



# COSYNE 2023 Tutorial:

## The DANDI Archive and Neurodata Without Borders Data Standard for Neurophysiology

Ryan Ly



Cody Baker





# COSYNE 2023 Tutorial:

## The DANDI Archive and Neurodata Without Borders Data Standard for Neurophysiology

Ryan Ly



Cody Baker



# Preamble

- Slides and notebooks: <https://bit.ly/nwb-cosyne-2023>
- No pressure to follow along with the code in real time. You can work through the notebooks at your own pace.
- Ask questions to the NWB & DANDI teams – they are online and ready to help!

	NWB	DANDI
<b>Slack</b>	<a href="https://bit.ly/nwb-slack">https://bit.ly/nwb-slack</a>	See email after registering on DANDI
<b>GitHub</b>	<a href="https://bit.ly/nwb-helpdesk">https://bit.ly/nwb-helpdesk</a>	<a href="https://bit.ly/dandi-helpdesk">https://bit.ly/dandi-helpdesk</a>

# Tutorial Outline

8:00 – 8:30: Introduction to NWB and DANDI

8:30 – 8:55: How to find relevant data on DANDI and use the DANDI Hub

8:55 – 9:00: Break

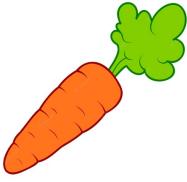
9:00 – 9:20: How to read and explore an NWB file in Python and MATLAB

9:20 – 9:50: How to analyze NWB data using popular data processing and visualization tools

# Introduction to NWB and DANDI: Outline

1. Motivation
2. What is NWB?
3. How does NWB work?
4. What is DANDI?
5. Overcoming the energy barrier of data standardization
6. How do I get started using NWB and DANDI?

# Motivation



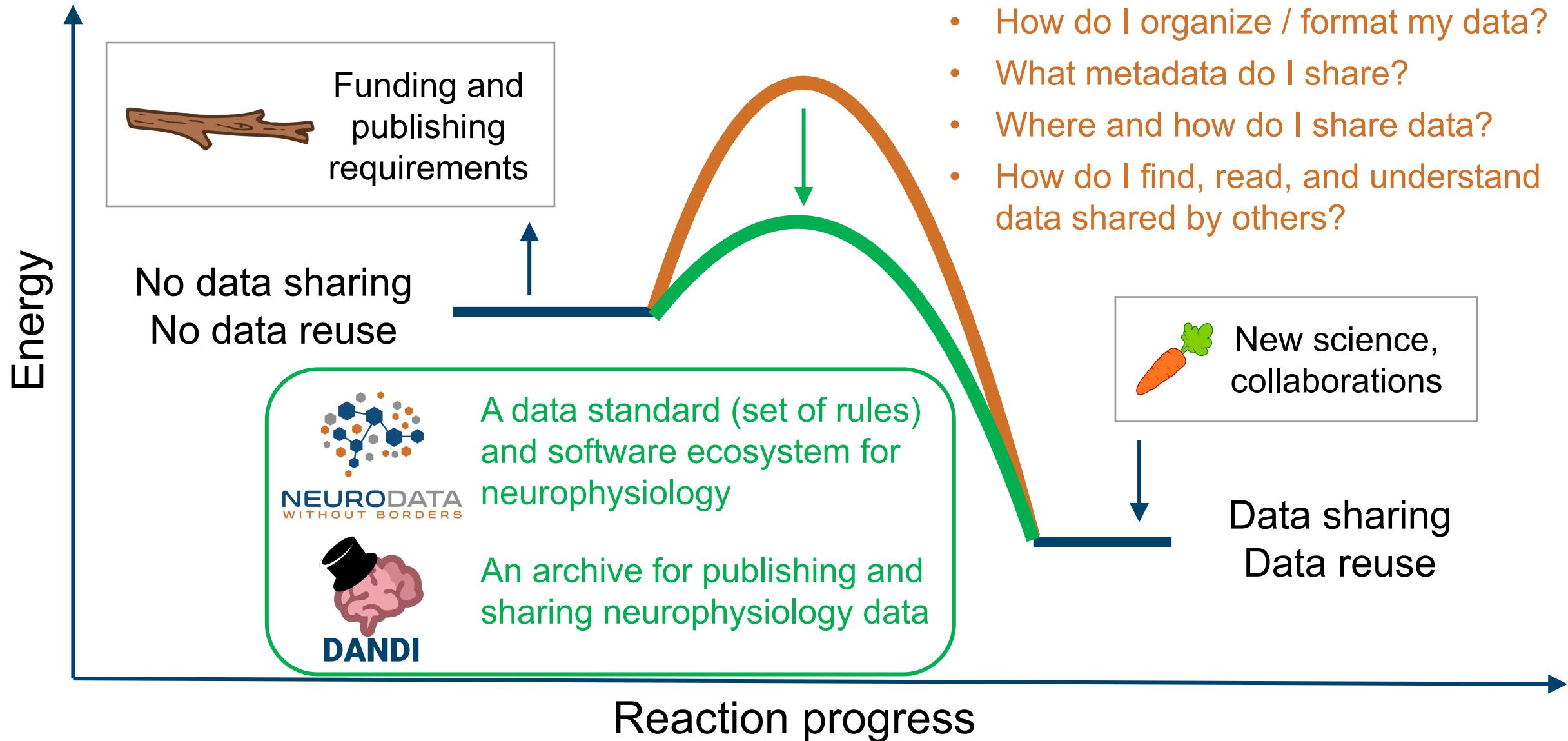
- Sharing data within a lab enables new science
- Sharing data outside a lab enables collaborations and new science
- Reusing shared data maximizes the return on investment of a dataset
  - Test generality of a finding
  - Explore follow-up questions to a study
  - Develop and evaluate new analysis methods and software tools
  - Also new science!

