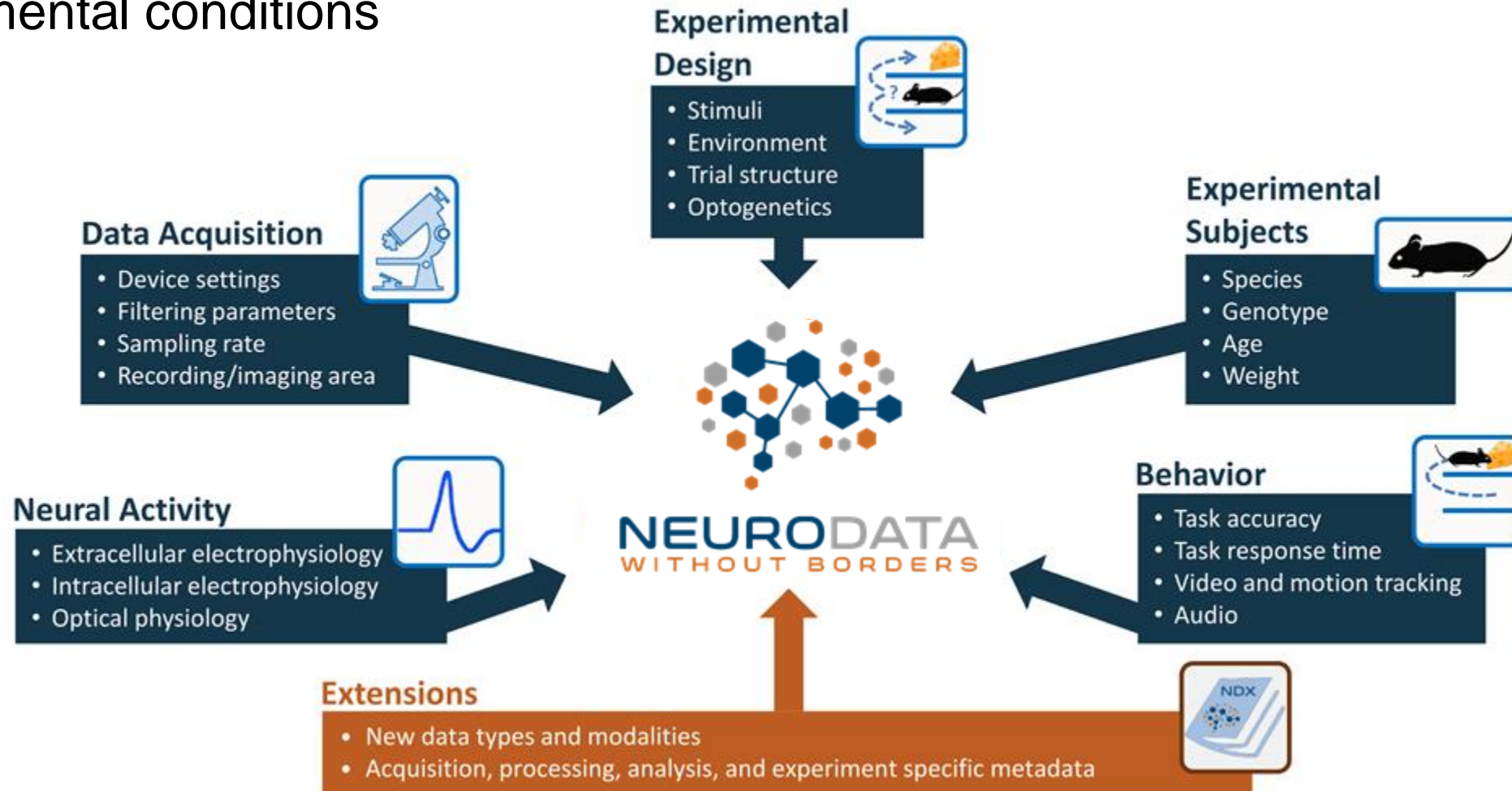
- CatalystNeuro
- DataJoint
- MBF Bioscience
- MathWorks
- Kitware

## Broader User and Developer Community

**Apologies if your name/team is missing!**  
This slide only shows a very rough cut of some of the teams and people that work on developing NWB. Pictures and names of many, many important members of the NWB community are missing!

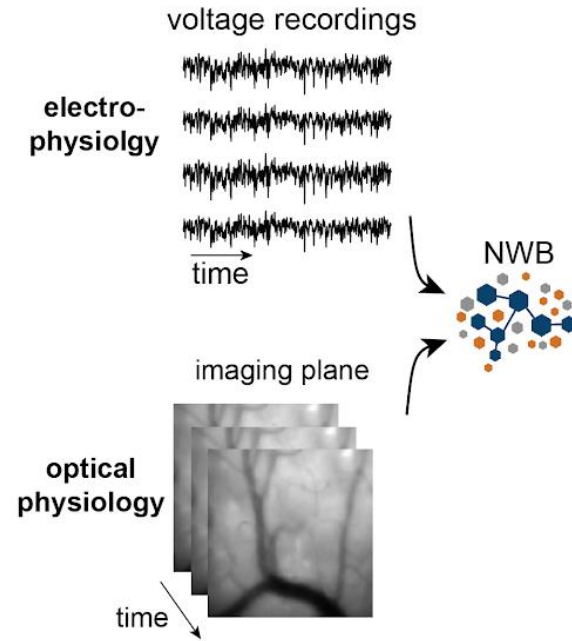
# A unified data standard for neurophysiology

The NWB data standard defines a unified data format for neurophysiology data, focused on the dynamics of groups of neurons measured under a large range of experimental conditions



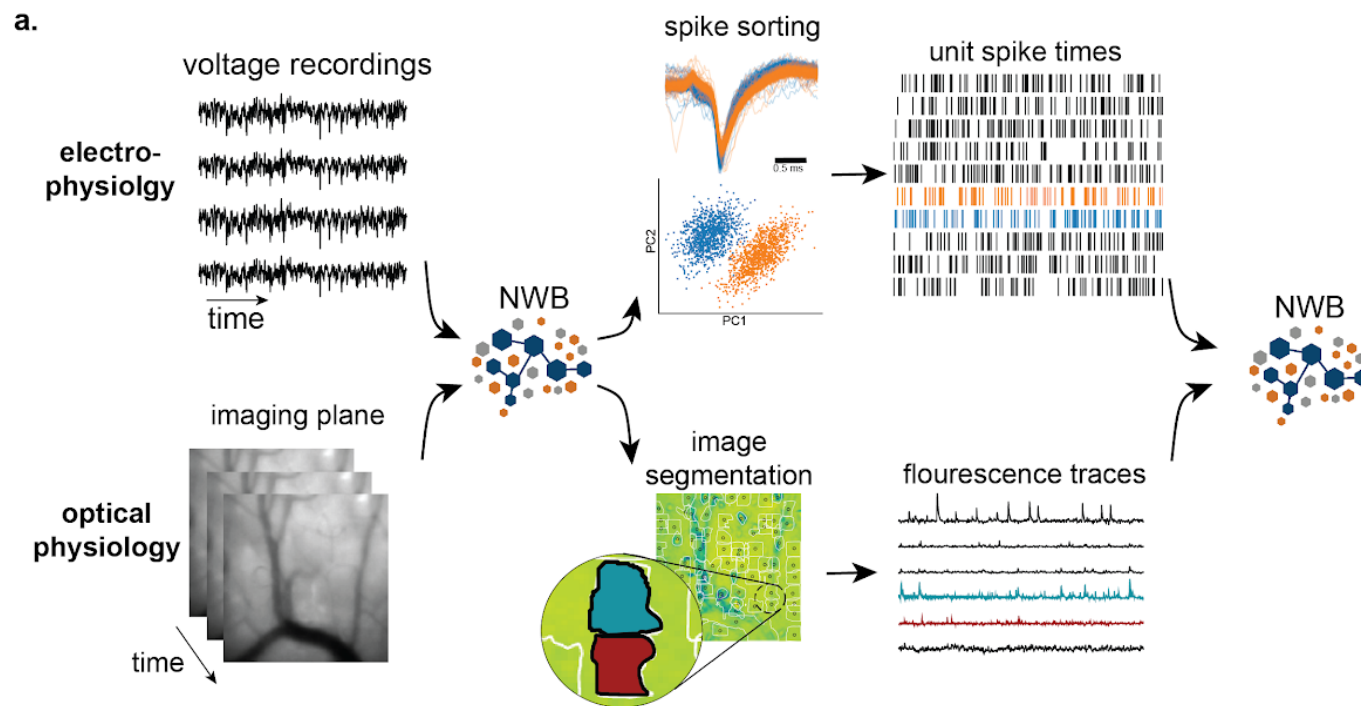
# NWB enables unified storage of multimodal raw and processed data

a.

