



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



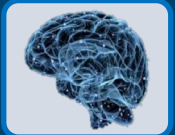
BERKELEY LAB
Bringing Science Solutions to the World

Overview

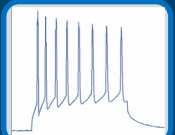
- Introduction to NWB
- Tools for Data Conversion & Inspection
- Select Current Developments

Multidisciplinary team science at work

Applications



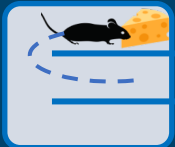
Extracellular
electrophysiology



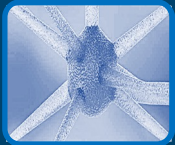
Intracellular
electrophysiology



Optical physiology



Behavior



Simulations

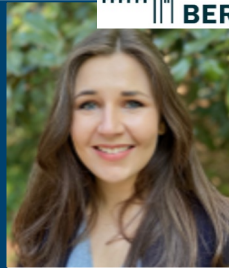
Technology Teams



BERKELEY LAB



O. Rübél



S. Prince



R. Ly



M. Avaylon



C. Baker



A. Weigl



L. Niu



S. De Vries



A. Buccino



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)



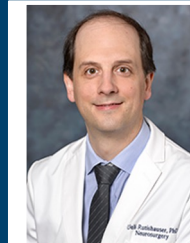
S. Ghosh
(MIT)



A. Kepecs
(WUSTL)



M. Murthy
(Princeton)



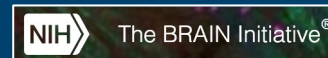
U. Rutishauser
(Cedars Sinai)

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

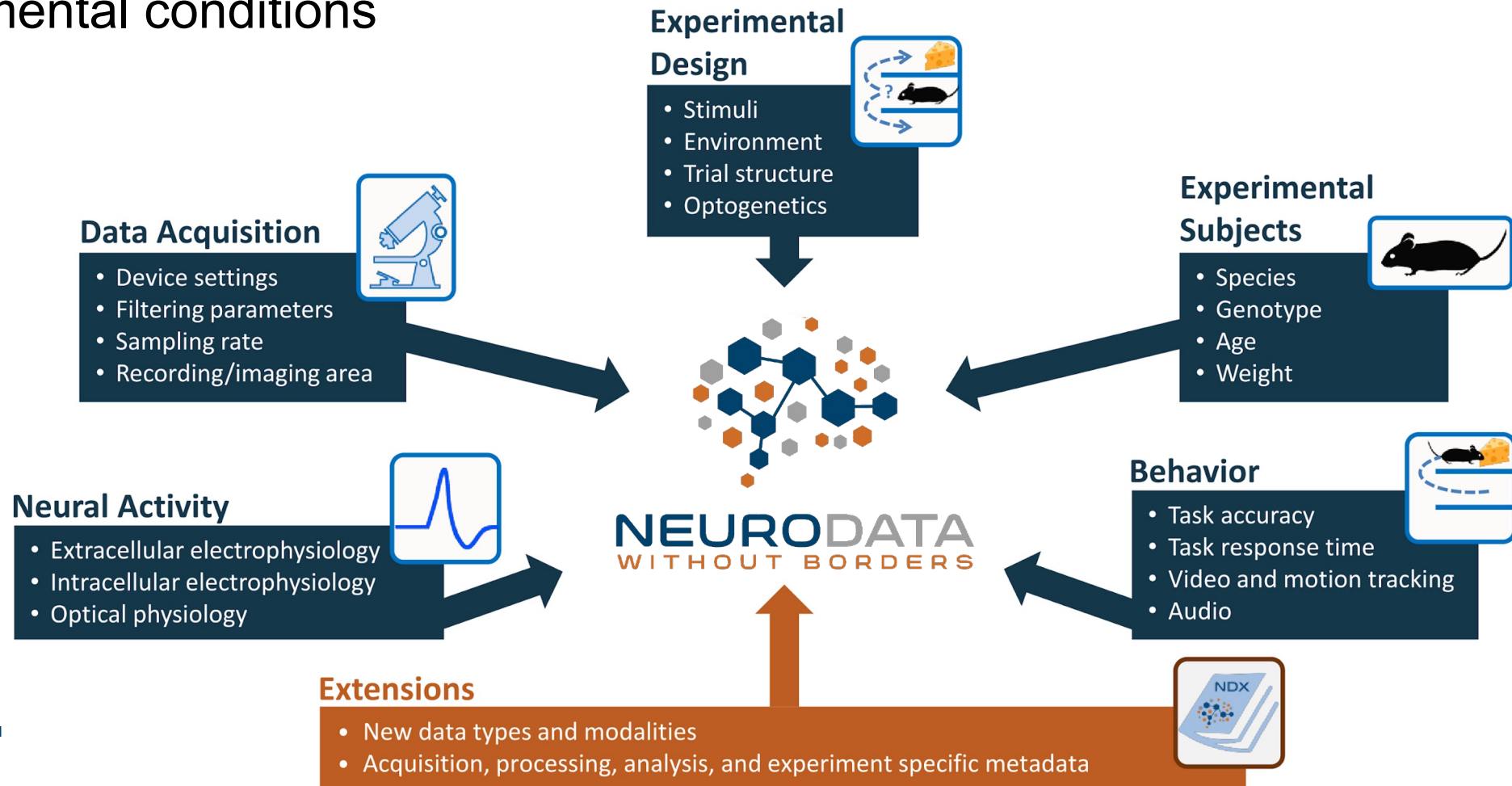
- CatalystNeuro
- DataJoint
- MBF Bioscience
- MathWorks
- Kitware

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!

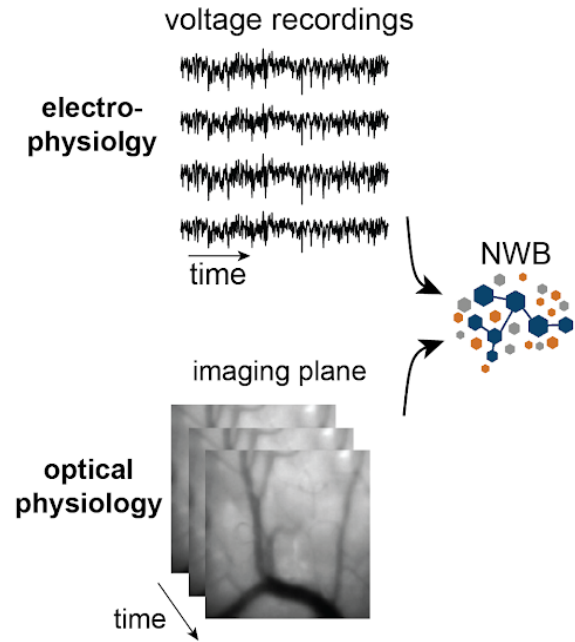
A unified data standard for neurophysiology