# Motivation

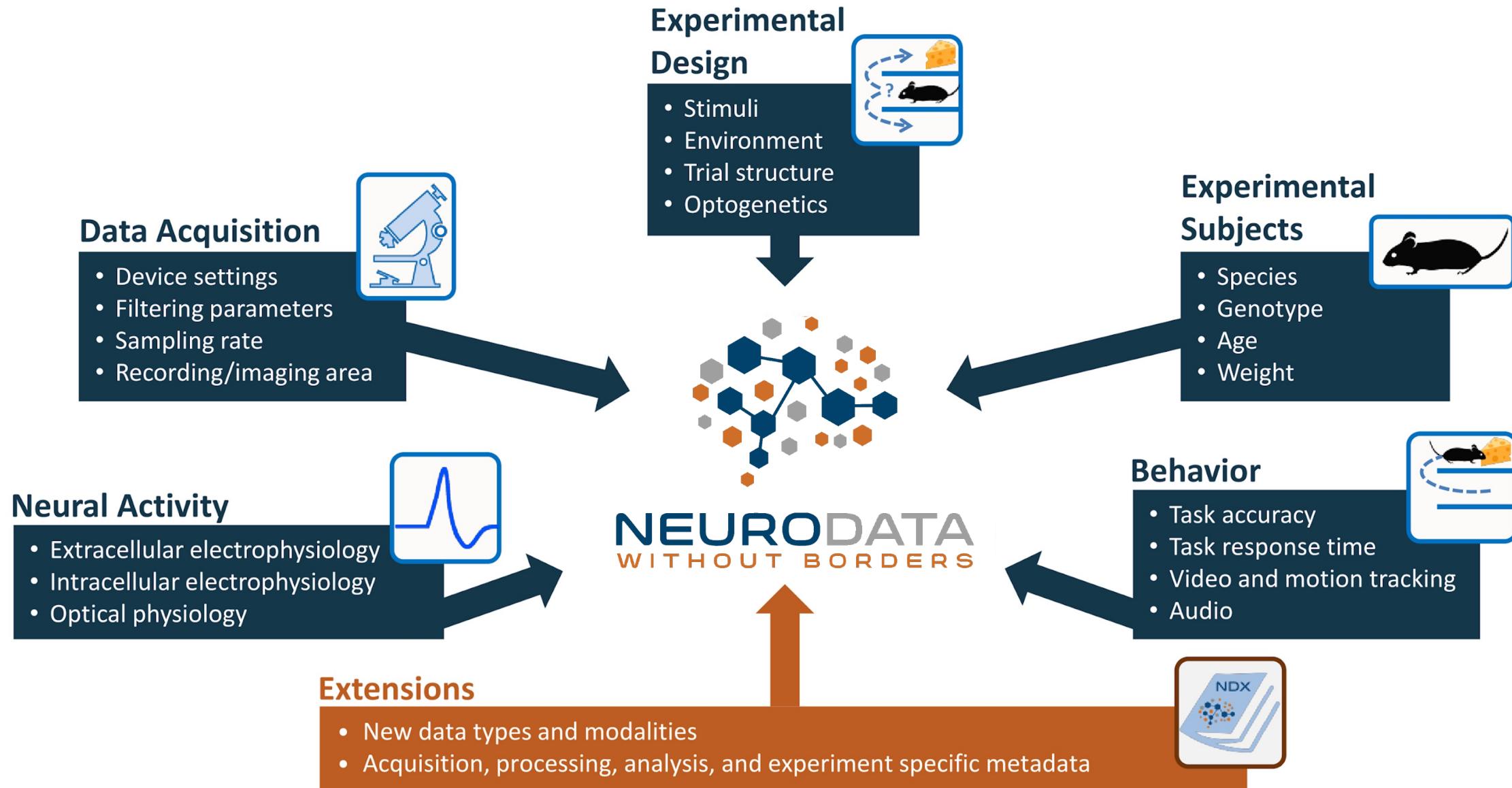


- Sharing data is required:
  - When publishing
  - When using US NIH funds (as of 2023 grants):
    - Under the new NIH Data Management and Sharing (DMS) policy, NIH expects that investigators and institutions:
      - “Plan and budget for the managing and sharing of data
      - Submit a DMS plan for review when applying for funding
      - Comply with the approved DMS plan”
    - DMS plans include:
      - Description of data standard
      - Specification of data archive

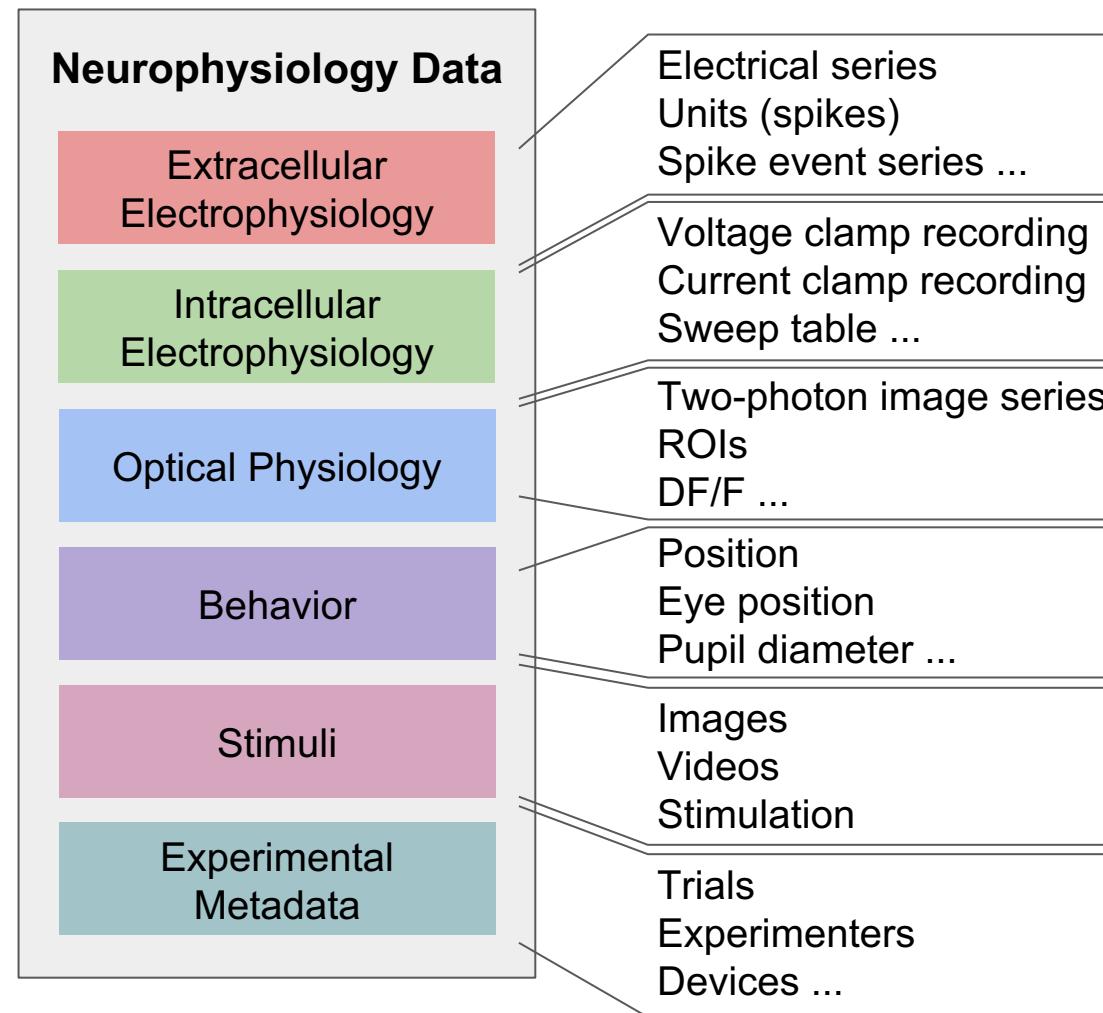
# The Energy Barrier of Data Sharing



# NWB defines a unified data format for neurophysiology



# NWB provides structured neurodata types for most of neurophysiology data



# NWB organizes your neurophysiology data and metadata



↳ general

↳ subject (neurodata\_type: Subject)

↳ species : Mus musculus

↳ experimenter :

- Ryan Ly
- Cody Baker

↳ extracellular\_ephys

↳ electrodes :

id	x	y	z	location	...
0	0.1	0.5	0.1	SSp-II1	...
1	1.1	0.5	2.1	SSp-II1	...
2	0.5	1.0	1.1	SSp-II2/3	...