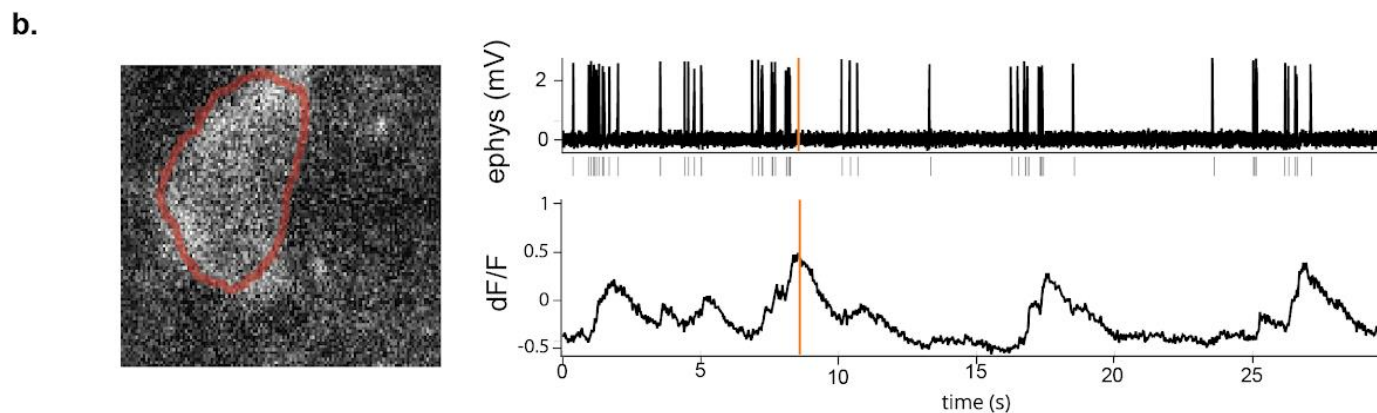
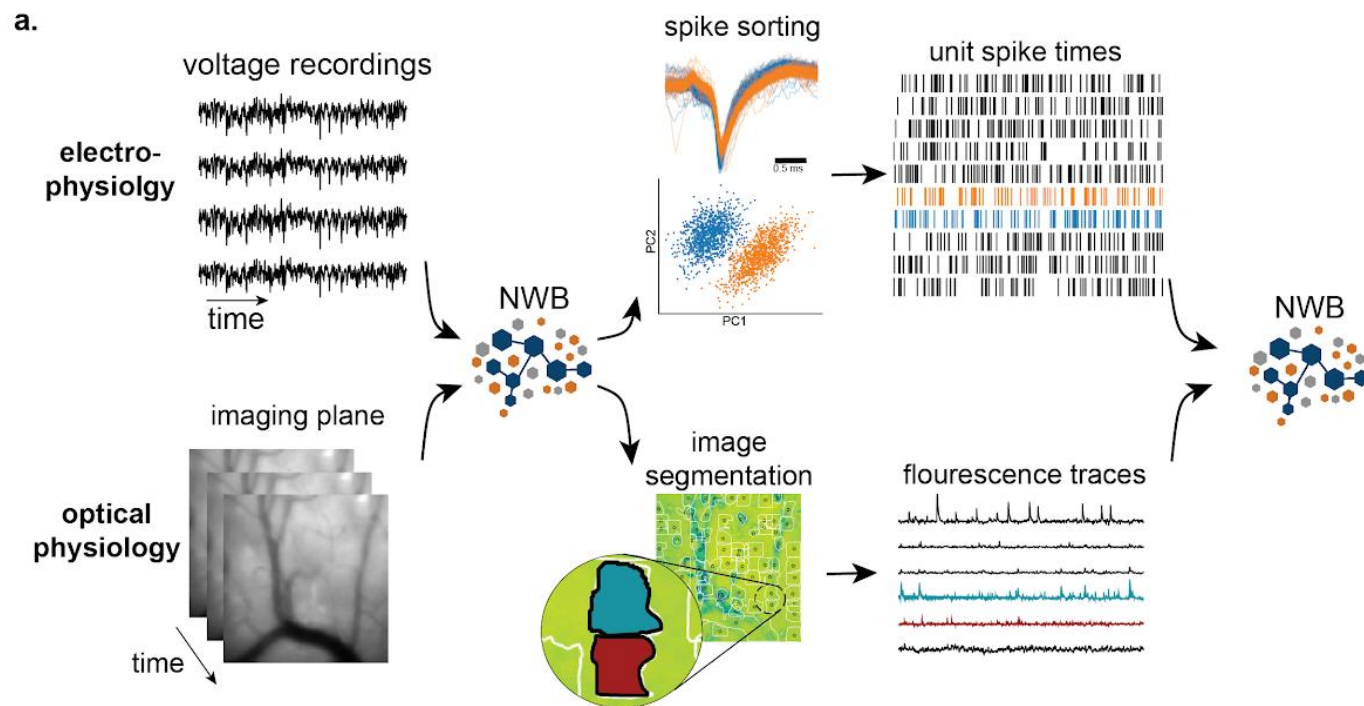




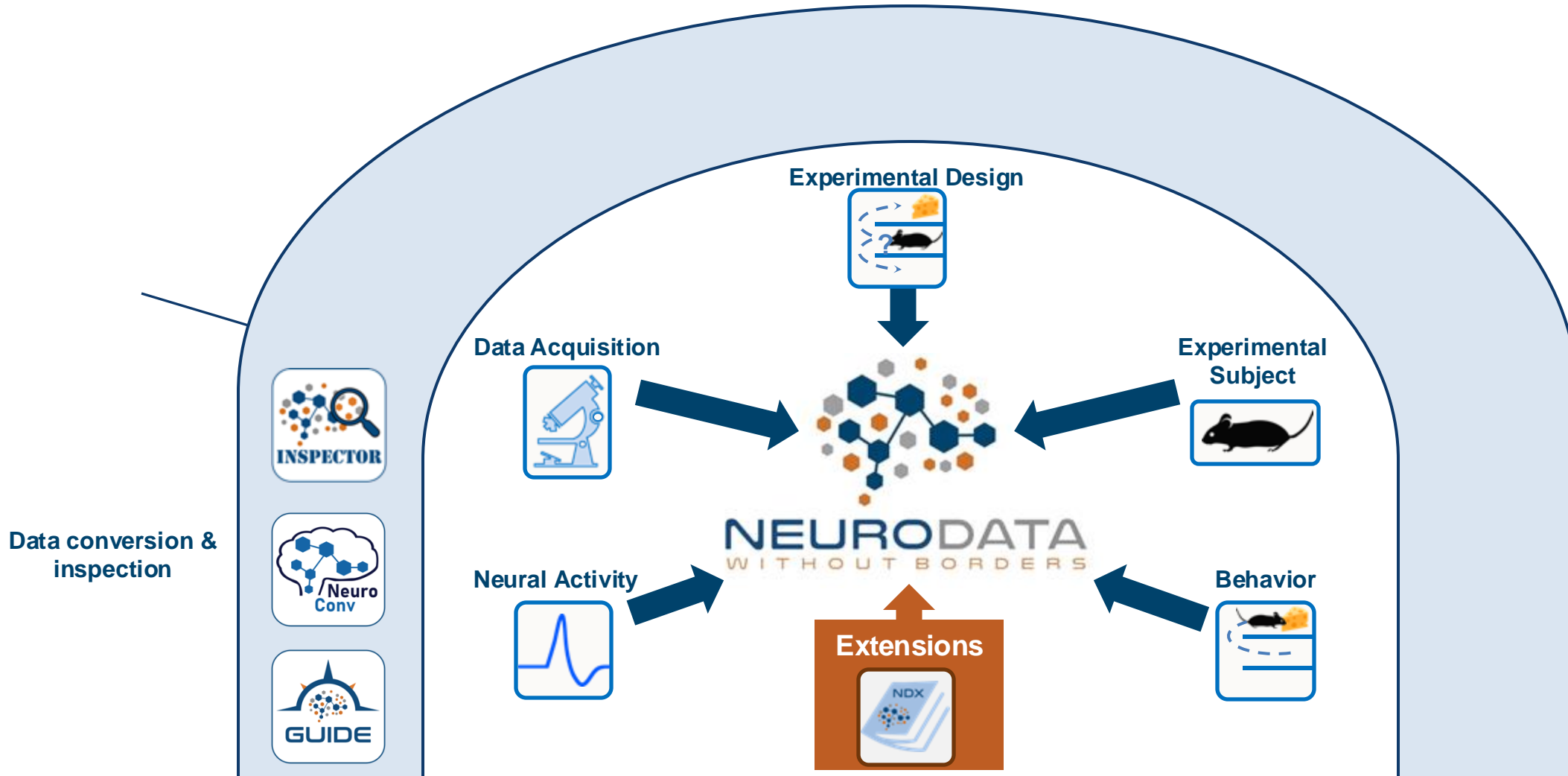
# NWB enables unified storage of multimodal raw and processed data