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

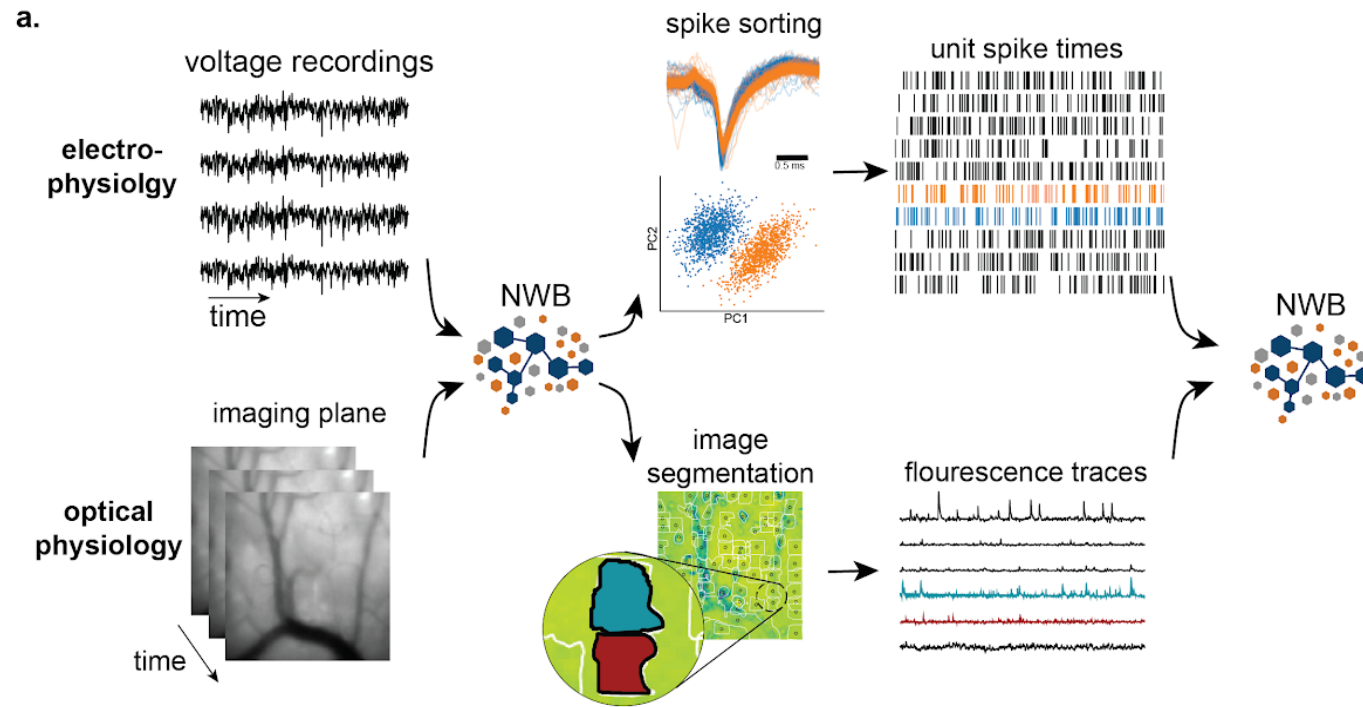


NWB enables unified storage of multimodal raw and processed data

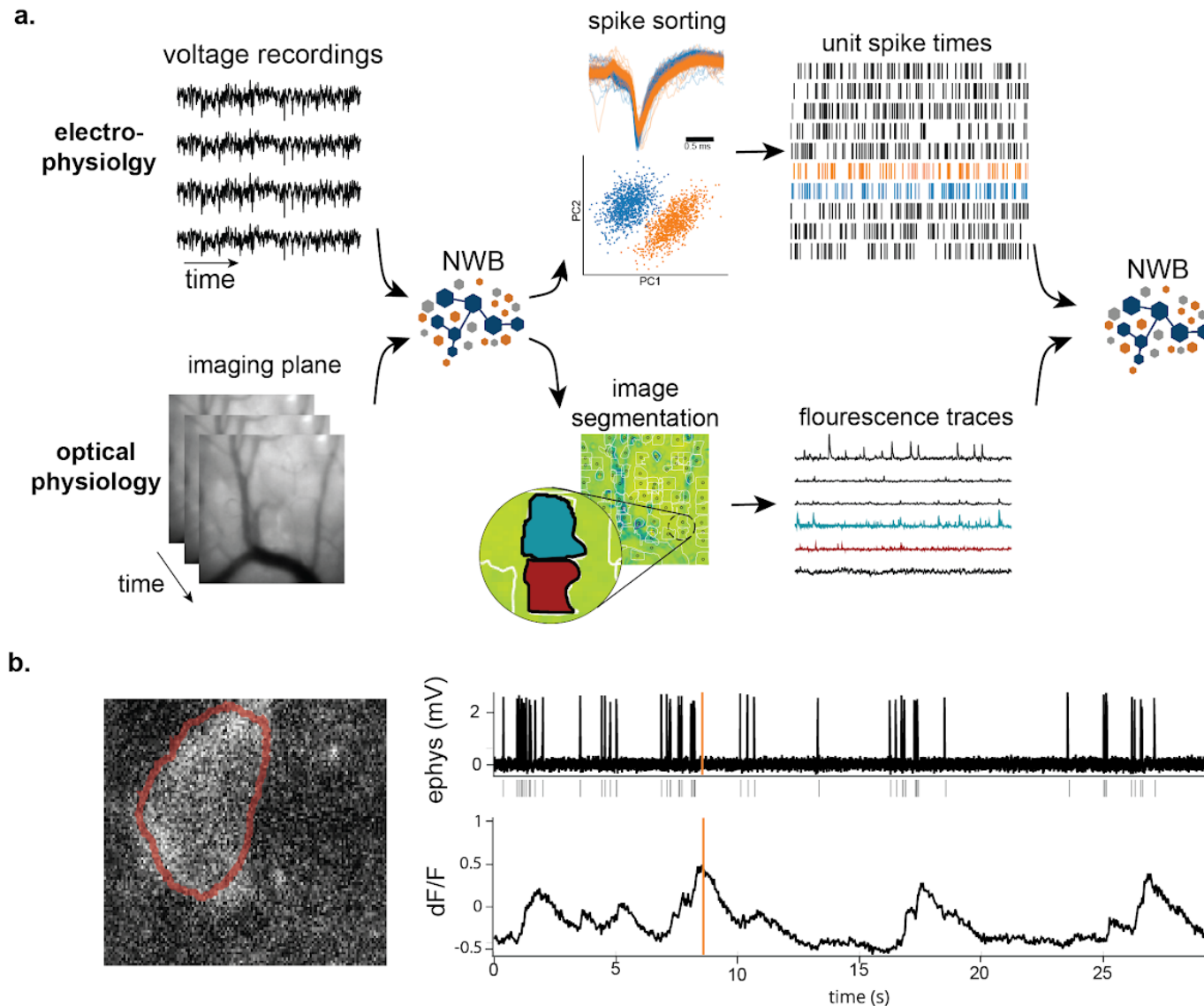
a.



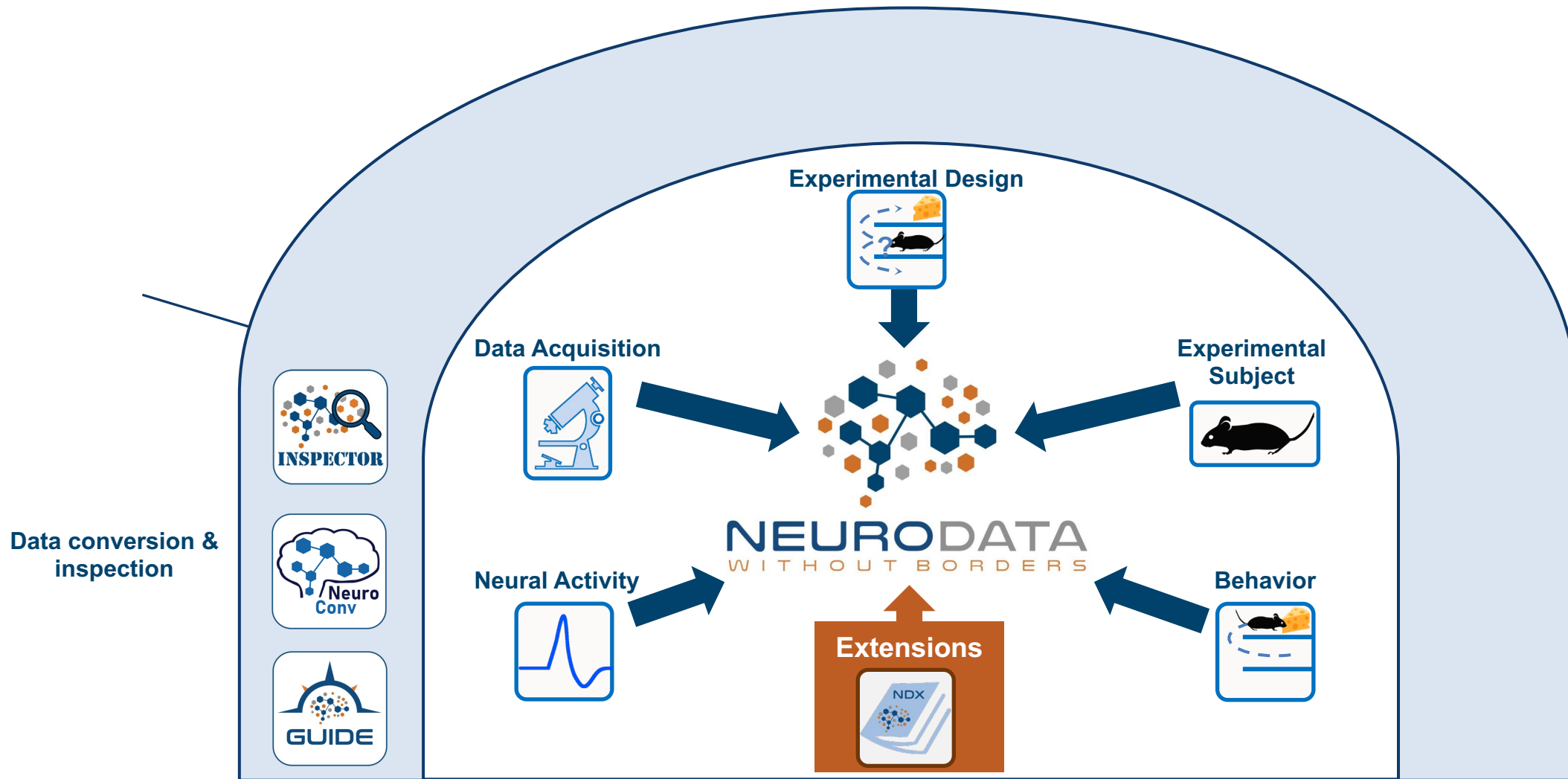
NWB enables unified storage of multimodal raw and processed data



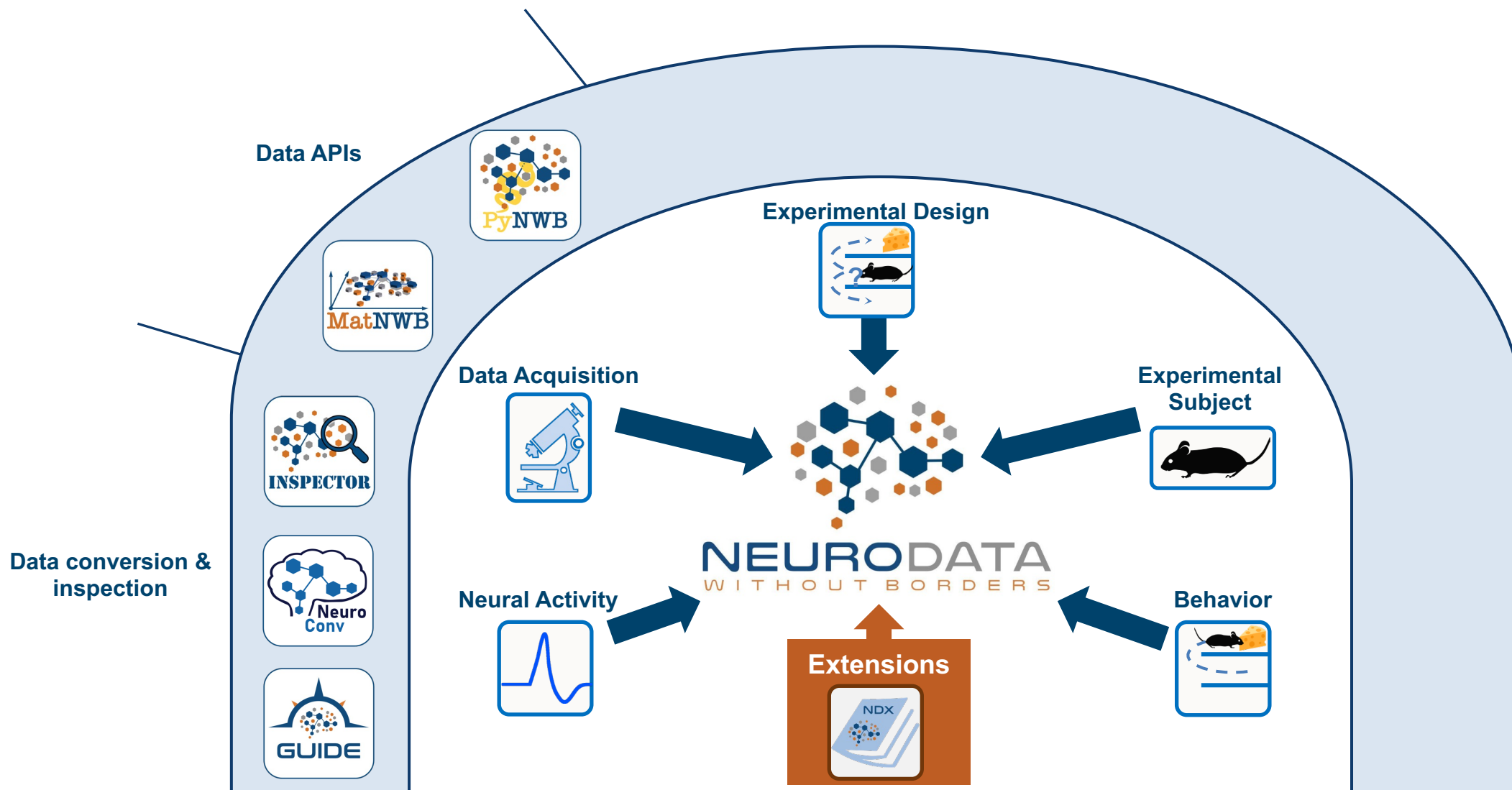
NWB enables unified storage of multimodal raw and processed data