↳ acquisition

↳ processing

↳ units

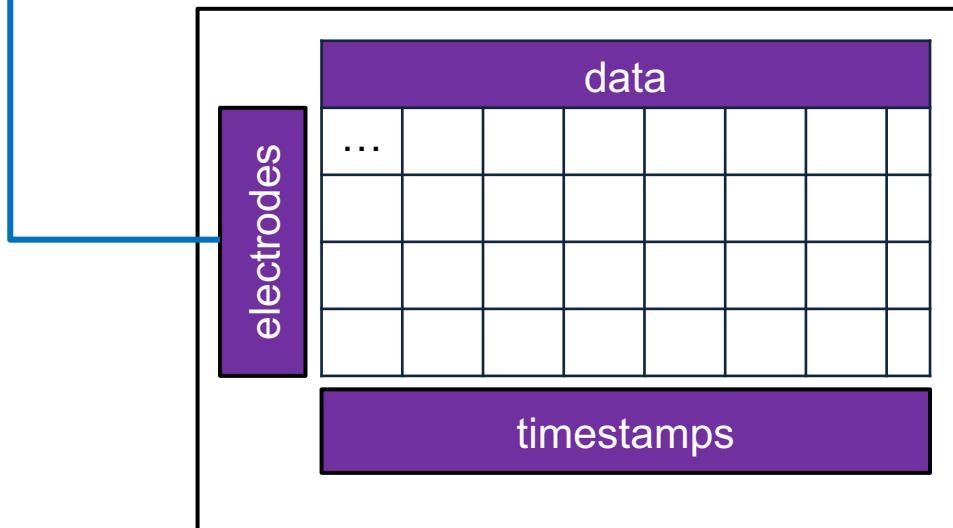
↳ intervals

...

# NWB organizes your neurophysiology data and metadata



↳ L general  
↳ L acquisition  
↳ L ElectricalSeries (neurodata\_type: ElectricalSeries)



- Description
- Unit of measurement
- Conversion factor
- Channel conversion factor
- Offset
- Sampling rate

↳ L processing  
↳ L units  
↳ L intervals  
...

# NWB organizes your neurophysiology data and metadata



- └ general ←
- └ acquisition
- └ processing
- └ units (neurodata\_type: Units)

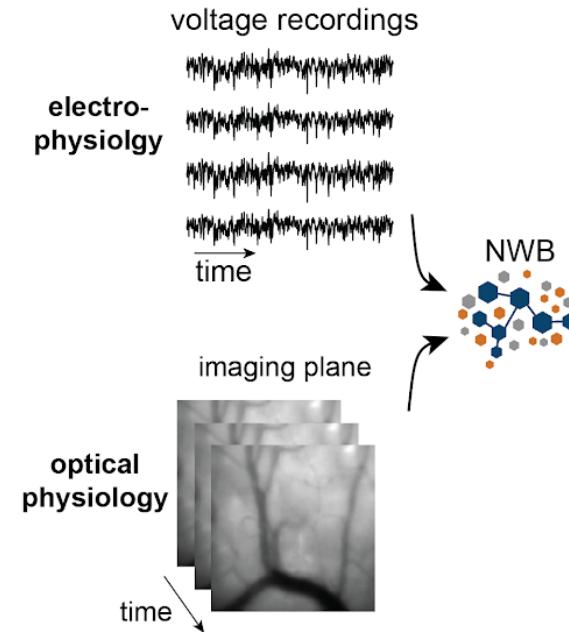
id	spike_times	waveform_mean	waveform_sd	electrodes	...
0	[0.03, 0.14]	[0.01, 0.03, ...]	[0.04, 0.06, ...]	[0, 1, 2]	...
1	[0.6, 0.71, 0.82]	[0.12, 0.13, ...]	[0.06, 0.08, ...]	[3, 5, 6, 8]	...
2	...	...	...	...	...

- └ intervals

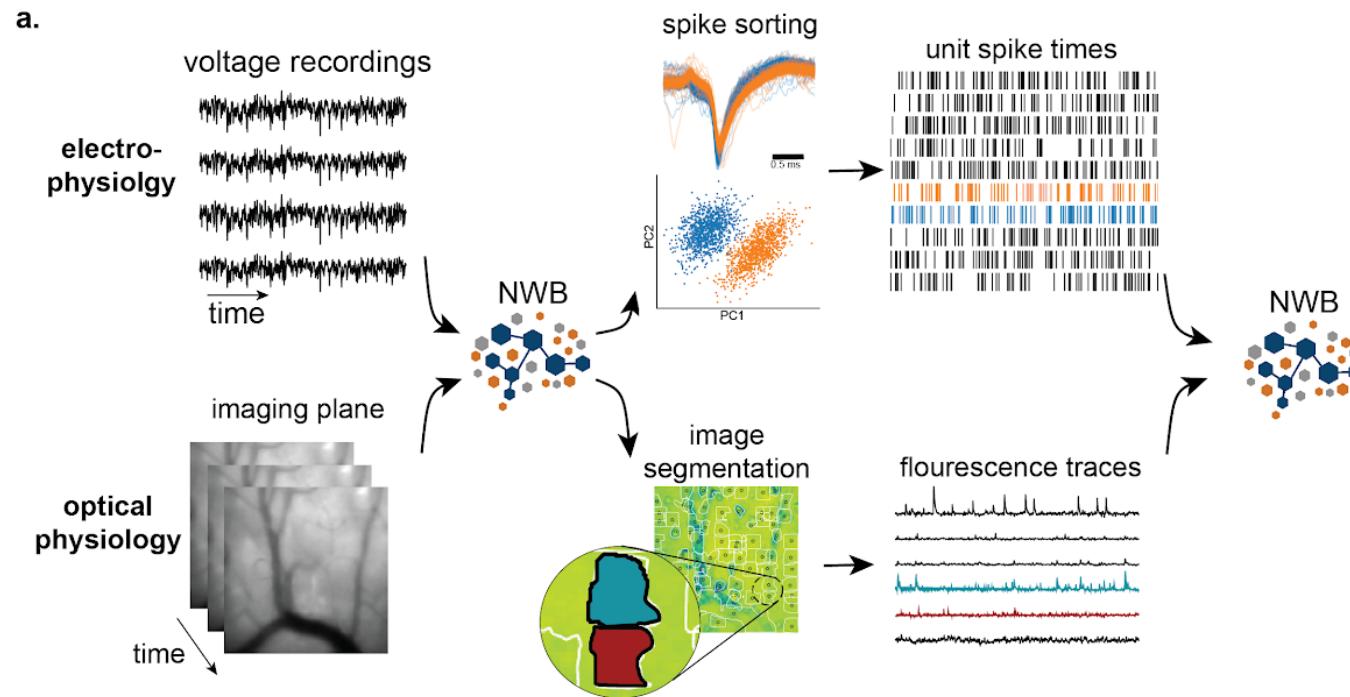
...