# NWB enables unified storage of multimodal raw and processed data

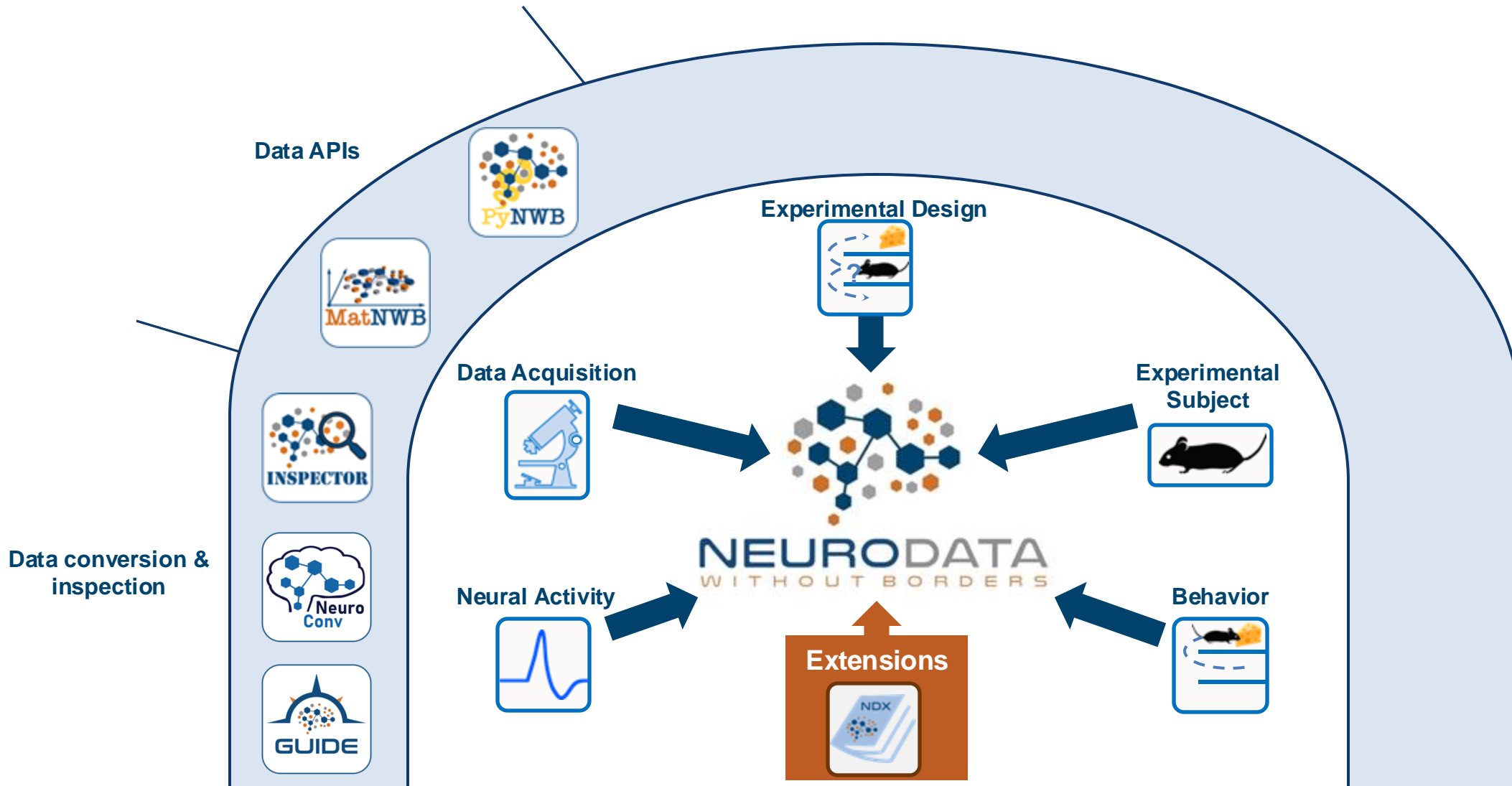


# An ecosystem for neuroscience data standardization

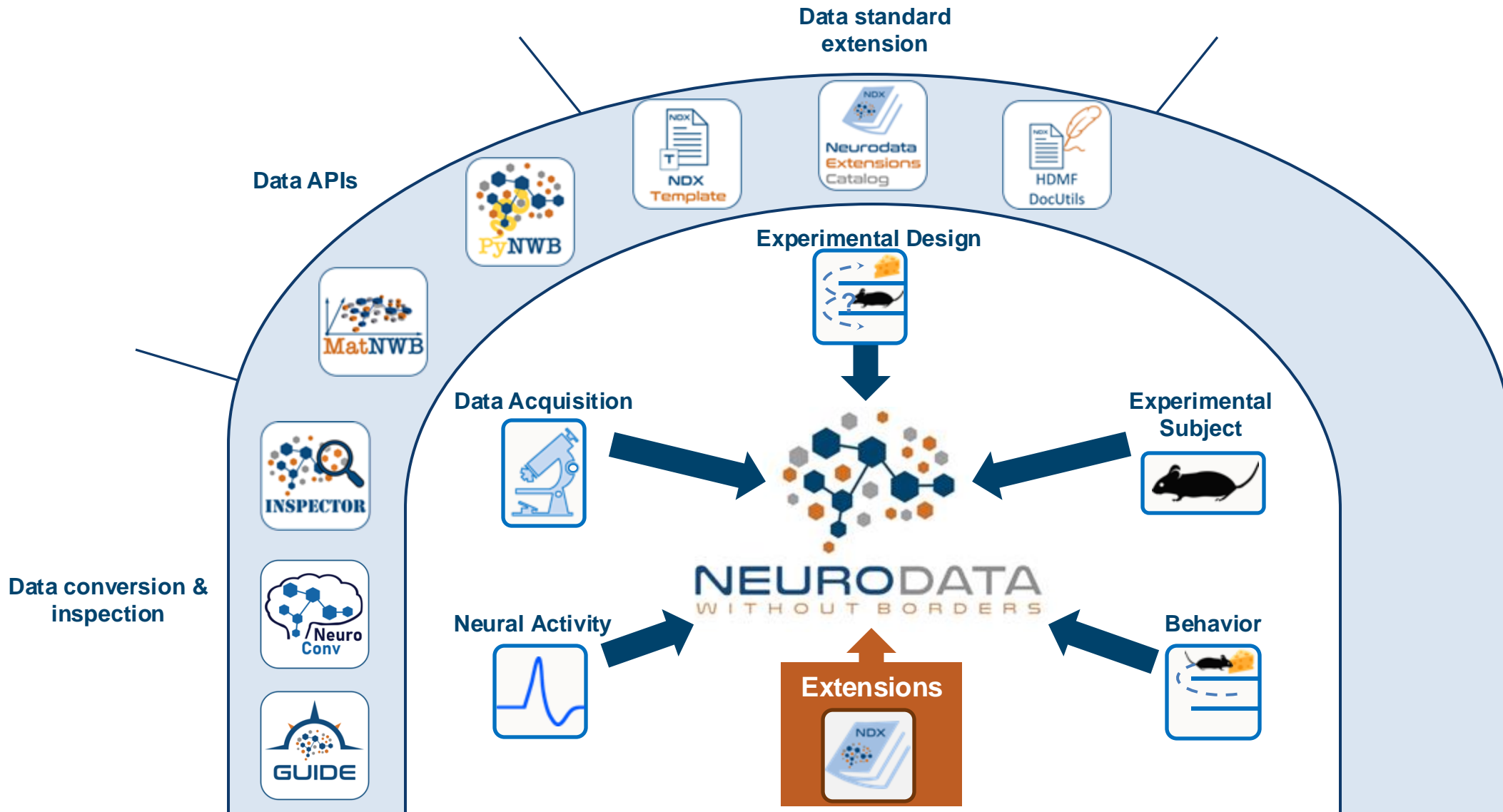




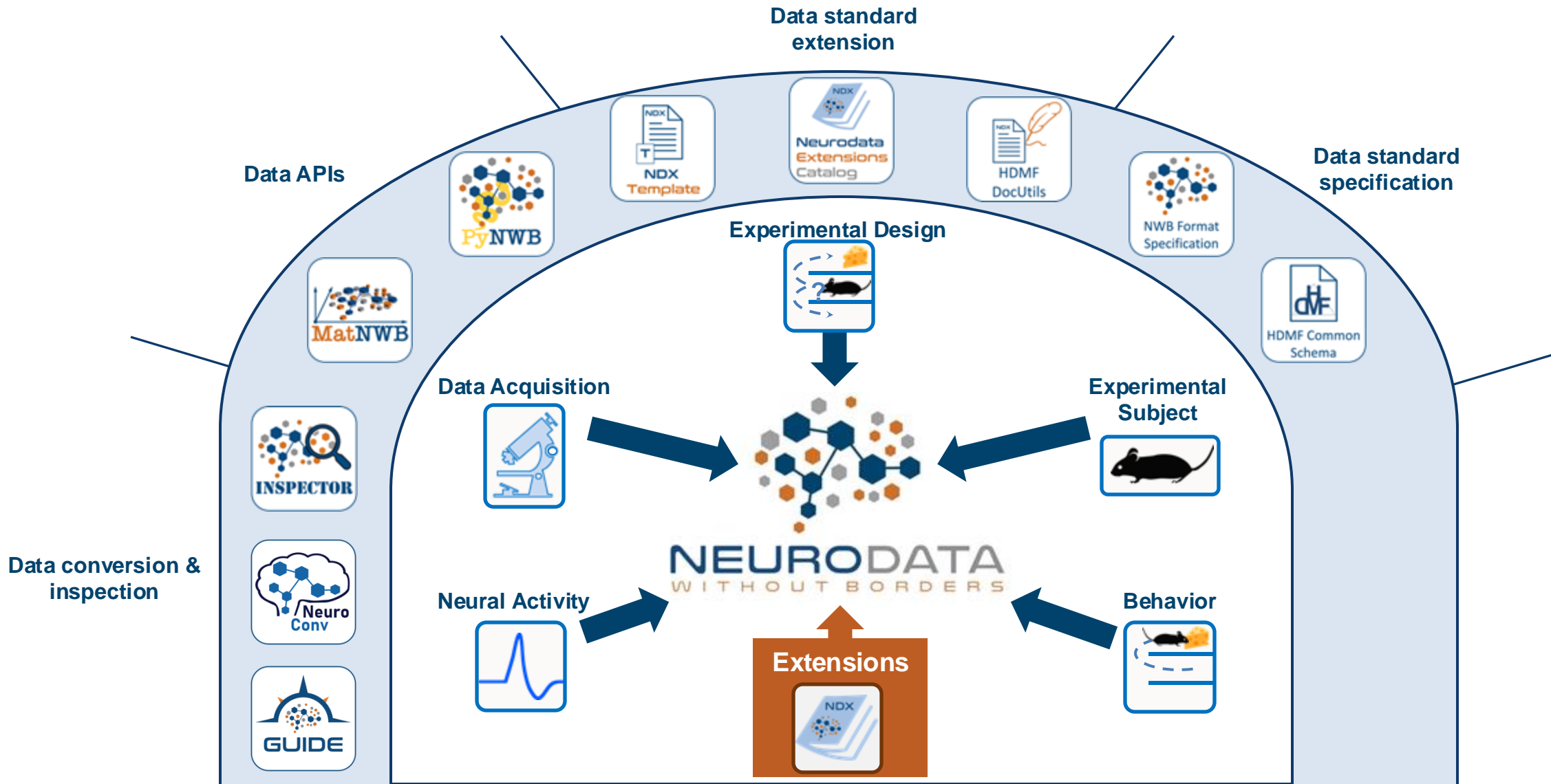
# An ecosystem for neuroscience data standardization



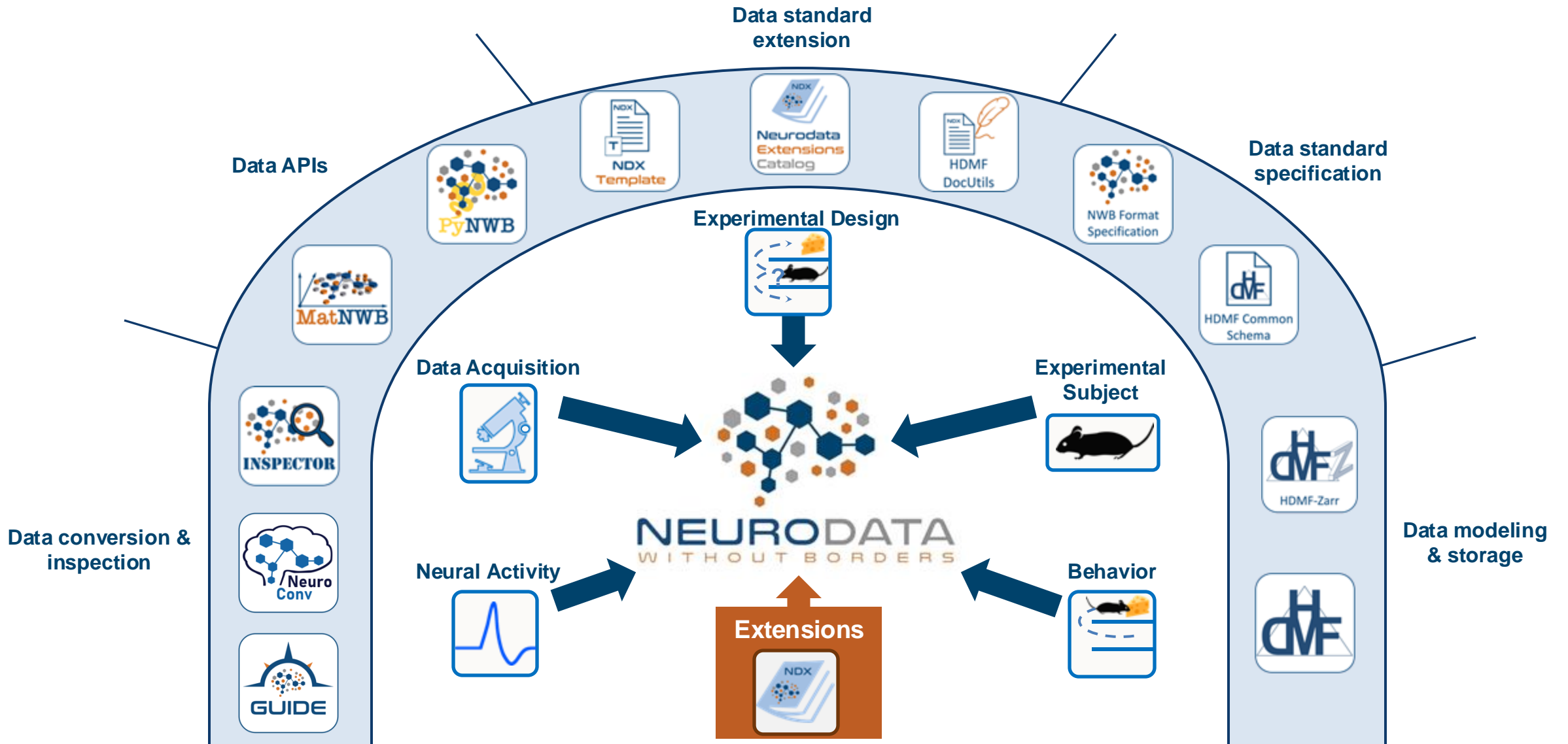
# An ecosystem for neuroscience data standardization