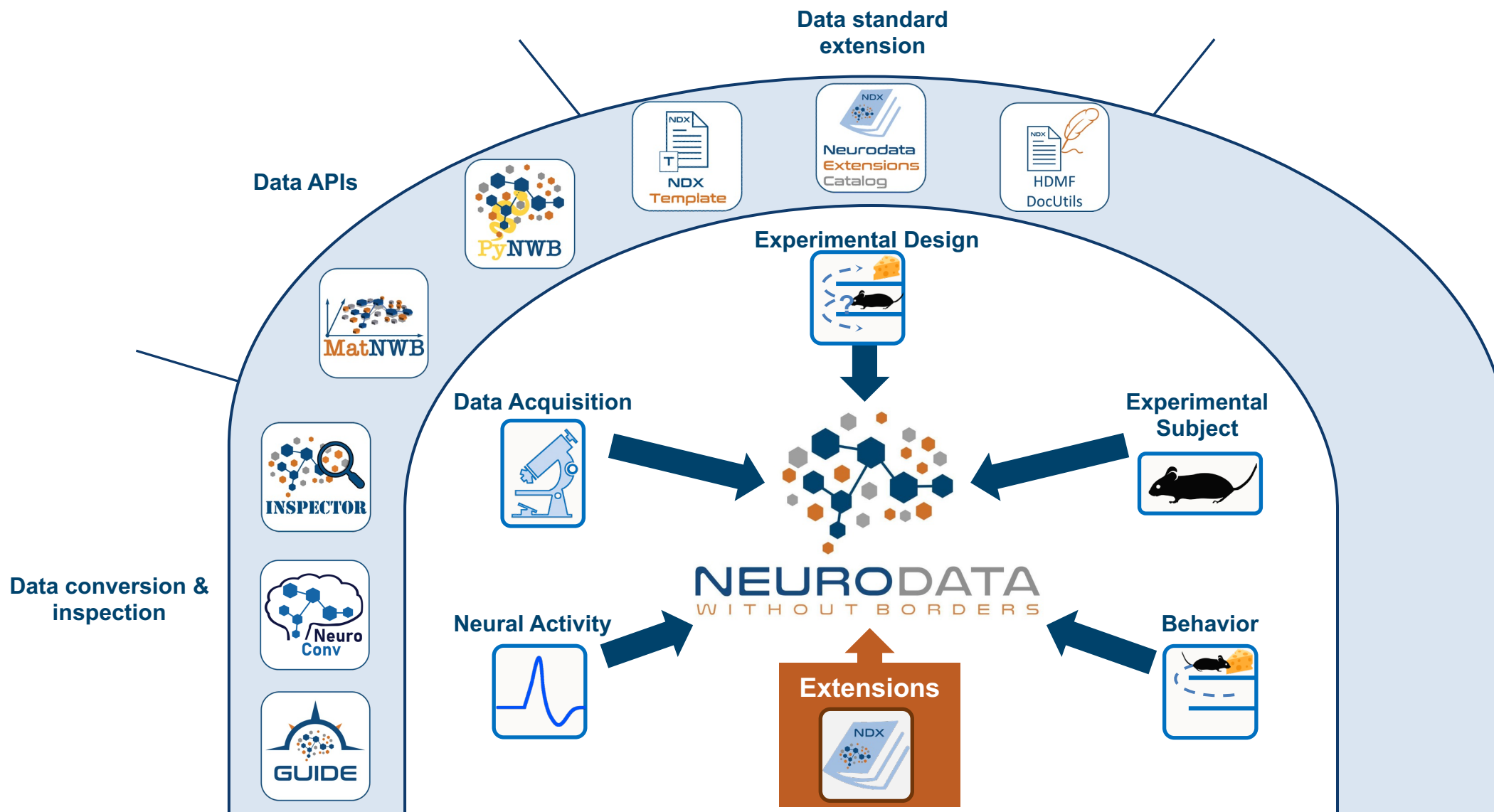
An ecosystem for neuroscience data standardization



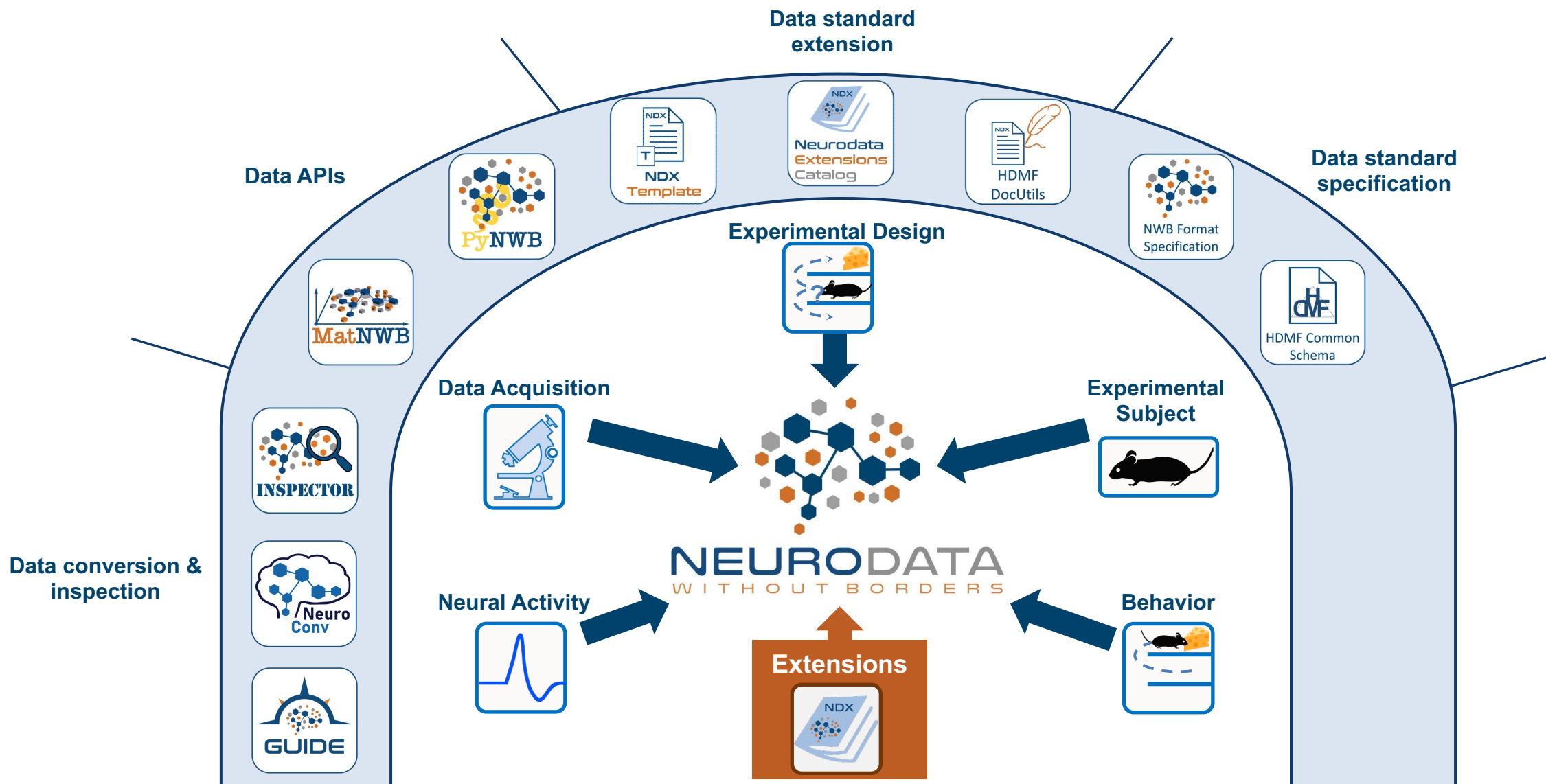
An ecosystem for neuroscience data standardization