# NWB enables unified storage of multimodal raw and processed data

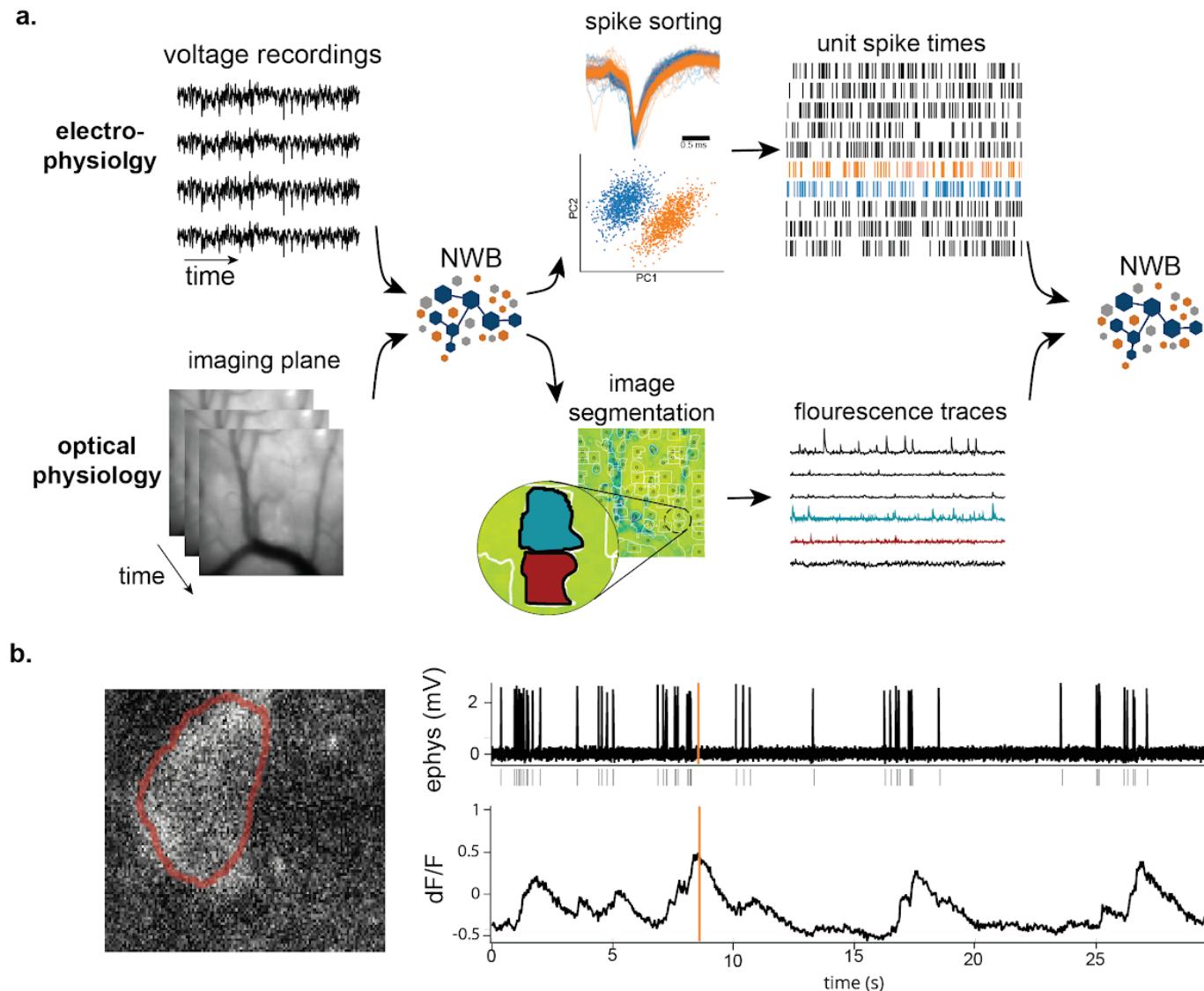
a.



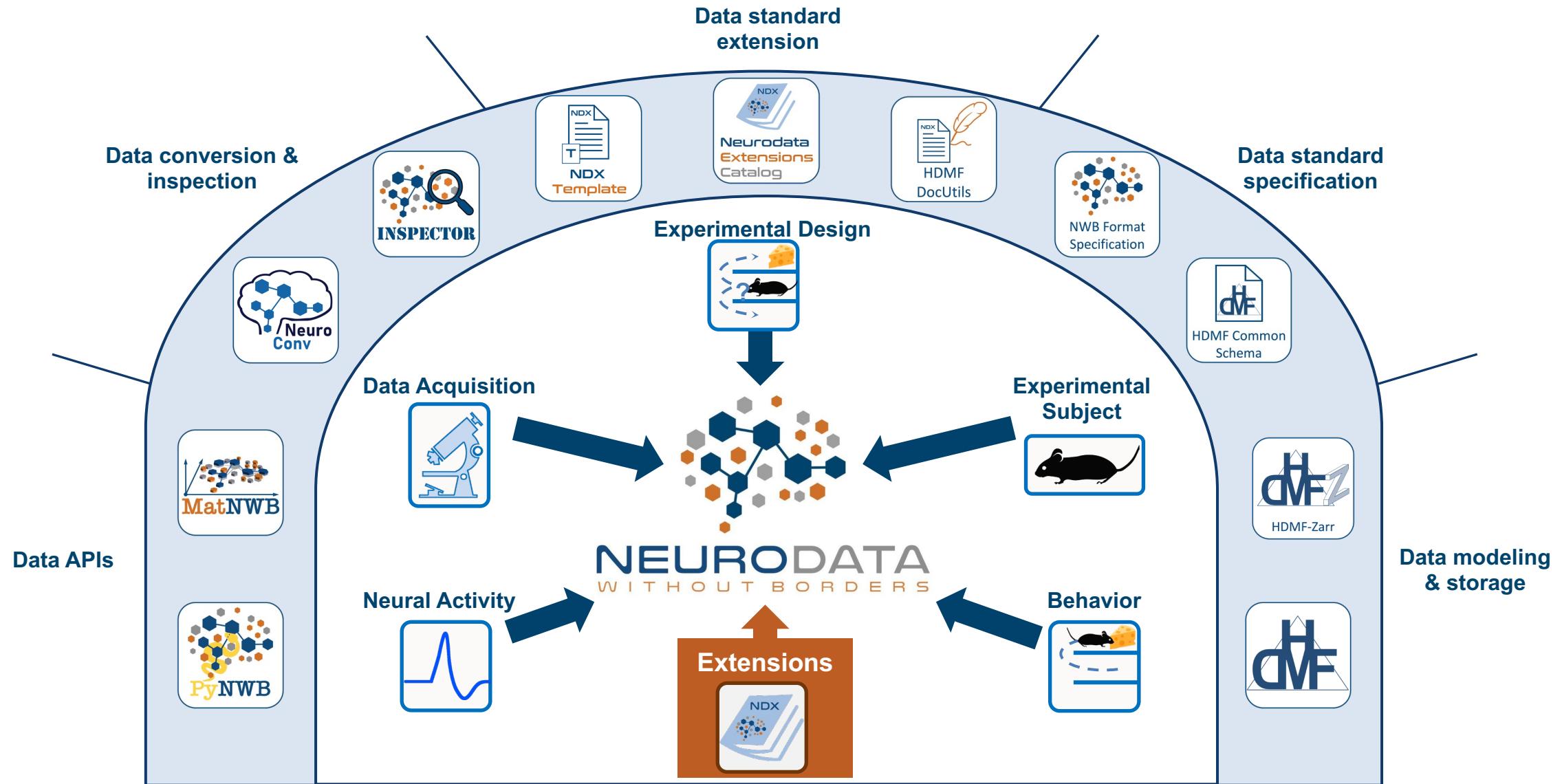
# NWB enables unified storage of multimodal raw and processed data