# An ecosystem for neuroscience data standardization



# An ecosystem for neuroscience data standardization



# NWB for neuroscience researchers

Enable researchers to effectively utilize and integrate their data with the NWB data standard and software

# Converting data to NWB



Custom code  
using NWB APIs



Low-code  
automated  
conversion  
for common input  
data formats



No-code automated  
conversion via graphical  
interface app  
(for select input data  
formats, in beta)

**Flexible**

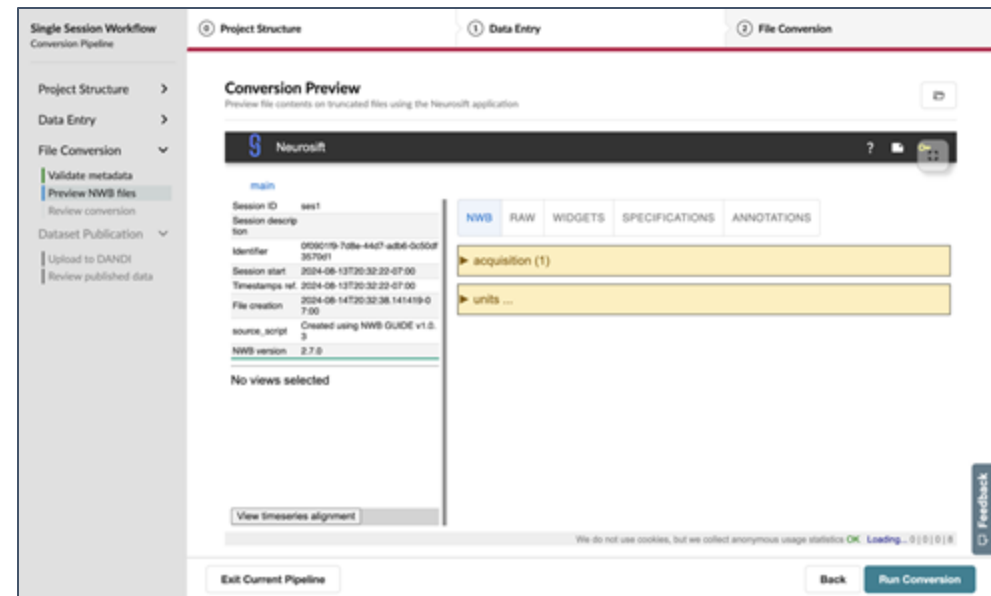
**Automated**





# TurboTax-like interface for converting data to NWB

- NWB GUIDE 1.0 officially released June 10, 2024 together with the Data Conversion Workshop
- Key Features:
  - Convert data from 40+ different formats to NWB using **NeuroConv**
  - Validate and scan NWB files for best practices using **NWB Inspector**
  - Visualize and interact with local NWB files using **Neurosift**
  - Upload data to **DANDI Archive**



# Automated NWB conversions with NeuroConv



## Electrophysiology

PLEXON

intan  
TECHNOLOGIES

Spike2

Neuralynx  
An FHC Company

open ephys

TDT  
TUCKER-DAVIS TECHNOLOGIES

SpikeGadgets

Blackrock  
Neurotech

## Optical physiology

Miniscope



## Behavior

LIGHTNING  
POSE

DeepLabCut<sup>™</sup>  
a software package for  
animal pose estimation



- Convert data to NWB from 44+ different formats with minimal time and effort
- Automatically extracts format-specific metadata
- Seamlessly support conversion of TB scale data and optimization of data via chunking and compression

```
from datetime import datetime
from dateutil import tz
from pathlib import Path
from neuroconv.converters import SpikeGLXConverterPipe

converter = SpikeGLXConverterPipe(
    folder_path="spikeglx/Noise4Sam_g0")

metadata = converter.get_metadata()
session_start_time =
metadata["NWBFile"]["session_start_time"].replace(
    tzinfo=tz.gettz("US/Pacific"))
metadata["NWBFile"].update(
    session_start_time=session_start_time)

converter.run_conversion(
    nwbfile_path="my_spikeglx_session.nwb",
    metadata=metadata)
```



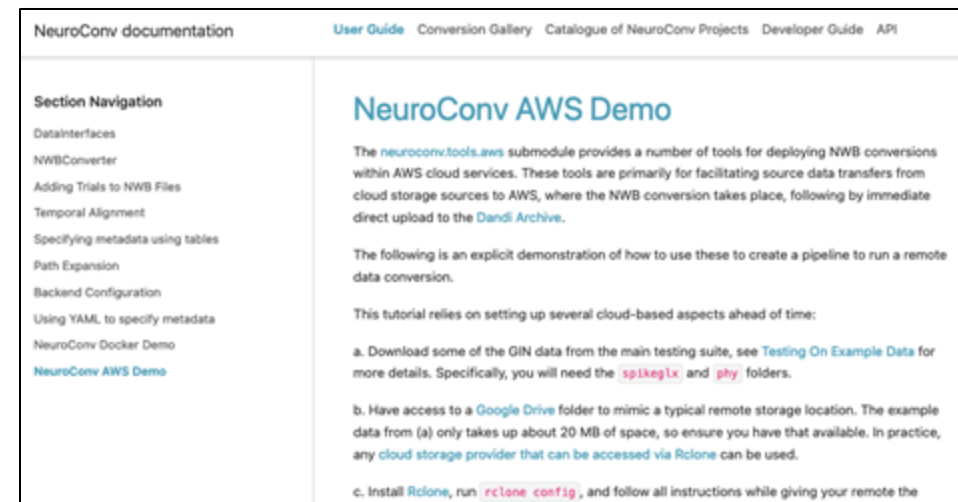
NeuroConv  
Conversion  
Gallery



# Enabling efficient data conversion using cloud-resources

- **Automated building and deployment of NeuroConv Docker images** via automated workflows for building and deploying NeuroConv Docker images for all releases and daily for the main development branch
- **Created `neuroconv.tools.aws` module** to simplify launching and managing NeuroConv data conversions on AWS EC2, incl:
  - Setting up AWS EC2 Batch infrastructure
  - Transferring data based on RClone for use in cloud environments
  - Tracking of batch job statuses via DynamoDB and AWS dashboard
- **Developed continuous integration `pytest` pipelines** to ensure proper functionality of all NeuroConv cloud modules
- **Created online documentation and tutorials** for using NeuroConv docker images and cloud data conversion
- **Impact:** Using these new capabilities users can easily launch container images on on-demand EC2 instances and:
  - Transfer data from remote cloud storage to EC2 using RClone
  - Convert data to NWB and upload results to DANDI
  - Automatically spin down and clean up resources to minimize

## AWS Demo



NeuroConv documentation [User Guide](#) [Conversion Gallery](#) [Catalogue of NeuroConv Projects](#) [Developer Guide](#) [API](#)

### Section Navigation

- Datainterfaces
- NWBConverter
- Adding Trials to NWB Files
- Temporal Alignment
- Specifying metadata using tables
- Path Expansion
- Backend Configuration
- Using YAML to specify metadata
- NeuroConv Docker Demo
- NeuroConv AWS Demo**

## NeuroConv AWS Demo

The `neuroconv.tools.aws` submodule provides a number of tools for deploying NWB conversions within AWS cloud services. These tools are primarily for facilitating source data transfers from cloud storage sources to AWS, where the NWB conversion takes place, following by immediate direct upload to the [Dandi Archive](#).

The following is an explicit demonstration of how to use these to create a pipeline to run a remote data conversion.

This tutorial relies on setting up several cloud-based aspects ahead of time:

- Download some of the GIN data from the main testing suite, see [Testing On Example Data](#) for more details. Specifically, you will need the `spikeglx` and `phy` folders.
- Have access to a [Google Drive](#) folder to mimic a typical remote storage location. The example data from (a) only takes up about 20 MB of space, so ensure you have that available. In practice, any cloud storage provider that can be accessed via Rclone can be used.
- Install [Rclone](#), run `rclone config`, and follow all instructions while giving your remote the

## DynamoDB table

id (String)	job_id	job_name	status	submitted_on
11...	6686f75-b49c-4e70-af65-c3e5e42a9a7	test_submit_aws_batch_job	Test passed	2024-09-05T19:03:39.758...
a3960f7...	8c60bc3-52c2-4af5-b674-b966f1cae205	test_submit_aws_batch_job_with_dependencies_1	Test passed	2024-08-08T18:42:33.996...
d667566f...	64e34405-be15-4177-b5f6-359927544b4f	test_submit_aws_batch_job_with_dependencies_1	Test passed	2024-09-28T16:00:56.956...
db922065...	e019d854-d866-4e17-82c2-4ba03caf81a5	test_rclone_transfer_batch_job_2024-09-28	Job submitted...	2024-09-28T15:05:23.722...

## Additional Resources

- **NeuroConv Dockerfiles:** <https://github.com/catalystneuro/neuroconv/tree/main/dockerfiles>
- **NeuroConv unit tests:** <https://github.com/catalystneuro/neuroconv/tree/main/tests>
- **Building/deploying docker images:**
  - **Workflows:** <https://github.com/catalystneuro/neuroconv/tree/main/github/workflows>
  - **Docs:** [https://neuroconv.readthedocs.io/en/main/developer\\_guide/docker\\_images.html](https://neuroconv.readthedocs.io/en/main/developer_guide/docker_images.html)
- **Docker demo:** [https://neuroconv.readthedocs.io/en/main/user\\_guide/docker\\_demo.html](https://neuroconv.readthedocs.io/en/main/user_guide/docker_demo.html)

# Use NWB APIs to write/read NWB data

- Most useful for converting custom data
- Also trial times, responses, and properties
- APIs support:
  - Iteratively reading and writing data
  - Custom chunking
  - Custom compression
  - Writing data in NWB extensions
- Lots of documentation and tutorials!