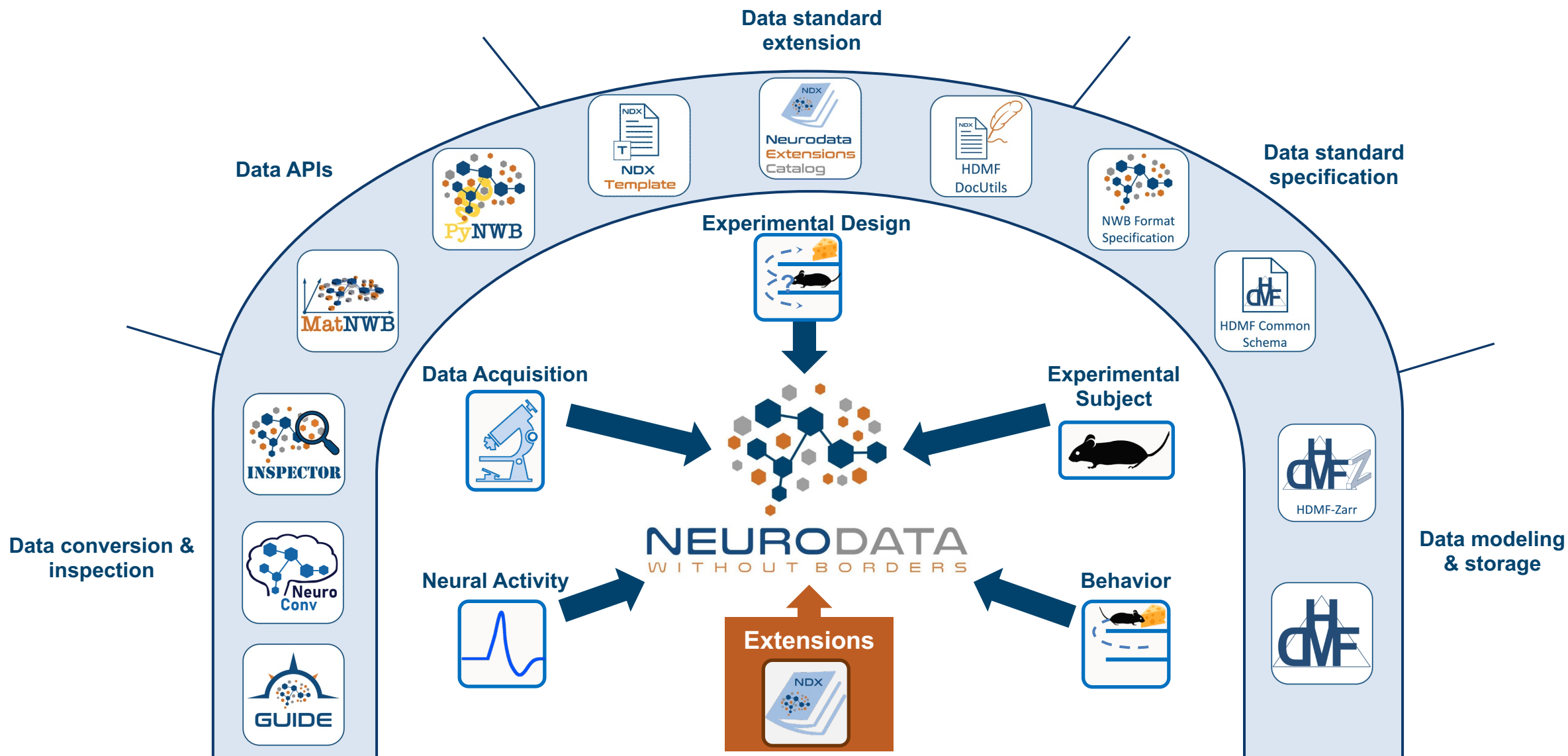
An ecosystem for neuroscience data standardization



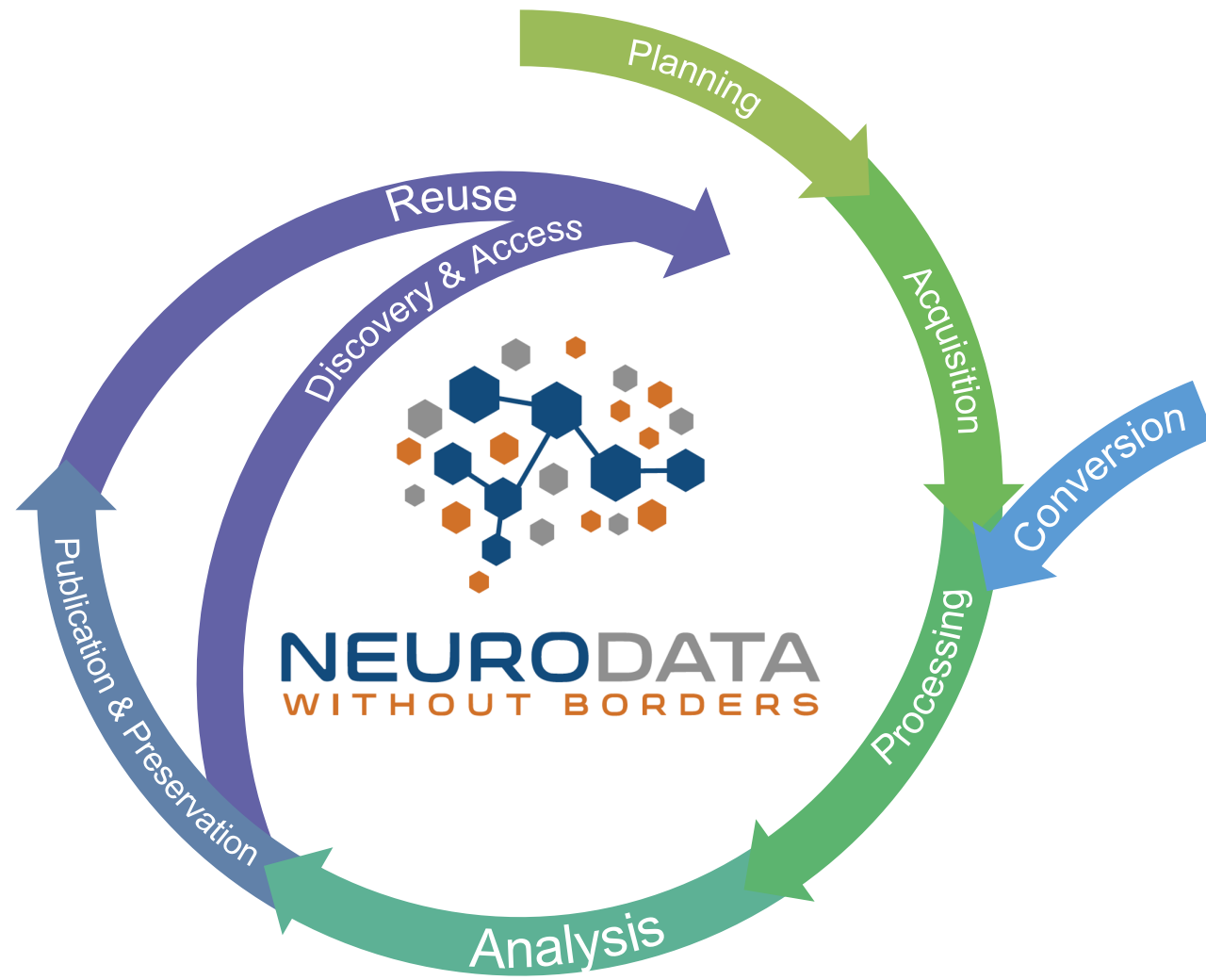
An ecosystem for neuroscience data standardization



An ecosystem for neuroscience data standardization

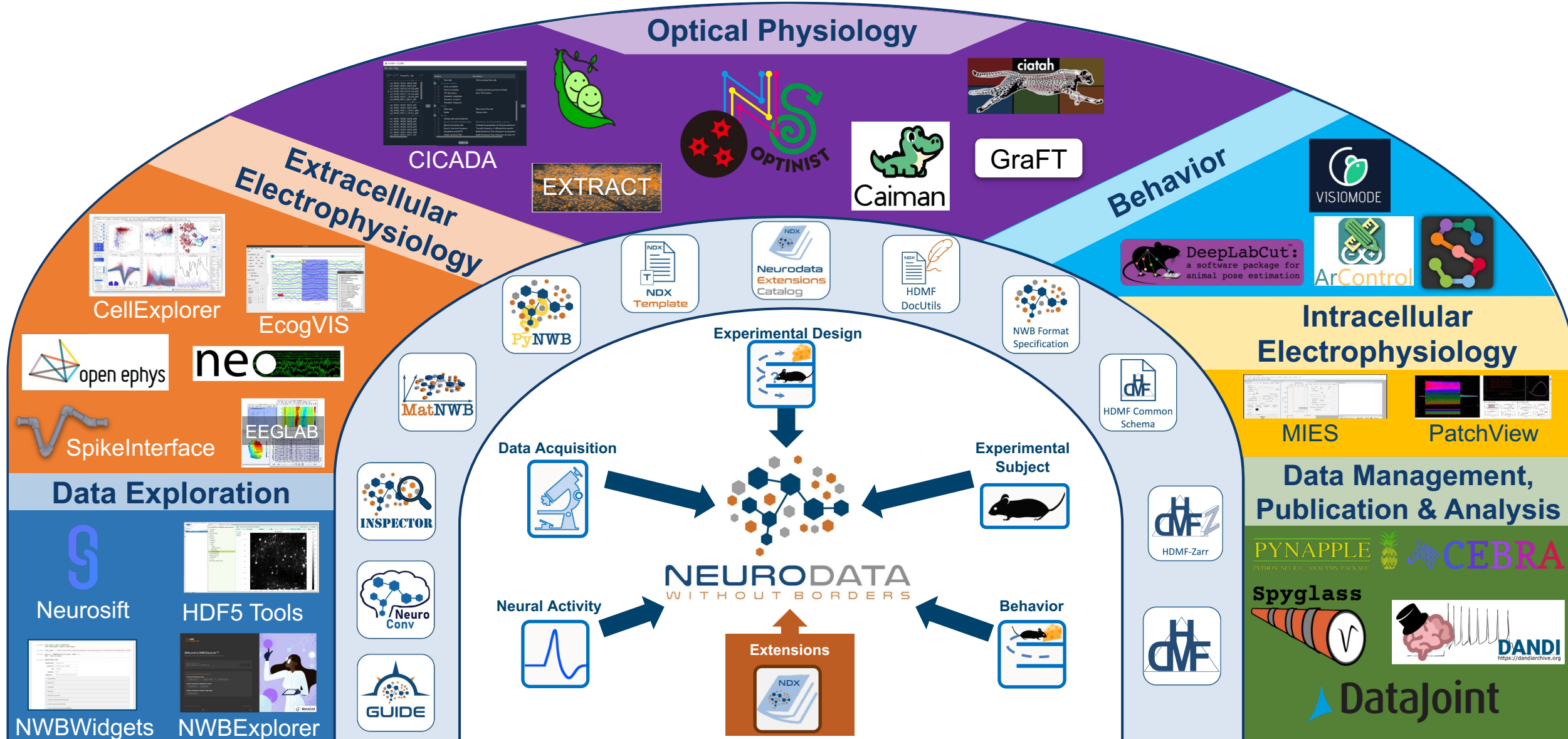


NWB technologies at the heart of the neurodata lifecycle



- 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

A unified data standard and software ecosystem for neurophysiology



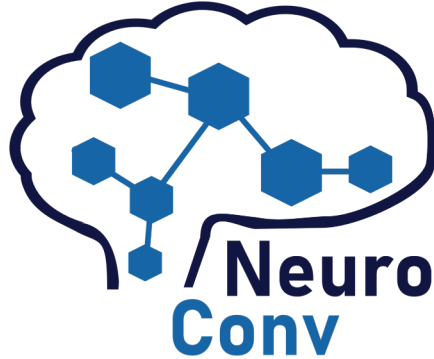
Tools for Data Conversion & Inspection

Enable researchers to effectively integrate their data with the NWB data standard