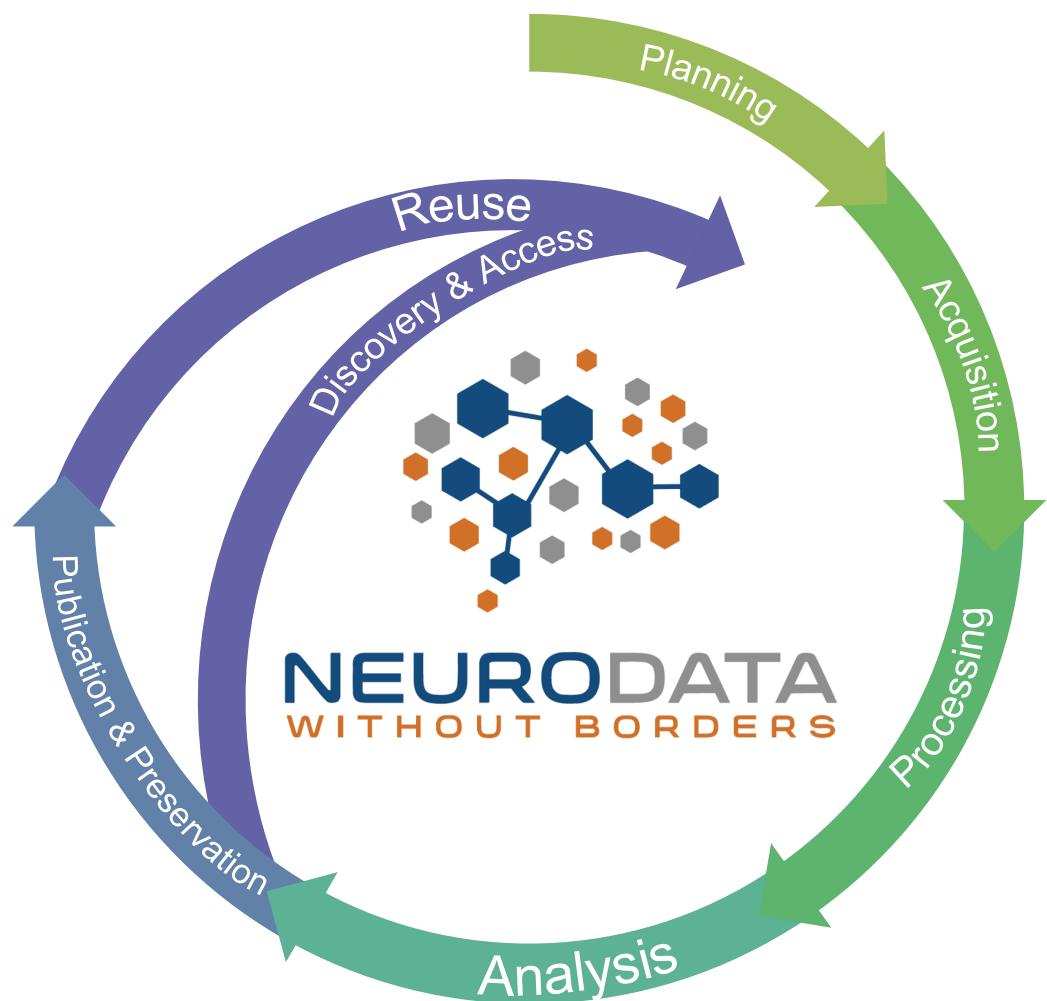
# NWB enables unified storage of multimodal raw and processed data