[pynwb.readthedocs.io](https://pynwb.readthedocs.io)



[neurodatawithoutborders.github.io/matnwb/](https://neurodatawithoutborders.github.io/matnwb/)

# NWB Inspector:

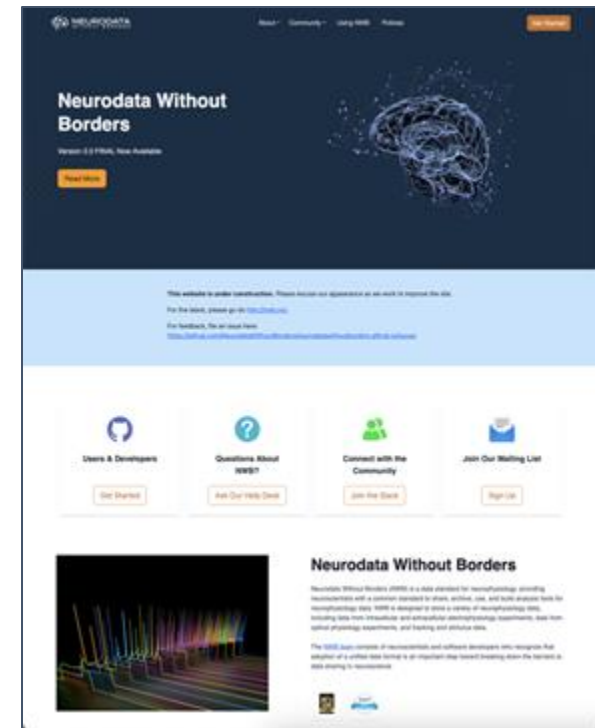
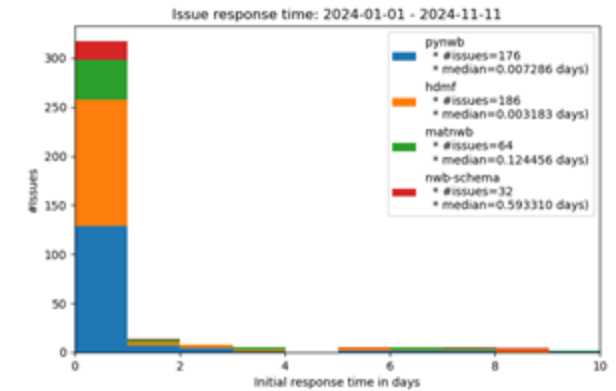
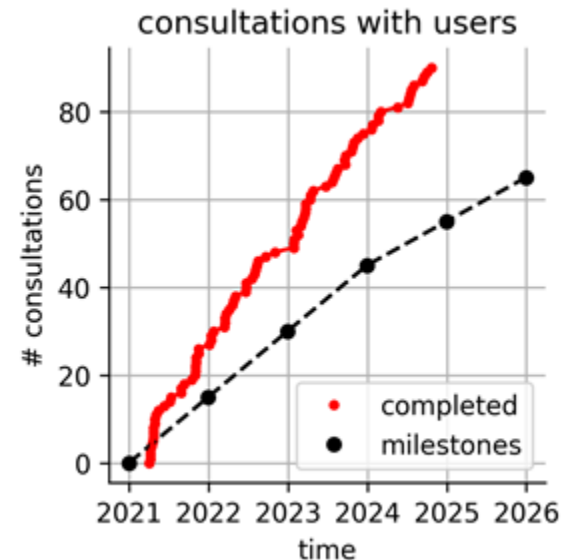
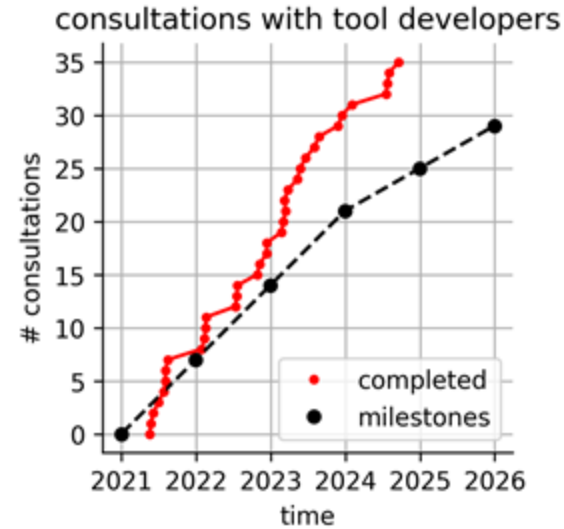
Extensible tool for inspecting NWB data files for compliance with best practices



- The NWB Inspector defines automated checks to:
    - find mistakes in data (e.g., bad units, inconsistent or transposed data ...)
    - check for compliance with NWB best practices
    - identify potential enhancements
    - ensure files are compliant with the schema (using the PyNWB validator)
  - Extensible and customizable (e.g., to add checks for extensions)
  - Integrated with DANDI upload process and NWB GUIDE
- 
- **Docs:** <https://nwbinspector.readthedocs.io>
  - **Repo:** <https://github.com/neurodatawithoutborders/nwbinspector>

# Ensure production readiness and support through optimization, standardization and user support

- Provide support for users via GitHub issues, online helpdesk, Slack etc.
- Conduct consultations with users and tool developers
- Enhance code test coverage, covering now **>90%** of PyNWB, MatNWB, and HDMF
- Enhance tutorials and documentation covering now **100%** of user-facing NWB data types and functionality
- Enhance accessibility of NWB documentation
  - We are modernizing our nwb.org website. See <https://neurodatawithoutborders.github.io> for the upcoming new website and provide feedback.

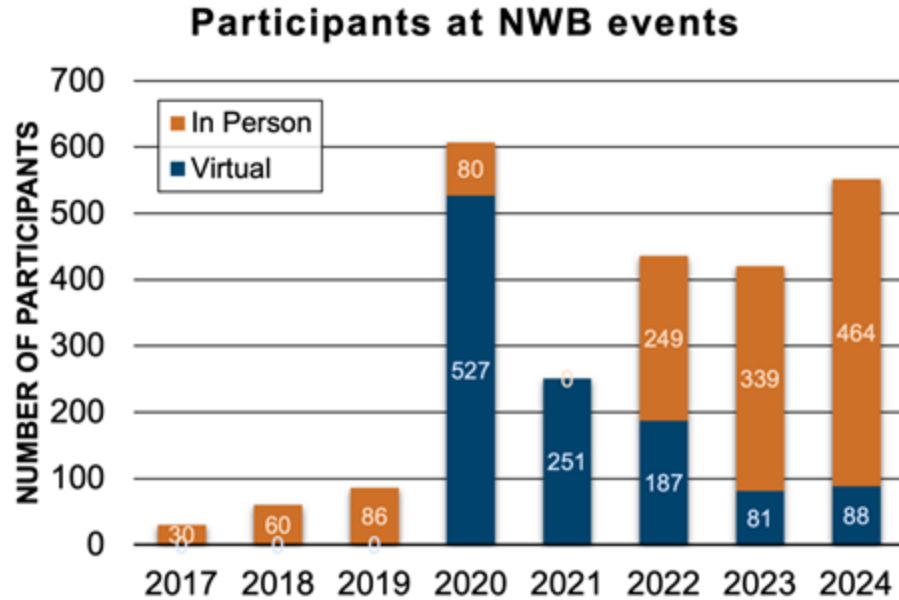




# NWB Community Events

[https://neurodatawithoutborders.github.io/nwb\\_hackathons](https://neurodatawithoutborders.github.io/nwb_hackathons)

Enable neuroscience researchers and developers to effectively utilize the NWB data standard and software



NWB Data Conversion Workshop 2025

May 12-14, 2025 (Virtual)

May 13: Open Neurodata and Tools Showcase



NeuroDataReHack 2025

July 14-18, 2025

Janelia Research Campus, Ashburn, VA



THE  
KAVLI  
FOUNDATION

NeuroData Discovery  
Award 2025



Cosyne 2025 Tutorial

March 27, 2025, 9:00 - 10:30 a.m.

Fairmount, the Queen Elizabeth  
Montreal, Canada



March 25-26: Pre-Cosyne Brainhack - Montreal, Canada

<https://pre-cosyne-brainhack.github.io/hackathon2025/>

- Oct. 5-9: Booth at SfN 2025, Chicago, IL, USA
- Roadshows 2025: Mexico City, (Tentative: UW, NYU, UCB)

NWB Community Hackathon 2025

August 4-6, 2025

Flatiron Institute, New York, USA

Registration  
not yet open.

Want to help with training and events?