Converting data to NWB



Custom code
using NWB APIs



Low-code
automated
conversion
for common input
data formats



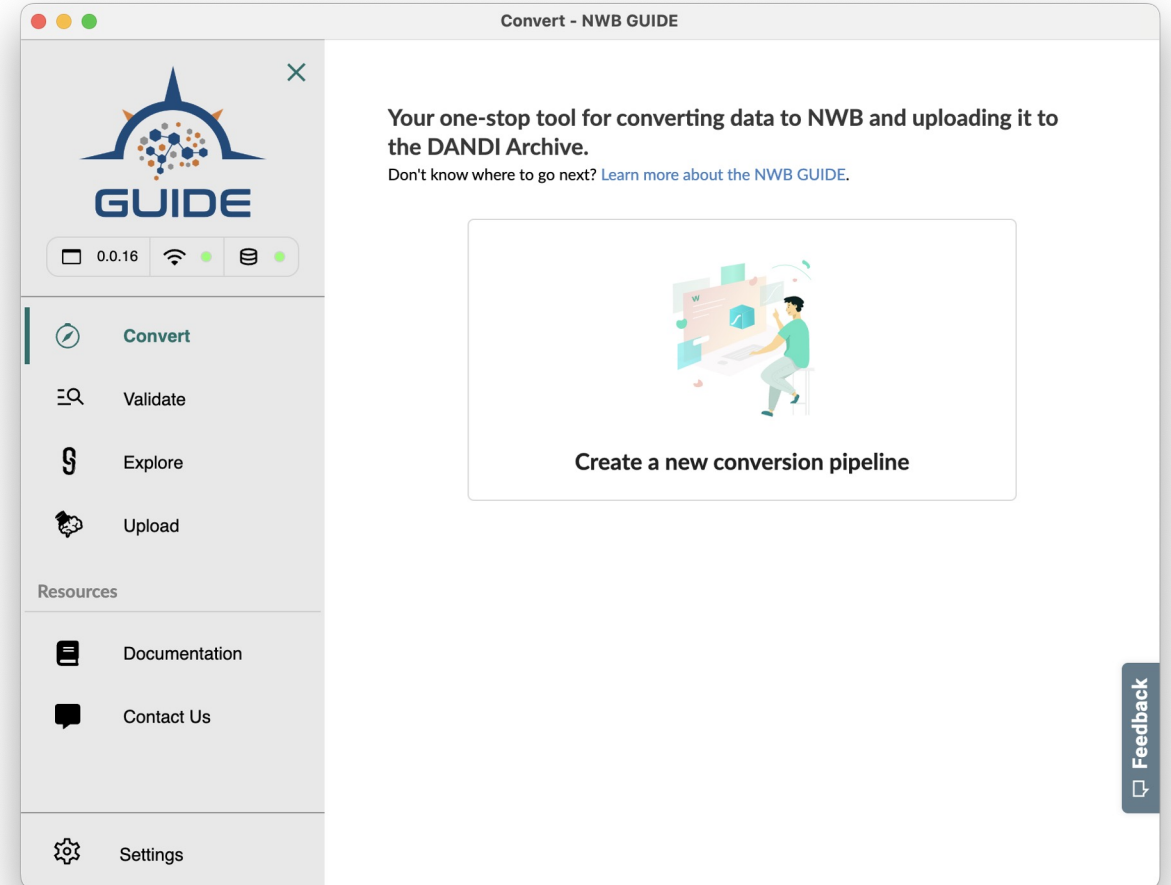
No-code automated
conversion via graphical
interface app
(for select input data
formats, in beta)

Flexible

Automated

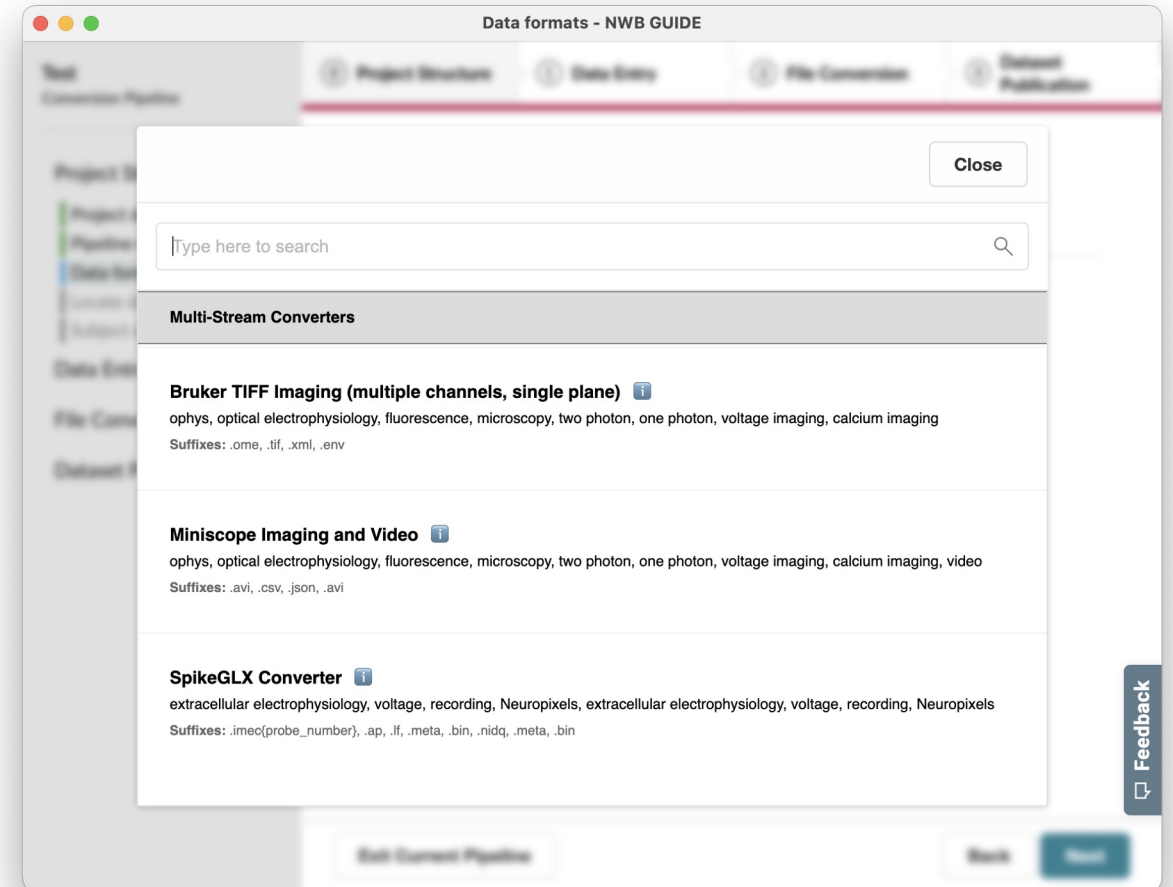
NWB GUIDE

- TurboTax-like interface for converting data to NWB



NWB GUIDE

- TurboTax-like interface for converting data to NWB
- **Select from 40+ data formats (backed by NeuroConv)**



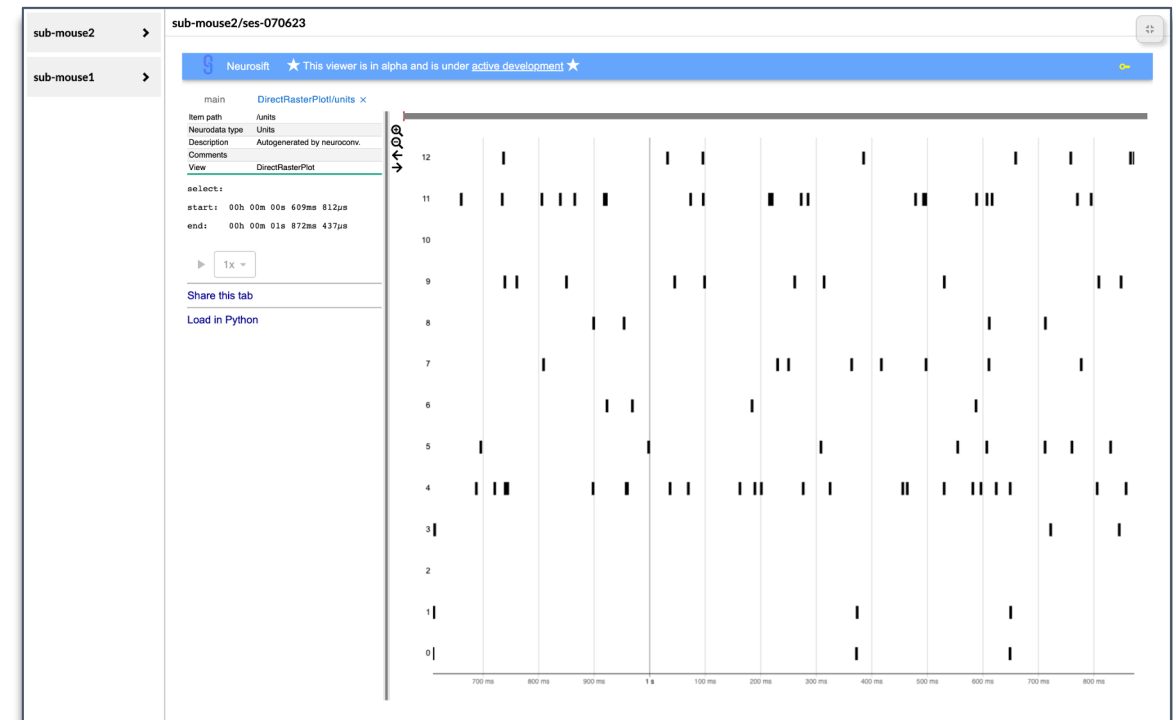
NWB GUIDE

- TurboTax-like interface for converting data to NWB
- Select from 40+ data formats (backed by NeuroConv)
- **Conveniently enter metadata with immediate validation and on-demand docs**