# An ecosystem for neuroscience data standardization

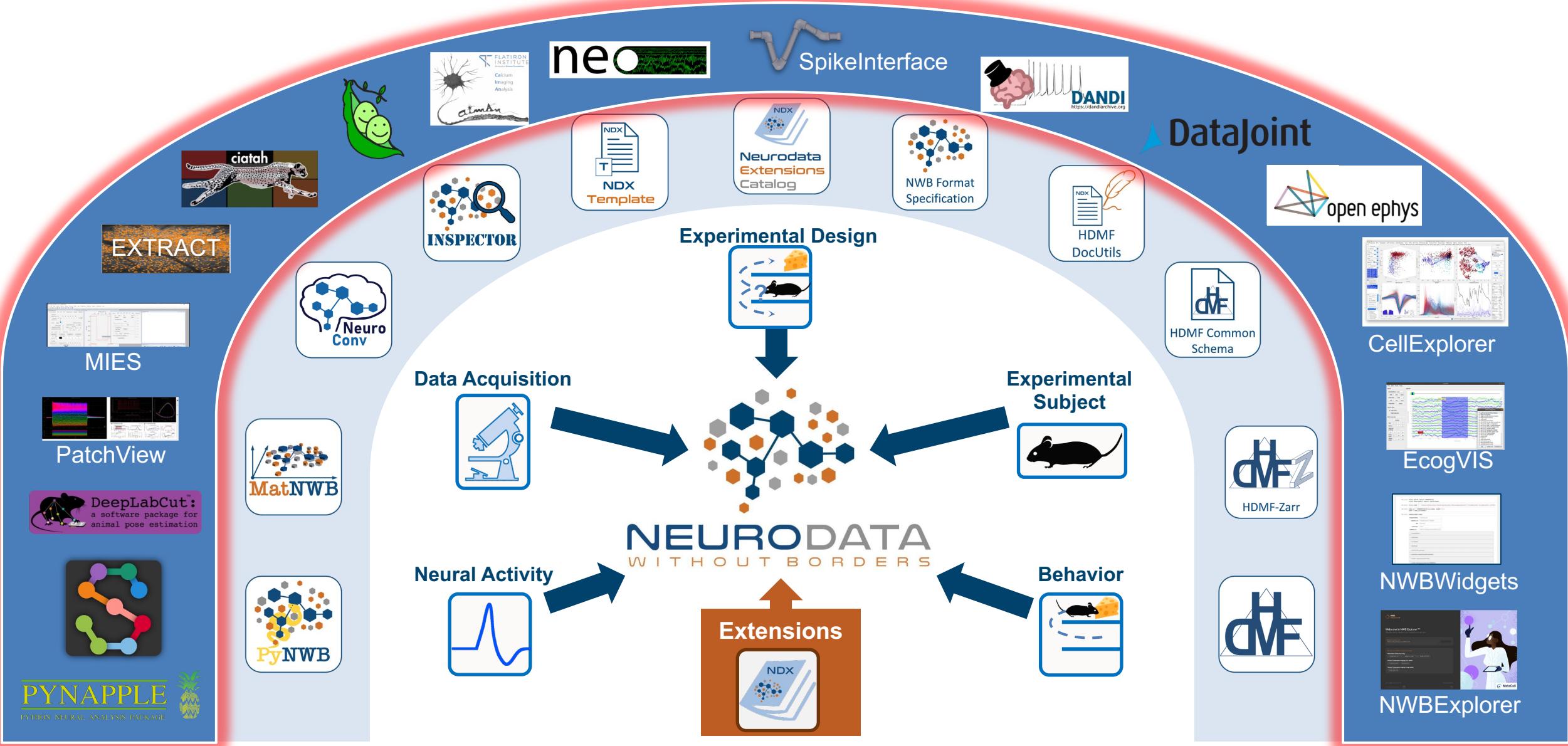


# NWB technologies at the heart of the neurodata lifecycle



- Data standards are a critical conduit that facilitate:
  - Flow of data throughout the data lifecycle
  - Integration of data and software across phases of the data lifecycle
- NWB must 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 Overview: <https://nwb-overview.readthedocs.io>

The screenshot displays two pages from the NWB Overview documentation:

- Glossary of Core NWB Tools**: This page provides a quick overview of key software packages. It includes sections for **FOR USERS** (Intro to NWB, Converting neurophysiology data to NWB, Reading NWB Files, Extending NWB), **FOR DEVELOPERS** (Accessing NWB Sources), and **Analysis and Visualization Tools**. It features sections for PyNWB (Python reference API) and MatNWB (MATLAB library). A large orange arrow points from this page to the right.
- Analysis and Visualization Tools**: This page is a collection of tools cataloged as a convenience reference for NWB users. It includes sections for **FOR USERS** (Intro to NWB, Converting neurophysiology data to NWB, Extending NWB, Glossary of Core NWB Tools, Reading NWB Files), **Analysis and Visualization Tools** (NWB Widgets, NWB Explorer, SpikeInterface, CellExplorer, EcogVIS, CalmAn, suite2p, CIAtah, EXTRACT, MIES, DeepLabCut, pynapple), and **Extracellular Electrophysiology Physiology Tools** (SpikeInterface, CellExplorer). A second large orange arrow points from this page back to the left.

# Convert your data to NWB using NeuroConv



- <https://neuroconv.readthedocs.io/>
- Reads data from 36 popular neurophysiology data formats and writes to NWB using best practices
- Handles large data volume by reading datasets piece-wise
- Minimizes file size by automatically applying chunking and compression

# Accessing NWB data on DANDI



<https://dandiarchive.org>

The DANDI Archive  
The BRAIN Initiative archive for publishing and sharing neurophysiology data including electrophysiology, optophysiology, and behavioral time-series, and images from immunostaining experiments.

Search Dandisets by name, description, identifier, or contributor n...

253 dandisets 765 users 486 TB total data size

Sort By: Modified Identifier Name Size

two photon series

**MICrONS Two Photon Functional Imaging**  
0.230307.2132 · DANDI:000402 · Contact Fahey, Paul G. · Updated on March 7, 2023 · 19 · 1.3 TB

**Guide to the construction and use of an adaptive optics two-photon microscope with direct wavefront sensing**  
0.230302.2331 · DANDI:000454 · Contact Kienfeld, David · Updated on March 2, 2023 · 4 · 7.4 GB

**Allen Institute Openscope - Credit Assignment project**  
 DANDI:000037 · Contact Gillon, Colleen J. · Updated on February 23, 2023 · 100 · 125.4 GB

Related resources

Experience-dependent contextual codes in the hippocampus  
ID: DOI:10.1038/s41593-021-00816-6  
Repo: Nature Neuroscience volume 24  
Relation: dcite:IsDescribedBy

Assets Summary

**Species**  
House mouse

**Approach**  
behavioral approach

microscopy approach; cell population imaging

**Data Standard**  
Neurodata Without Borders (NWB)

RRID:SCR\_015242

**Number Of Subjects**  
10

**Variable Measured**

PUBLIC DANDISETS ABOUT DOCUMENTATION HELP DANDIHUB LOG IN

000054 / sub-F1 Size

.. 29.8 GB

sub-F1\_ses-20190407T210000\_behavior+ophys.nwb 24.8 GB

sub-F1\_ses-20190409T210000\_behavior+ophys.nwb 27.5 GB

**Hippus. Nat** Dandiset Actions DOWNLOAD

Download full dandiset HELP ?

Use this command in your DANDI CLI

dandi download DANDI:000054/0.210819.1547

Download a different version?