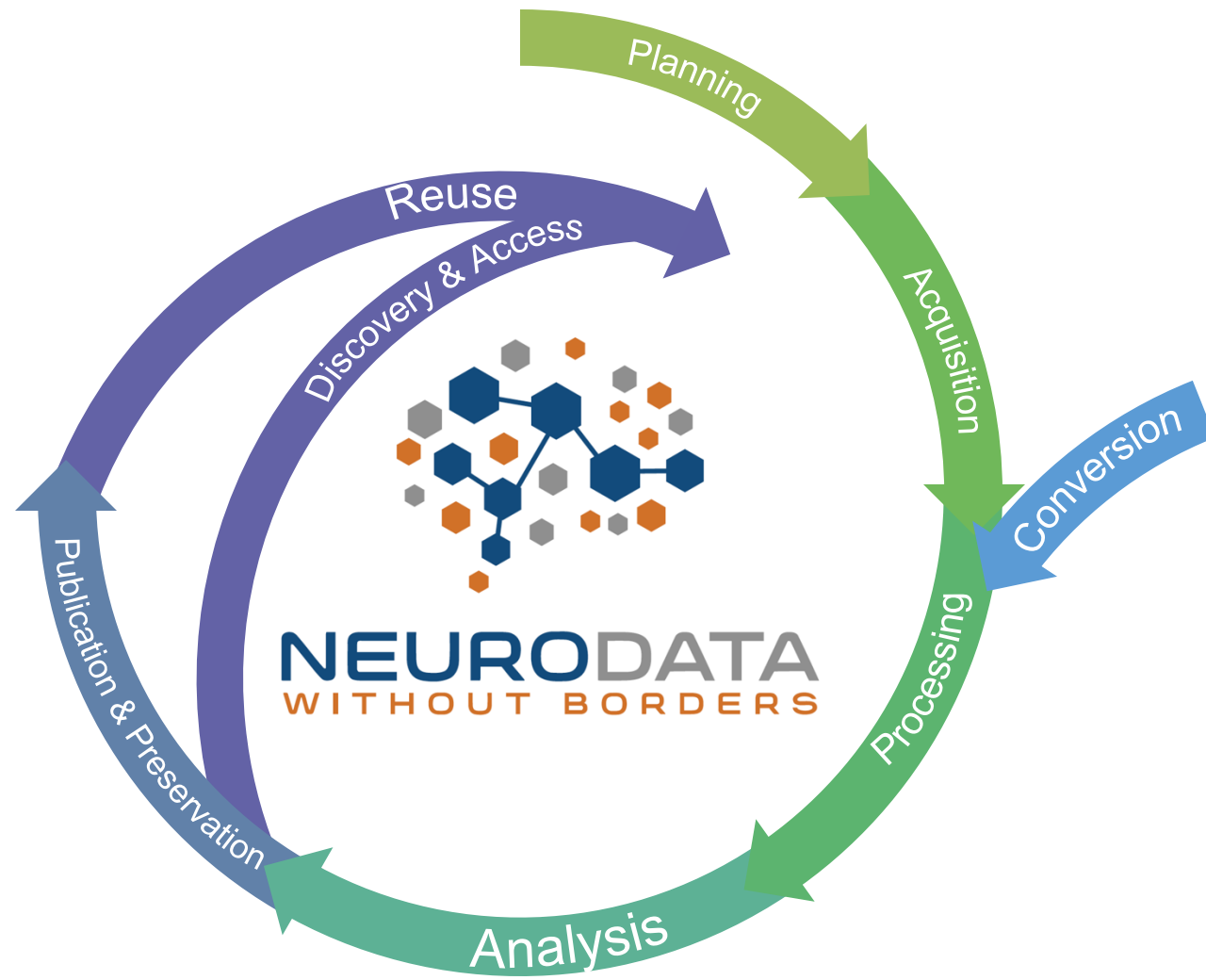
Want to host an NWB Roadshow? Want to help teach others about NWB? Please fill out this form to let us know about your interest to give back to the community.



# NWB for neuroscience tools and technologies

Enable integration of NWB with neurophysiology data analysis and management tools

# NWB technologies at the heart of the neurodata lifecycle



- Data standards are a critical conduit that facilitate the:
  - Flow of data throughout the data lifecycle
  - Integration of data and software across phases of the data lifecycle
- NWB needs to support the needs of, and integrate with, technologies across the data lifecycle:
  - Work with (not compete with) existing and emerging data technologies
- NWB is a data standard for (not a standard of) neurophysiology experiments

# DANDI: An Archive and Collaboration Space for Neurophysiology Projects

- **DANDI:** The BRAIN Initiative archive for publishing and sharing neurophysiology
  - Stores neurophysiology data and associated behavioral data in NWB format
  - Validates NWB, checks for best practices, automatically extracts metadata
- **Upload 100s of TBs for FREE**
  - Can handle all raw data for most experiments
  - AWS Open Data Sponsorship Program
- Open licenses: CC-0 or CC-BY
- Can upload data in embargo mode
- Create DOIs by “publishing” and re-publishing



<https://dandiarchive.org>

778

[dandisets](#)

1521

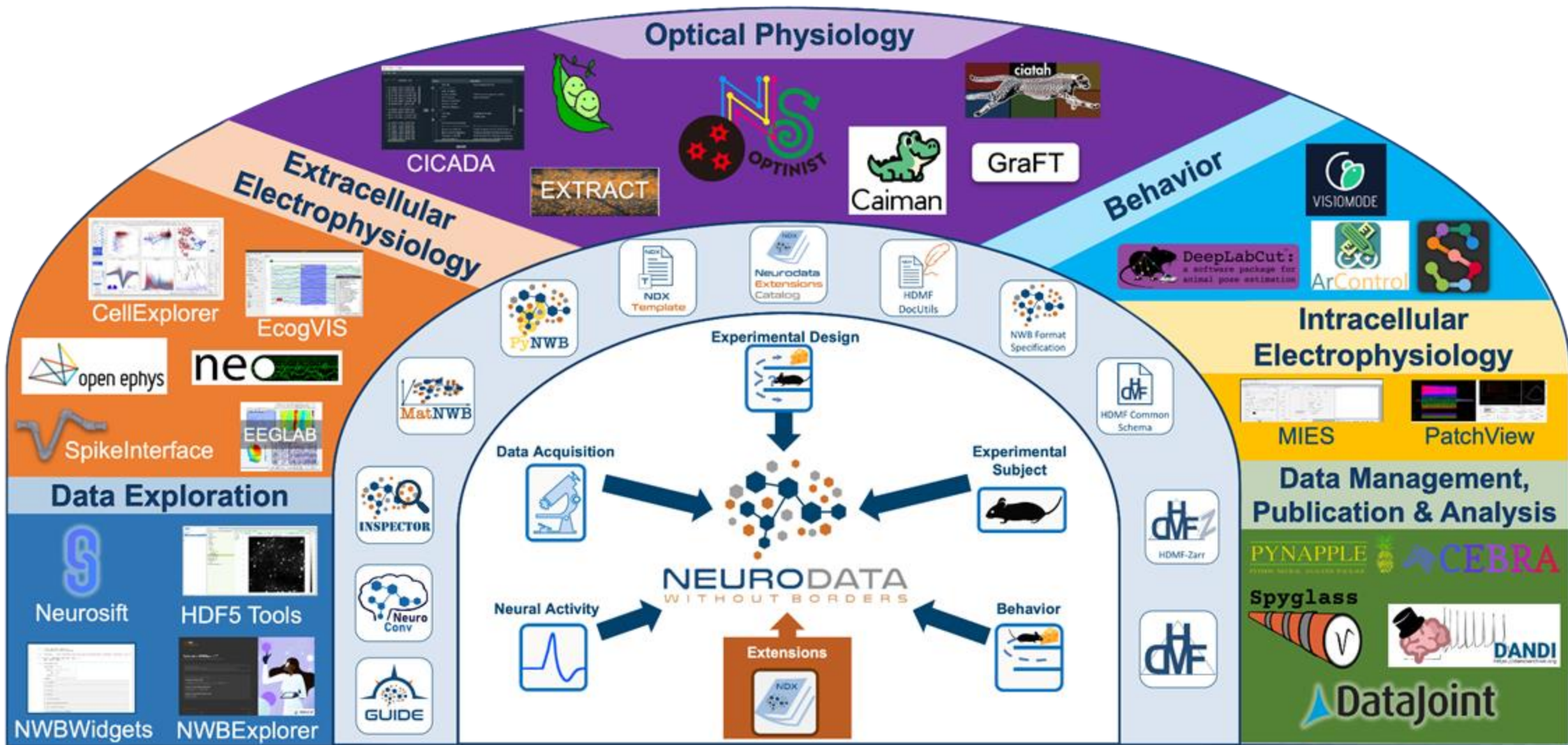
users

868 TB

total data size

# NWB Community Tools

30+ tools support NWB

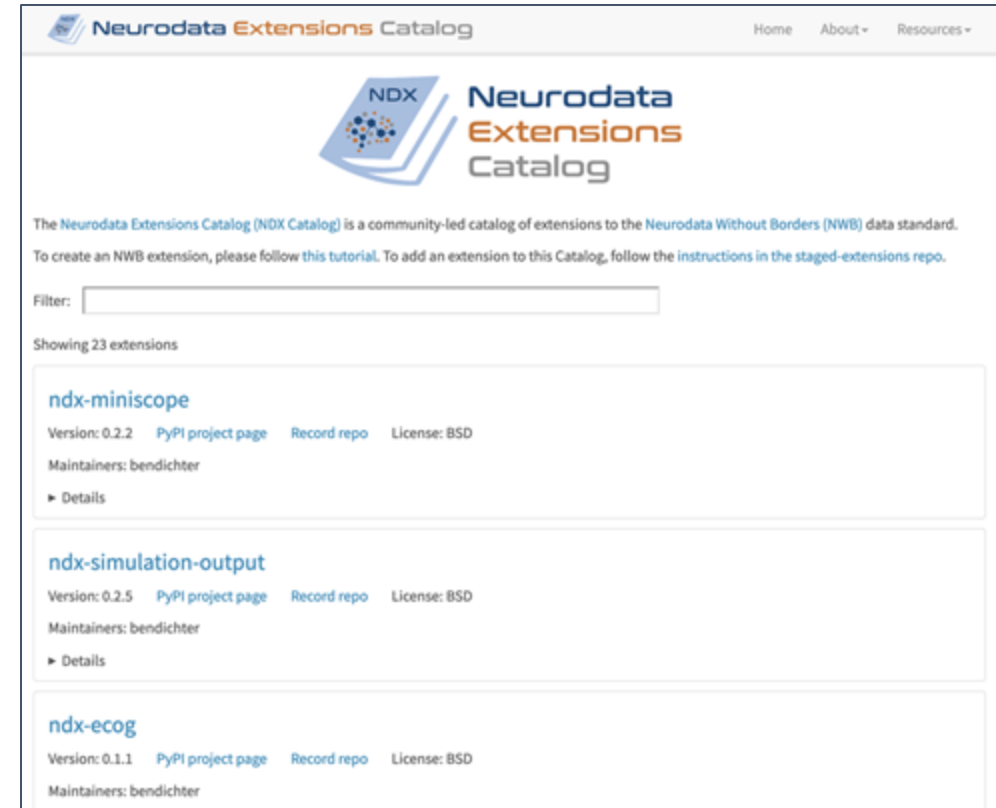


[https://nwb-overview.readthedocs.io/en/latest/tools/analysis\\_tools\\_home.html](https://nwb-overview.readthedocs.io/en/latest/tools/analysis_tools_home.html)



# NWB Extensions and Enhancement Proposal

- **Neurodata extensions (NDX)** allow the community to define new data types to accommodate new technologies and data:
  - NDX can be easily shared via NWB files and the NDX catalog at <https://nwb-extensions.github.io/>
  - 23+ extensions published in the NDX catalog
- **NWB Enhancement Proposals (NWBEF)** allow the community to propose enhancements to NWB:
  - Developed with the NWB Technical Advisory Board (TAB) policies and processes for NWBEFs
  - Created forms and formal process for managing and reviewing NWBEFs Completed review of NWBEF001 to enhance support for event-based data
  - Upcoming reviews for (**NWBEF002** Probe devices and channel mapping in extracellular ephys, **NWBEF003** Multichannel volumetric imaging)



Extension Tutorial



NWB Policies



NEP Review



# Community Call for TAB and NWBEP Contributors

- **NWB Technical Advisory Board Nominations**
  - The NWB TAB will have a community election in January to replace outgoing member Lawrence Niu. If you are interested in joining the NWB TAB, please nominate yourself for the election.
- **Volunteer to serve as a NWBEP reviewer or organize an NWB Working Group**
  - The NWB TAB plans to organize community reviews of several NWB Enhancement Proposals. If you are interested, please volunteer as a reviewer.
  - Do you have ideas for an NWB Enhancement Proposal and want to organize people to discuss them? Let us know and we can help connect you with anyone we know who has similar interests.