The screenshot shows the 'File metadata - NWB GUIDE' application window. The interface is divided into a left sidebar and a main content area. The sidebar, titled 'Test Conversion Pipeline', lists four steps: 'Project Structure', 'Data Entry', 'Source data', 'File metadata' (which is highlighted with a blue bar), 'File Conversion', and 'Dataset Publication'. The main content area has a top navigation bar with four steps: '0 Project Structure', '1 Data Entry' (which is active), '2 File Conversion', and '3 Dataset Publication'. Below the navigation bar, the 'File Metadata' section is titled 'Edit all metadata for this conversion at the session level'. The main form area is for 'sub-test/ses-test' and includes a 'Preview' button and a 'Save' button. The form contains several fields: 'General Metadata' (a dropdown menu), 'Subject' (a dropdown menu), 'Sex *' (a text input field with 'Female' entered), 'Species *' (a text input field with 'Mus musculus - House mouse' entered), and 'Subject Age' (a text input field). A 'Feedback' button is located in the bottom right corner. At the bottom of the window, there are three buttons: 'Exit Current Pipeline', 'Back', and 'Next'.

NWB GUIDE

- TurboTax-like interface for converting data to NWB
- Select from 30+ data formats (backed by NeuroConv)
- Conveniently enter metadata with immediate validation and on-demand docs
- **Visualize and inspect sample converted data**
- **Batch convert all data**



NWB GUIDE

- TurboTax-like interface for converting data to NWB
- Select from 30+ data formats (backed by NeuroConv)
- Conveniently enter metadata with immediate validation and on-demand docs
- Visualize and inspect sample converted data
- Batch convert all data
- **Create a Dandiset and upload NWB data to DANDI**

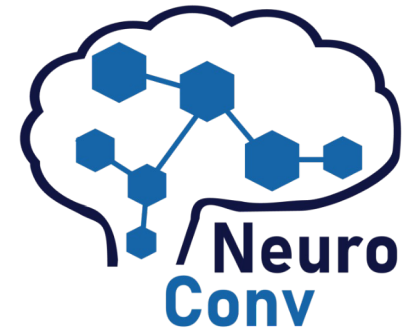


1.0 Release - <https://nwb-guide.readthedocs.io/>
Tutorial: <https://www.youtube.com/watch?v=EhhdDVuHAZ0>

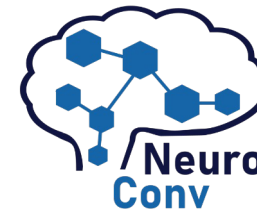
A screenshot of the 'Create a Dandiset' form in the NWB GUIDE application. The form is titled 'Create a Dandiset' and has a 'Close' button in the top right corner. It contains four main sections: 'Title *' with a text input field containing 'Neuropixels Recordings from Mouse CA3'; 'Archive *' with a dropdown menu showing 'Main Archive'; 'Description *' with a text input field containing 'Data used in Student et al. (2024). Neuropixels Recordings from Mouse CA3.'; and 'License *' with a dropdown menu. Below the 'License *' section is a 'Create' button. The background shows the 'Upload to DANDI - NWB GUIDE' window with tabs for 'Project Structure', 'Data Entry', 'File Conversion', and 'Dataset Publication'. A 'Feedback' button is visible on the right side of the window.

Automated NWB conversions with NeuroConv

- Convert data to NWB with minimal time and effort
- Supports 44+ common neurophys data formats
- Automatically extracts format-specific metadata
- Seamlessly integrated data engineering
 - Converts on the TB scale
 - Can reduce file size by ~35%
 - Optimizes chunking for streaming from DANDI



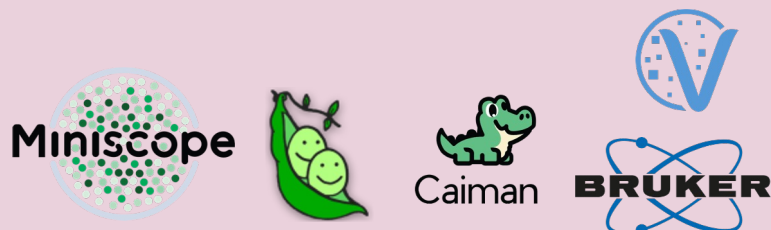
NeuroConv support 44+ common input formats



Electrophysiology



Optical physiology



Behavior



```
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)
```

A few lines of code...



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



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>

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



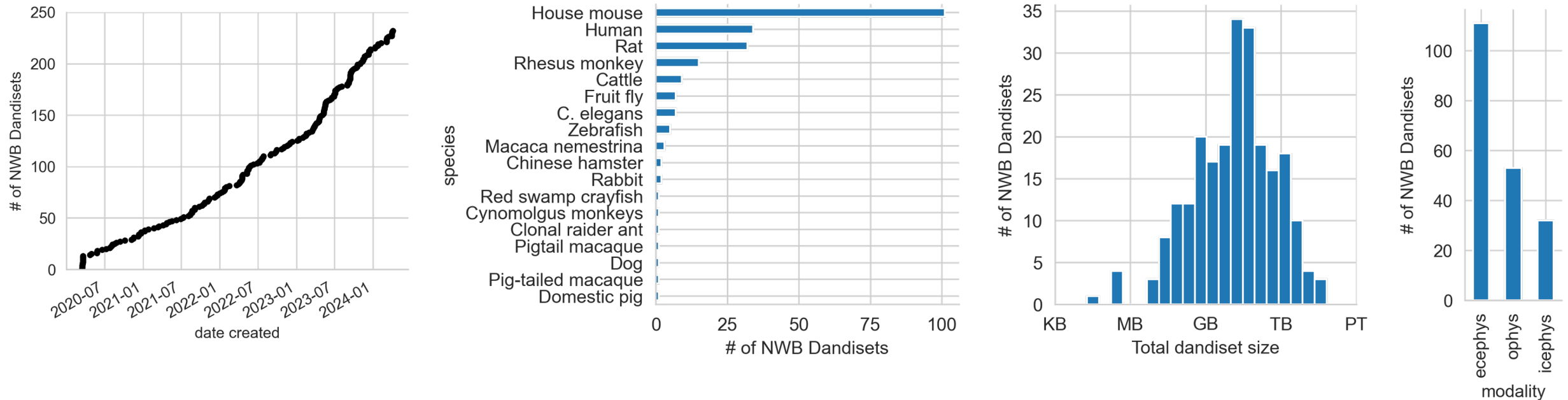
<https://dandiarchive.org>

642
[dandisets](#)

1690
users

816 TB
total data size

Public neurophysiology datasets on DANDI are growing!



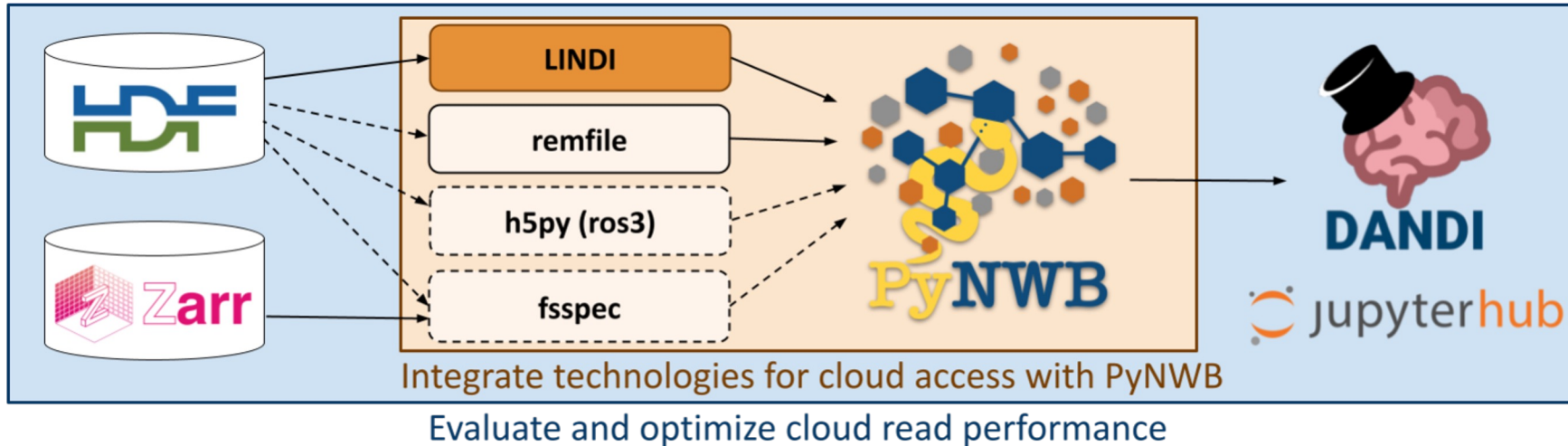
- **Growing availability of public data** including from large, multi-disciplinary efforts
 - **IBL - Brain Wide Map** ([000409](#)): 33.6 TB dataset from the International Brain Laboratory including Neuropixel electrophysiology, eye tracking, and behavioral monitoring for a visual decision making task
 - **MICrONS Two Photon Functional Imaging** ([000402](#)): 1.3 TB of optical physiology data for a visual task. The functional data was co-registered with electron microscopy (EM) data.
 - **Allen Institute OpenScope** ([000036](#), [000037](#), [000488](#), [000535](#)): Several studies from the Allen Institute Open Scope program that use optical physiology recordings of the visual cortex.