Draft  
 Latest version  
 Other version

dandi download https://dandiarchive.org/dandiset/

# Visualize and analyze data on DANDI Hub

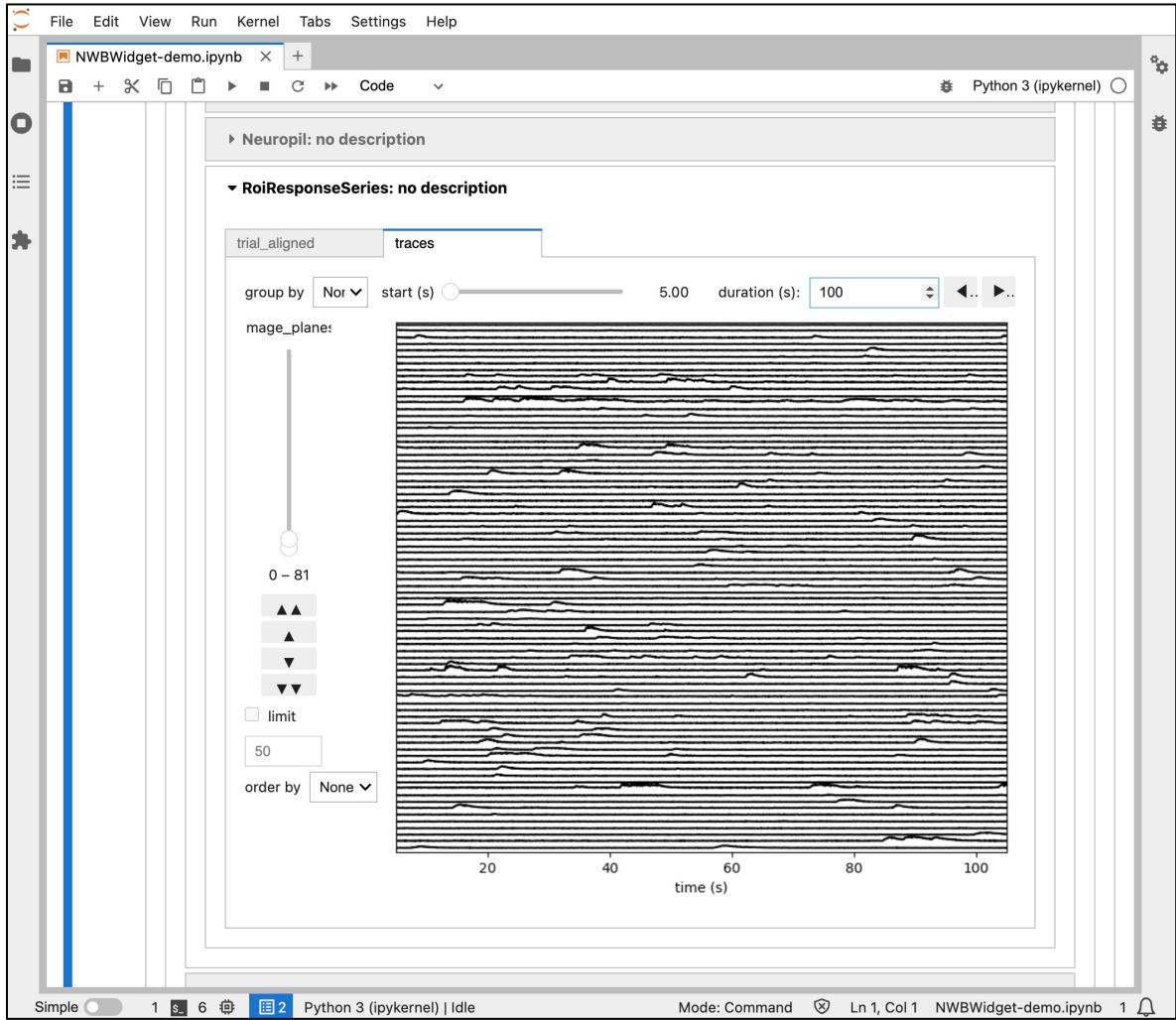
<https://hub.dandiarchive.org/>



A screenshot of the JupyterHub interface showing "Server Options". It lists several server configurations:

- Tiny. Useful for many quick things  
0.5 CPU / 1 GB
- Base**  
6 CPU / 16 GB upto 12C/32G. May take up to 15 mins to start.
- Large  
12C/32G upto 24C/64G. May take up to 15 mins to start.
- Extra large  
24C/64G upto 48C/96G. May take up to 15 mins to start.
- T4 GPU for inference  
4 CPU / 15 GB / 1 T4 GPU. May take up to 15 mins to start.
- Beta (MATLAB - Bring your own license)  
6 CPU / 16 GB upto 12C/32G. May take up to 15 mins to start. This requires your own license.