Nominate yourself for  
the NWB TAB



Sign up as a  
Volunteer

# Evaluating and optimizing reading of NWB data from cloud storage



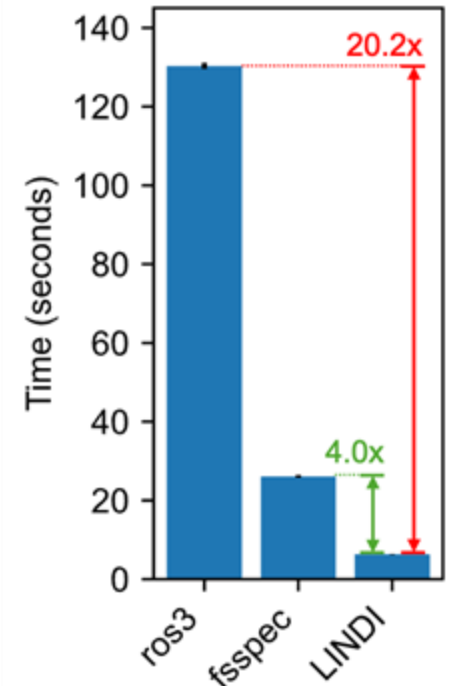
## Developed new **NWB Benchmarks** library based on ASV

- Developed benchmarks to evaluate and compare the runtime performance for reading ecephys, ophys, and icephys NWB data from DANDI using:
  - 1: ROS3 (h5py)
  - 2, 3: fsspec (w/o cache)
  - 4, 5: RemFile (w/o cache)
  - 6: LINDI
  - 7, 8: Zarr (w/o consolidated metadata)
- Developed network traffic benchmarks to profile web traffic of the different methods
- **Docs:** <https://nwb-benchmarks.readthedocs.io/>
- **Source:** [https://github.com/NeurodataWithoutBorders/nwb\\_benchmarks](https://github.com/NeurodataWithoutBorders/nwb_benchmarks)

## Enhanced **HDMF-Zarr** to support consolidated metadata to enhance cloud read

## Developed new **Linked Data Interface (LINDI)** together with J. Magland

- Index existing files using a Zarr-compatible file format (similar to kerchunk)
- Enable efficient access to remote HDF5 files via consolidated metadata
- Developed h5py compatible API to support direct integration with PyNWB
- **Source:** <https://github.com/NeurodataWithoutBorders/lindi>



Open and read slice from ecephys file

# AqNWB

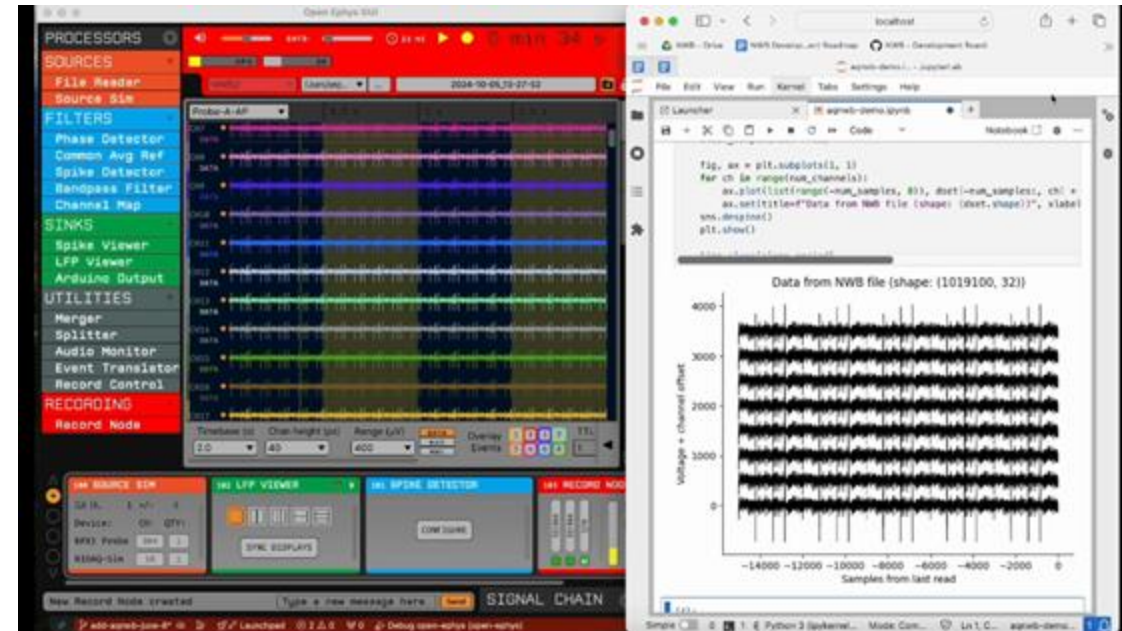
**Goal:** A lightweight C++ API to integrate with existing acquisition systems and acquire data directly into the NWB format.

## Updates

- Developed initial integration with Open Ephys with support for electrophysiological data write
- Ongoing discussions with several other acquisition systems
- First stable release planned for summer 2025

## Resources:

- **Docs:**  
<https://neurodatawithoutborders.github.io/aqnwb>
- **Source:**  
<https://github.com/NeurodataWithoutBorders/aqnwb>



**Support:** Funded by the Kavli Foundation

# HERD: Linking to Ontologies and External Resources

**Goal:** Enable users to attach and store contextual metadata to link NWB data to ontologies and external resources

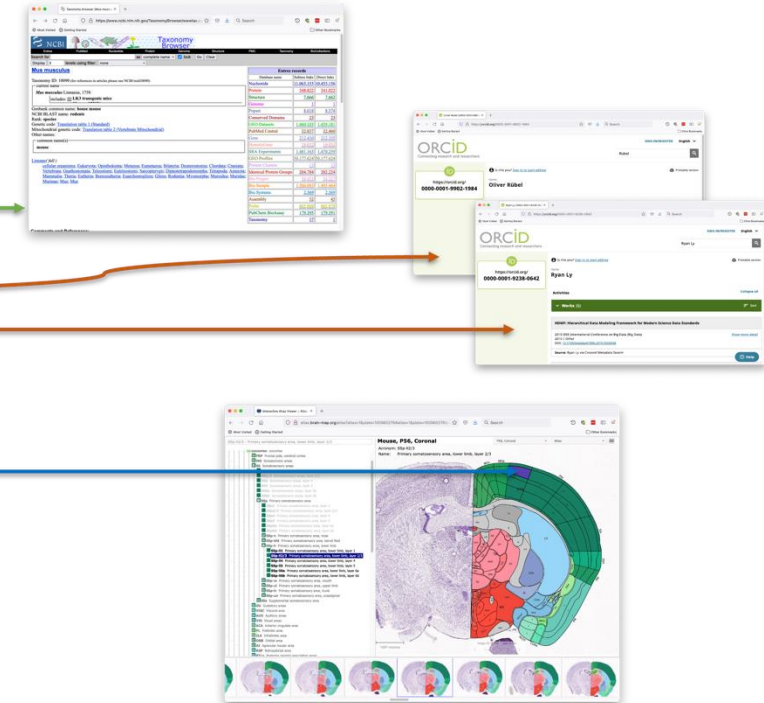
**Approach:** Developed the HDMF External Resources Data (HERD) tool in HDMF, which supports:

- **TermSet:** Define terms via LinkML schema
- **TermSetWrapper:** Wrapper for datasets and attributes to validate and link data according to a TermSet.
- **Configuration Schemas:** Automate the usage of controlled terminologies by pre-configuring which TermSets to use with an NWB file
- **HERD IO:** Automatically collect and save linkages from wrapped datasets and support saving, reading, and editing of data annotations.

NEURODATA  
WITHOUT BORDERS

```
L general
  L subject
    neurodata_type: Subject
  L species : Mouse
  L experimenter :
    ● Oliver Rübel
    ● Ryan Ly
  L extracellular_ephys
    L electrodes :
      x   y   z   imp  location  ...
      .. .. .. ..  SSp-II1  ...
      .. .. .. ..  SSp-II1  ...
      .. .. .. ..  SSp-II2/3 ...
  L acquisition
    L neuropixel_recordings
      neurodata_type: ElectricalSeries
  L ...
```

x	y	z	imp	location	...
..	..	..	..	SSp-II1	...
..	..	..	..	SSp-II1	...
..	..	..	..	SSp-II2/3	...