Select Current Developments

- Storage & Cloud
- Extensions
- Acquisition API
- HERD

Evaluating and Optimizing NWB for the Cloud

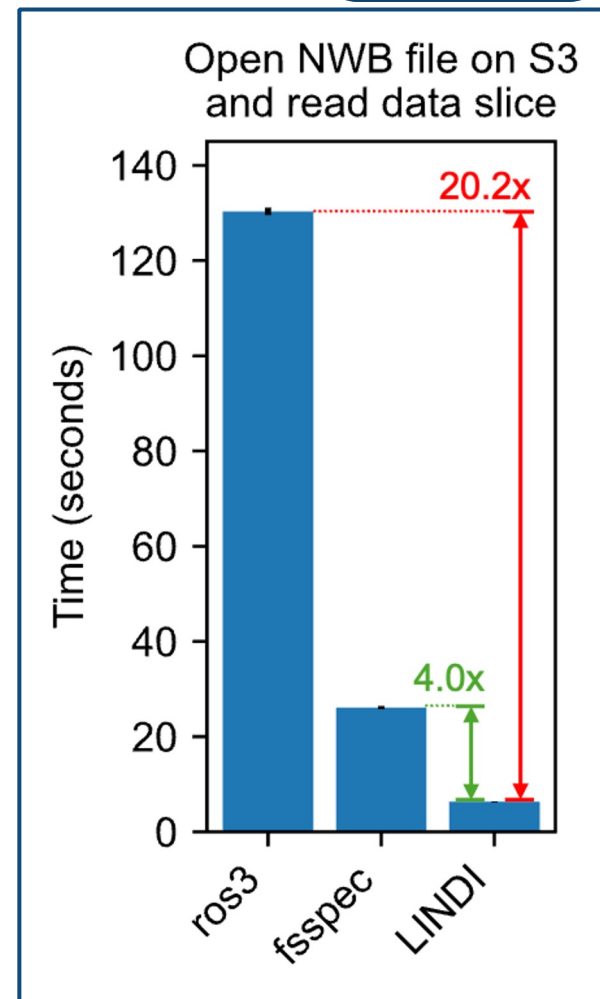
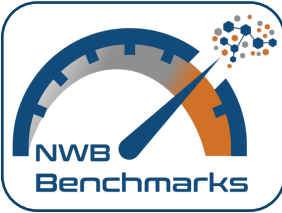
- **Goal:** Enable NWB users to harness the immense data storage and parallel processing capabilities offered by cloud infrastructure
- **Approach:**



HDMF Zarr: <https://hdmf-zarr.readthedocs.io>

NWB Benchmarks: <https://nwb-benchmarks.readthedocs.io>

LINDI: <https://github.com/NeurodataWithoutBorders/lindi>



Support the integration of new data types and metadata

- NWB supports the **formal extension** of the data standard to define new data types and metadata
- Extensions are immediately usable and shared directly with the data and can be published via the NDX Catalog
- Developed with the NWB Technical Advisory Board policies for formal NWB Enhancement Proposals (NEP)
- Examples:
 - ndx-events (NWBEP-001)
 - ndx-extracellular-channels (NWBEP-002)
 - ndx-multichannel-volume (NWBEP-003)
 - ndx-structured-behavior
 - ndx-pose
 - ...



Extension
Tutorial



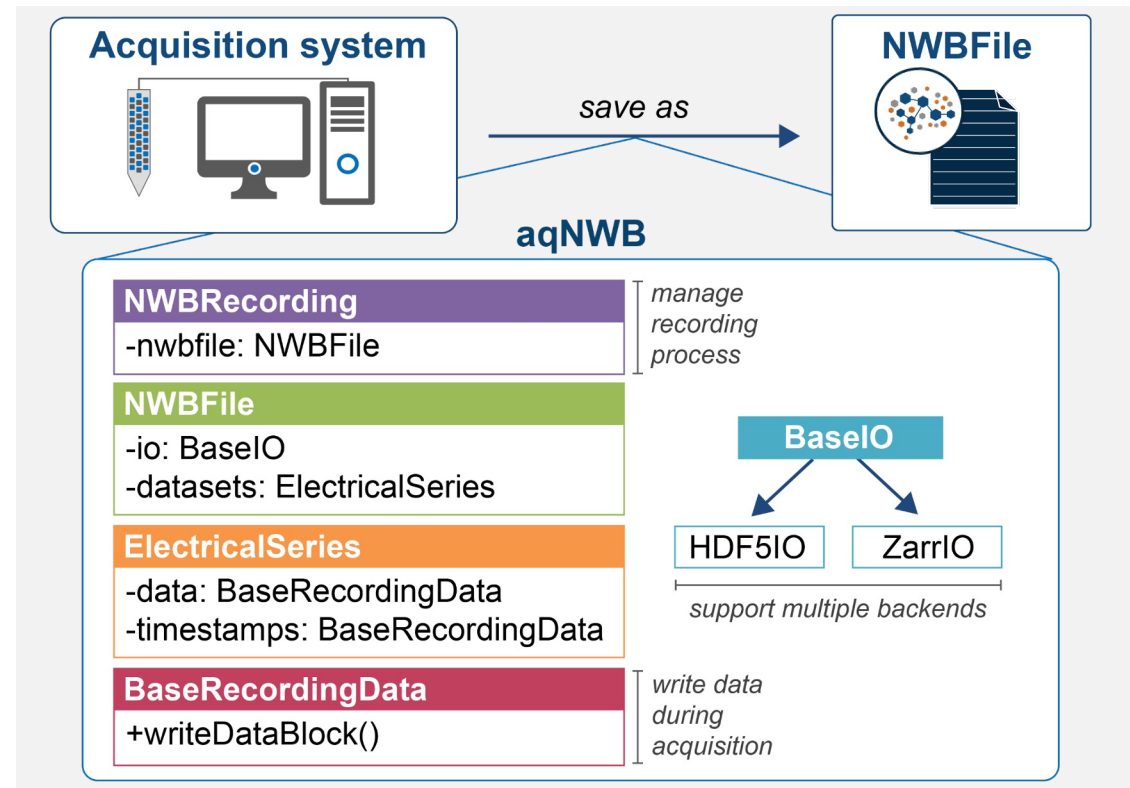
NWB Policies



NEP Review

Enabling Acquisition of Neurophysiology Data in NWB

- **aqNWB** is a C++ acquisition API currently in development that will provide a lightweight API to integrate NWB with existing acquisition systems to allow users to record their data directly into the NWB format.

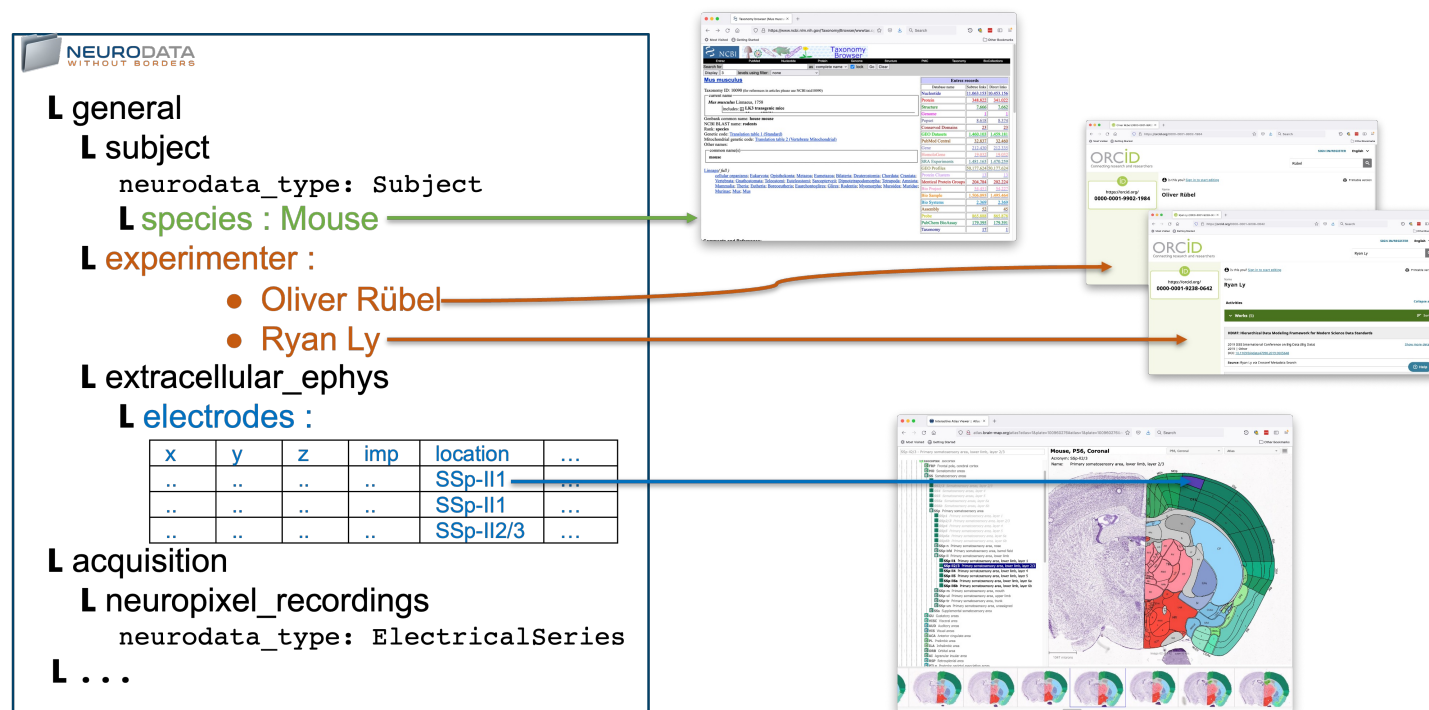


HERD: Linking to Ontologies and External Resources

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.