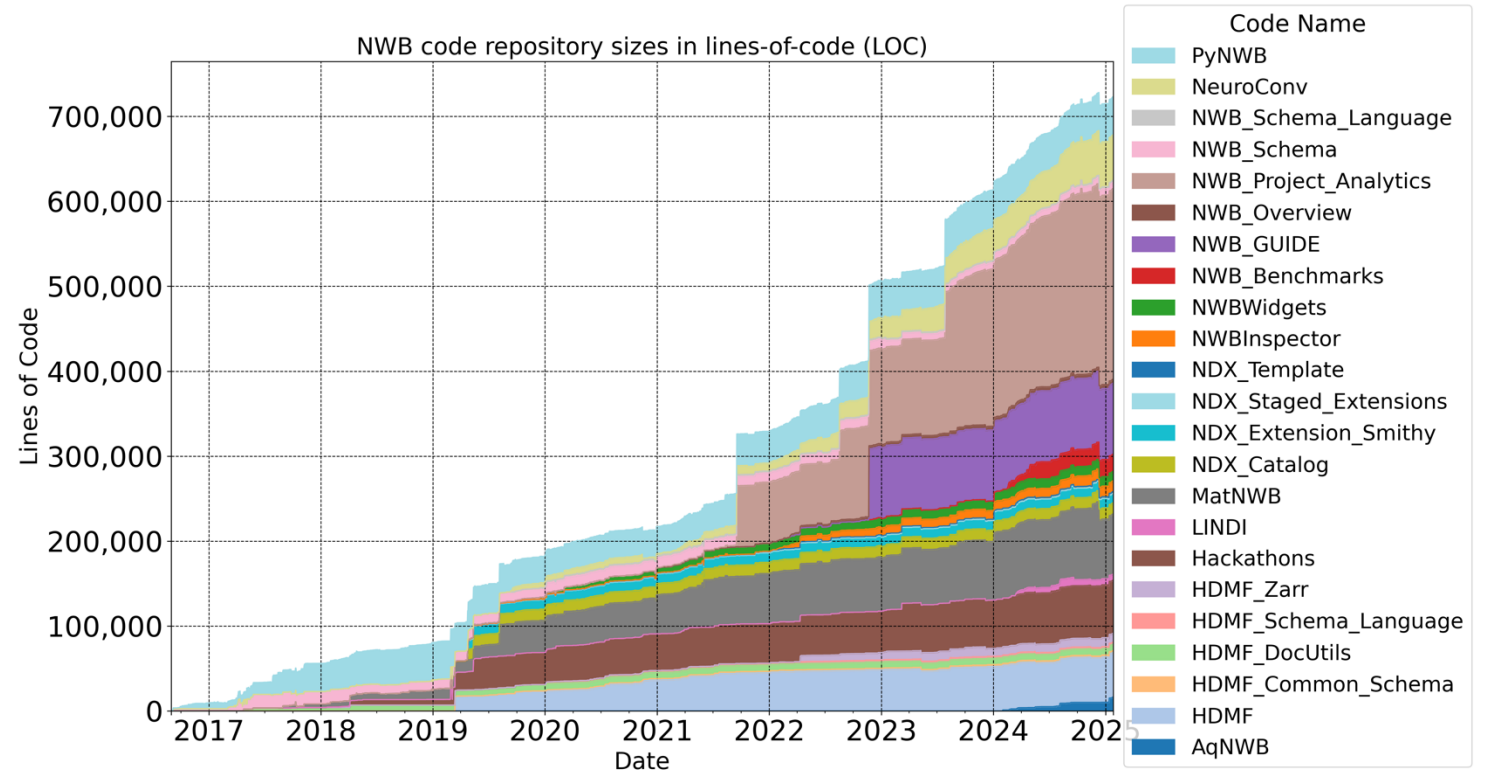
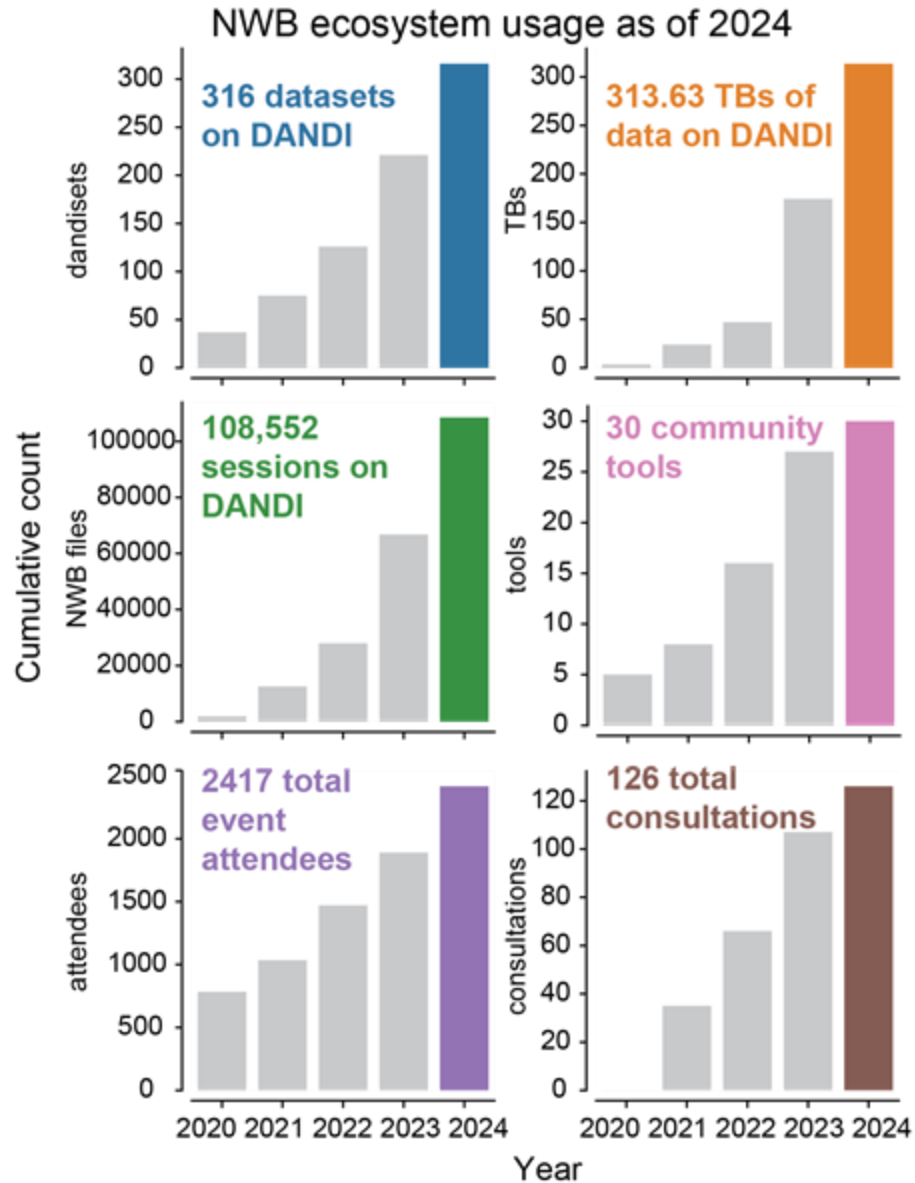
M.Avaylon, R.Ly, A.Tritt, B.Dichter, K.E. Bouchard, C.J. Mungall, O.Ruebel, “**Methods for Linking Data to Online Resources and Ontologies with Applications to Neurophysiology,**” arXiv (May 2024), DOI: <https://doi.org/10.48550/arXiv.2406.00063>



# Summary

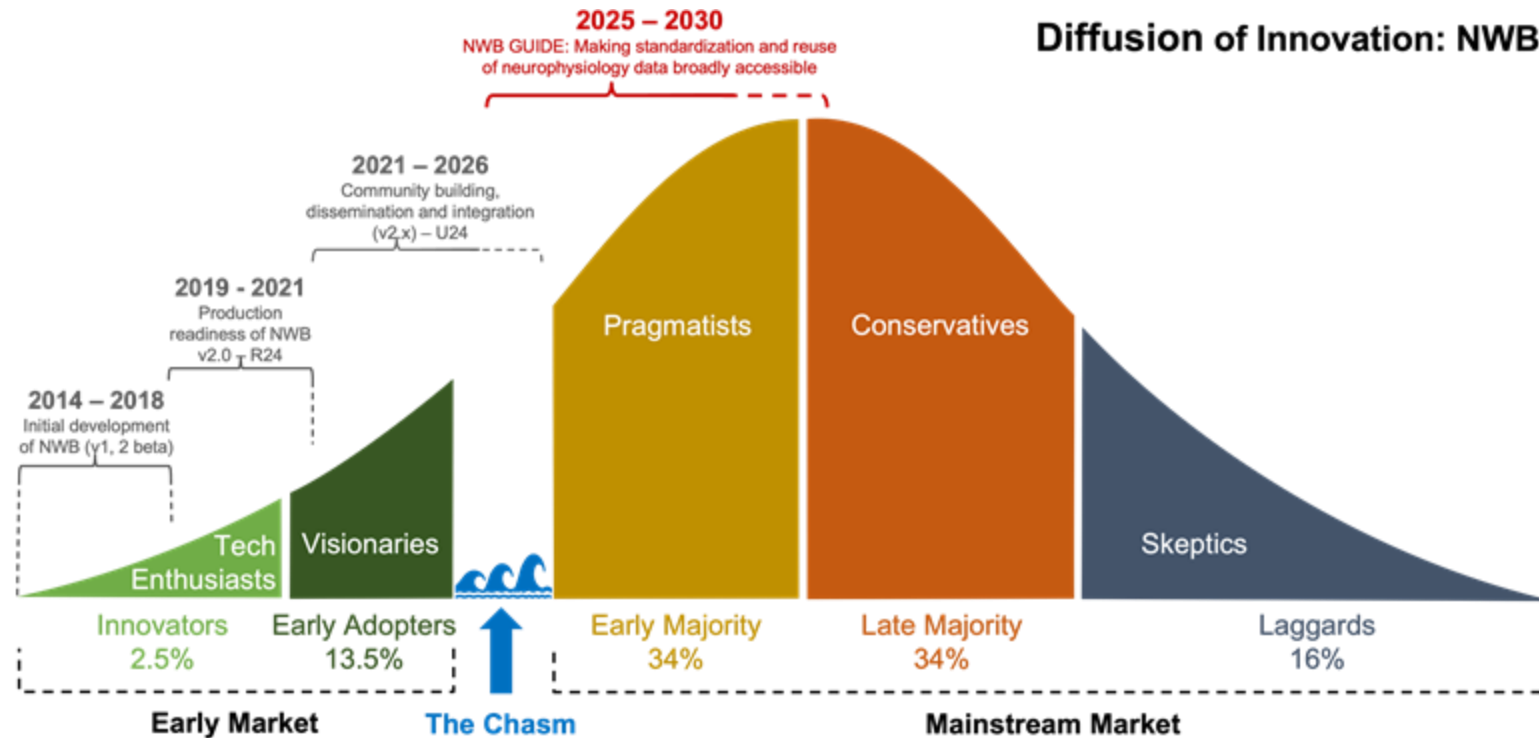
How has NWB grown since the first release of NWB 2.0 in Jan. 2019?

# Current state of NWB





# NWB is growing!



- Our current users are “*Tech Enthusiasts*” and “*Visionaries*” who are willing to put in extra work to utilize state-of-the-art technology
- We need to be able to reach “*Pragmatists*” who generally follow the path that is most convenient and cost-effective and require end-to-end solutions
- For NWB to be successful we need to be able to scale with the community, adapt to growing and changing community needs, and continue to ensure readiness of NWB for production use





**NEURODATA**  
WITHOUT BORDERS

Visit us at [NWB.org](https://nwb.org) and  
<https://nwb-overview.readthedocs.io>



**BERKELEY LAB**  
Bringing Science Solutions to the World



- **Disclaimer:** This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.
- **Copyright notice:** All rights reserved. This manuscript has been authored by an author at Lawrence Berkeley National Laboratory under Contract No. DE-AC02-05CH11231 with the U.S. Department of Energy. The U.S. Government retains, and the publisher, by accepting the article for publication, acknowledges, that the U.S. Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for U.S. Government purposes.