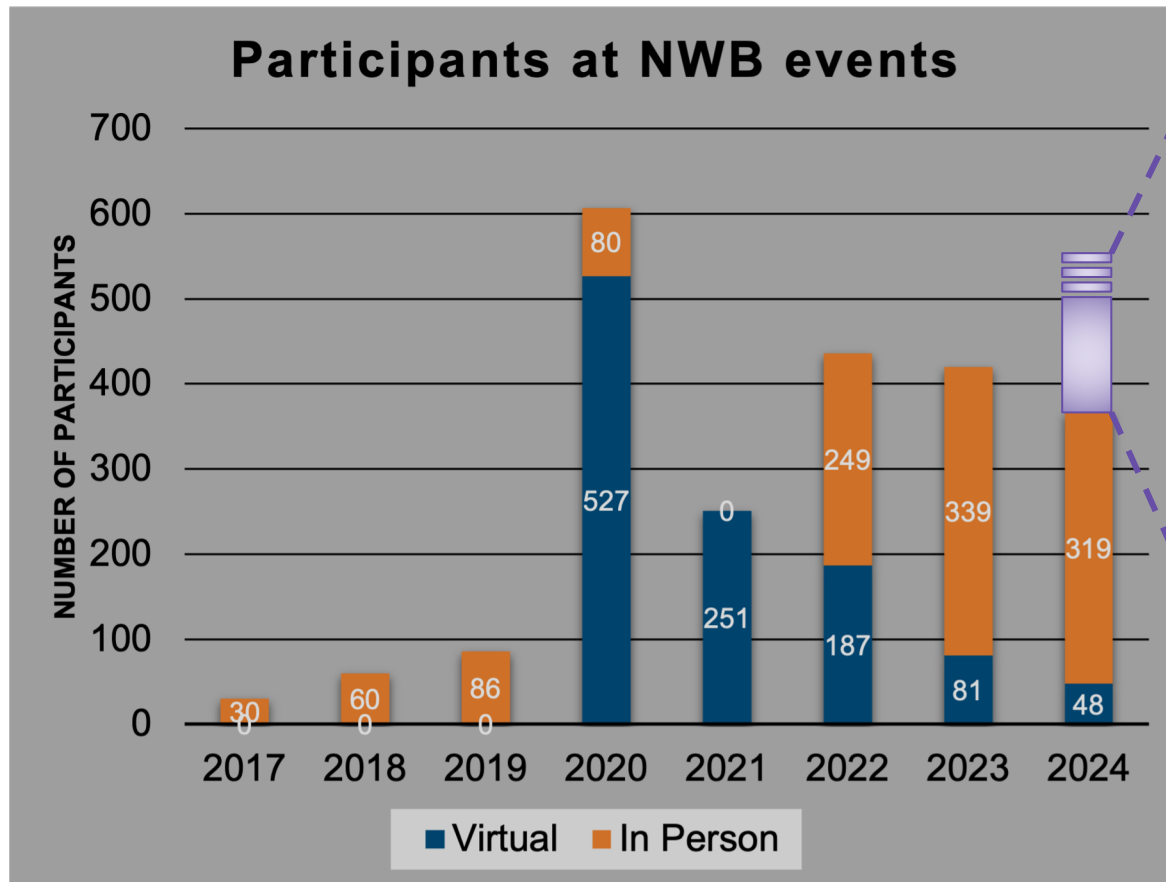


M.Avaylon, R.Ly, A.Tritt, B.Dichter, K.E. Bouchard, C.J. Mungall, O.Ruebel, “**Methods for Linking Data to Online Resources and Ontologies with Applications to Neurophysiology**,” arXiv (May 2024), DOI: <https://doi.org/10.48550/arXiv.2406.00063>



NWB Community Events

- **Goal:** Enable neuroscience researchers and developers to effectively utilize the NWB data standard and software, lower the barriers of adopting, and promote broad dissemination



- June 20-21: NWB Roadshow at JHU
- July 2: Open Neurodata Showcase (Virtual) ★
- 

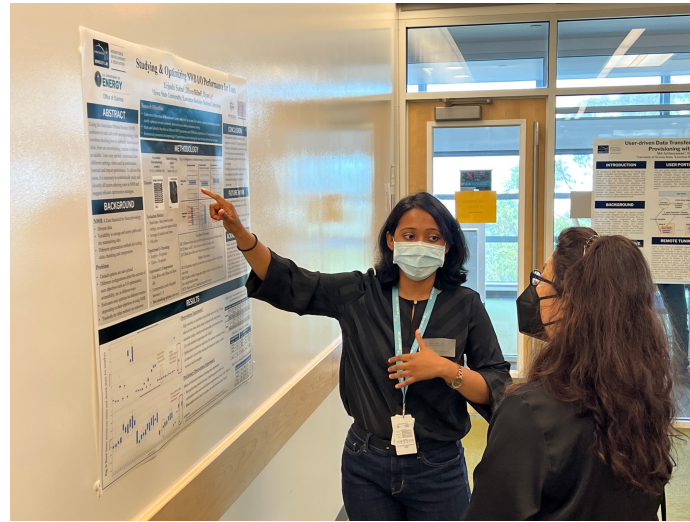
NeuroDataReHack
Generating new insights from existing neurophysiology data through secondary analysis
July 8-12, 2024
Janelia Research Campus
Ashburn, VA, USA
- Sep. 26-27: NWB Workshop at the INCF Neuroinformatics Assembly (Austin, TX) ★
- Oct. 5 - 9: Booth at SfN 2024 (Chicago, IL)



LBNL Internships and Summer Program

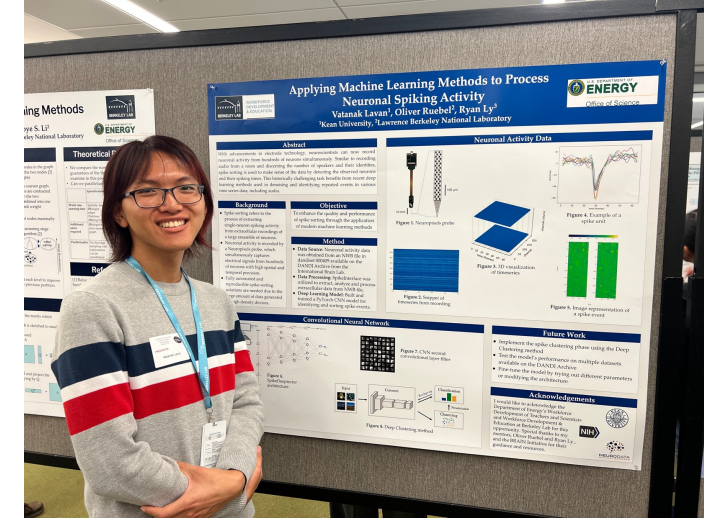
Urjoshi Sinha
Iowa State University
Graduate Student

**Studied and optimized
I/O performance for
large-scale NWB data
using chunking and
compression in HDF5**



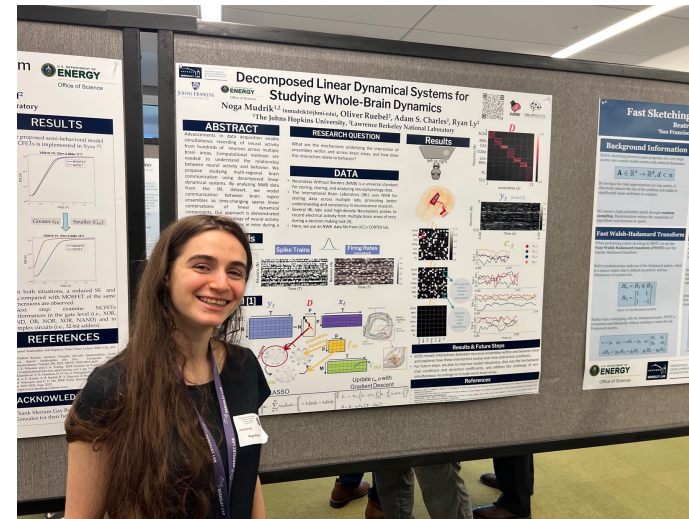
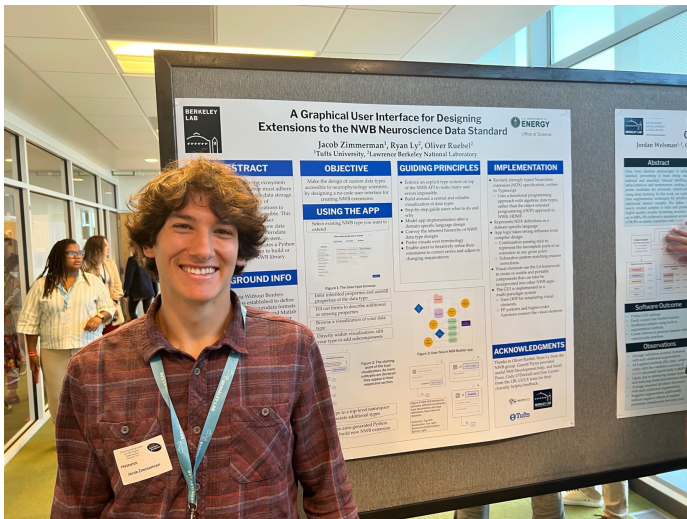
Vatanak Lavan
Kean University
Senior Undergraduate Student

**Developed a deep
convolutional neural
network for classification
of spike events using
public NWB data on
DANDI for training**



Jacob Zimmerman
Tufts University
Senior Undergraduate Student

**Developed a web-based
graphical user interface
(GUI) for designing NWB
extensions**



Noga Mudrik
Johns Hopkins University
PhD Candidate

**Studied brain multi-
regional communication
using decomposed
linear dynamics
systems using public
NWB data on DANDI**



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.