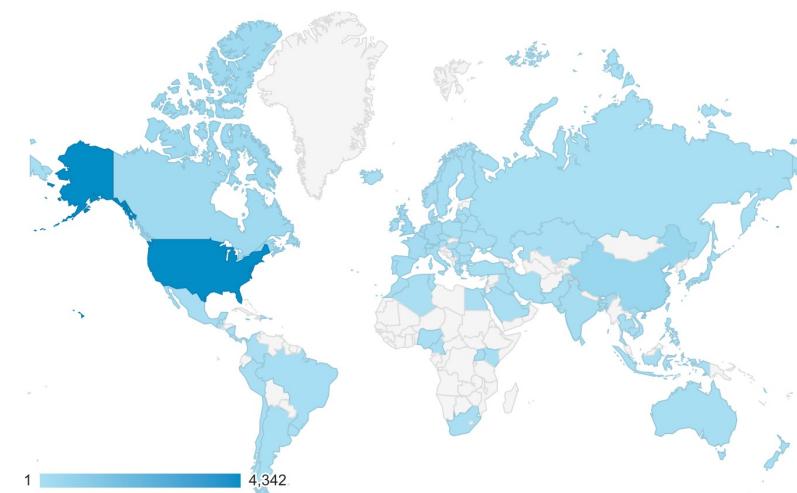
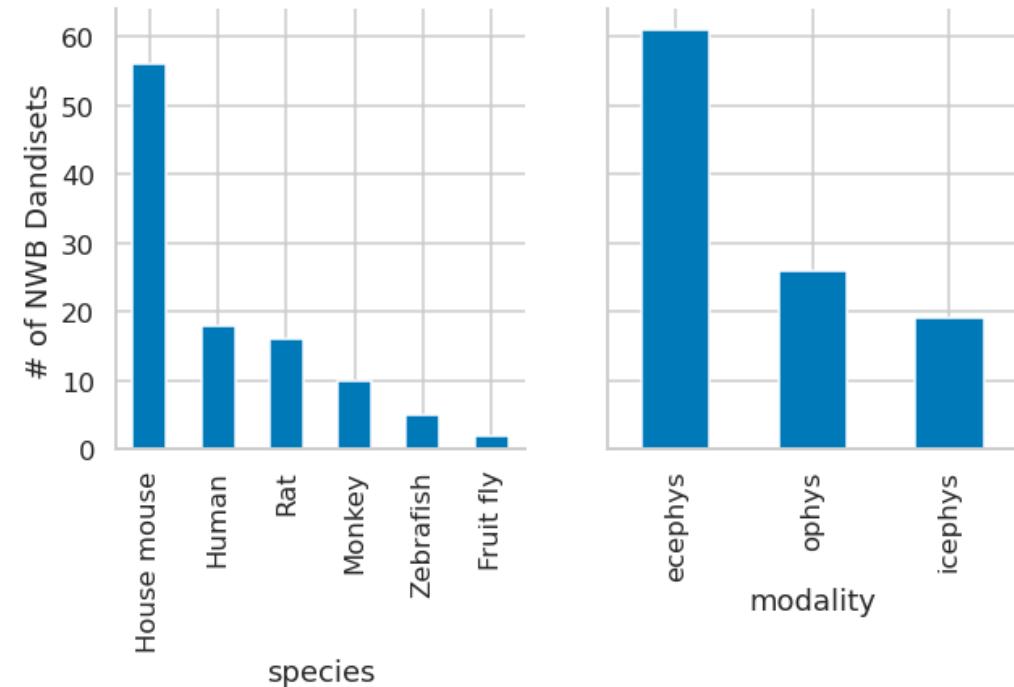
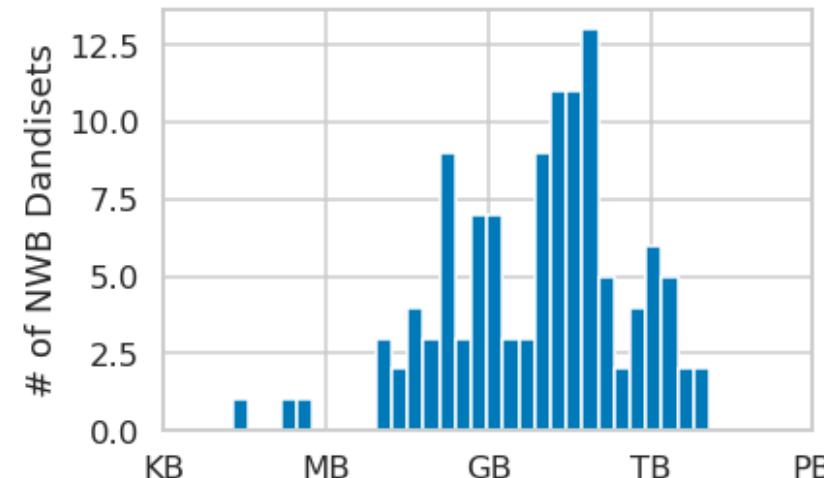
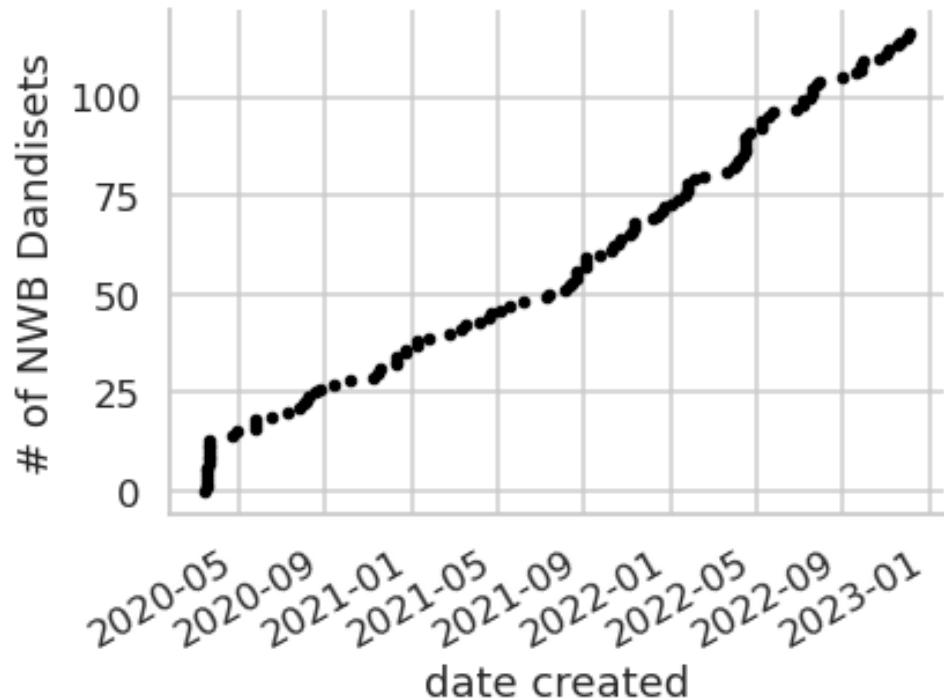
An orange arrow points from the left towards the "Base" option. At the bottom is a large orange "Start" button.





DANDI

# Growth and diversity of NWB data in DANDI



# We want to help you standardize, share, and reuse data!

- Online and in-person training
- July 24 – 26: [NWB User Days](#) (training workshop) @ HHMI Janelia (Ashburn, VA, USA)
- July 27 – 29: [NWB Developer Days](#) (developer hackathon) @ HHMI Janelia (Ashburn, VA, USA)
- Sep. 5 – 8: [NeuroDataReHack](#) (reuse hackathon) @ IBRO World Congress (Granada, Spain)

Go to <https://bit.ly/nwb-events> for a full list

- Consulting and support for labs and tool developers to integrate with NWB
- Working groups to refine NWB to meet evolving needs of the community

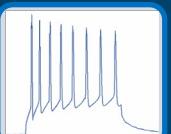
# **End of Tutorial Slides**

# Multidisciplinary team science at work

## Applications



Extracellular  
electrophysiology



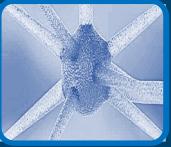
Intracellular  
electrophysiology



Optical physiology



Behavior



Simulations

## Technology Teams



O. Rübel



A. J. Tritt



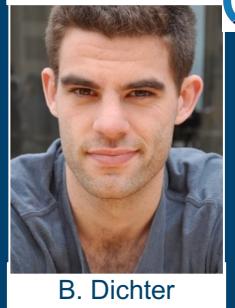
R. Ly



M. Avaylon



CATALYST NEURO



B. Dichter



C. Baker



A. Weigl



L. Niu



ALLEN INSTITUTE for  
BRAIN SCIENCE



S. De Vries



A. Buccino



L. Ng



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. Kepcs  
(WUSTL)



M. Murthy  
(Princeton)



H.D. Mansvelder  
(VU Amsterdam)



U. Rutishauser  
(Cedars Sinai)

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

## Sponsors



The BRAIN Initiative®



## Industry Engagement

- CatalystNeuro
- DataJoint
- MBF Biosciences
- 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!



# Multidisciplinary team science at work

## Dartmouth

- Yaroslav O. Halchenko
- John T. Wodder II
- Matt Van der Meer

## Kitware

- Roni Choudhury
- Daniel Chiquito
- Mike VanDenburgh
- Jake Nesbitt
- Brian Helba

## MIT

- Satrajit Ghosh
- Dorota Jarecka
- Sanu Abraham
- Jakob Voigts
- Mark Harnett

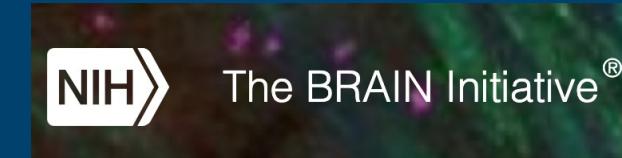
## CatalystNeuro

- Ben Dichter
- Luiz Tauffer
- Alessio Buccino
- Cody Baker
- Daniel Sotoude

## Collaborators

- Neurodata Without Borders
- Brain Cell Data Center  
(BCDC/BICCN)
- Allen Institute for Brain Science
- BIDS community
- OpenScope project
- INCF
- International Brain Initiative

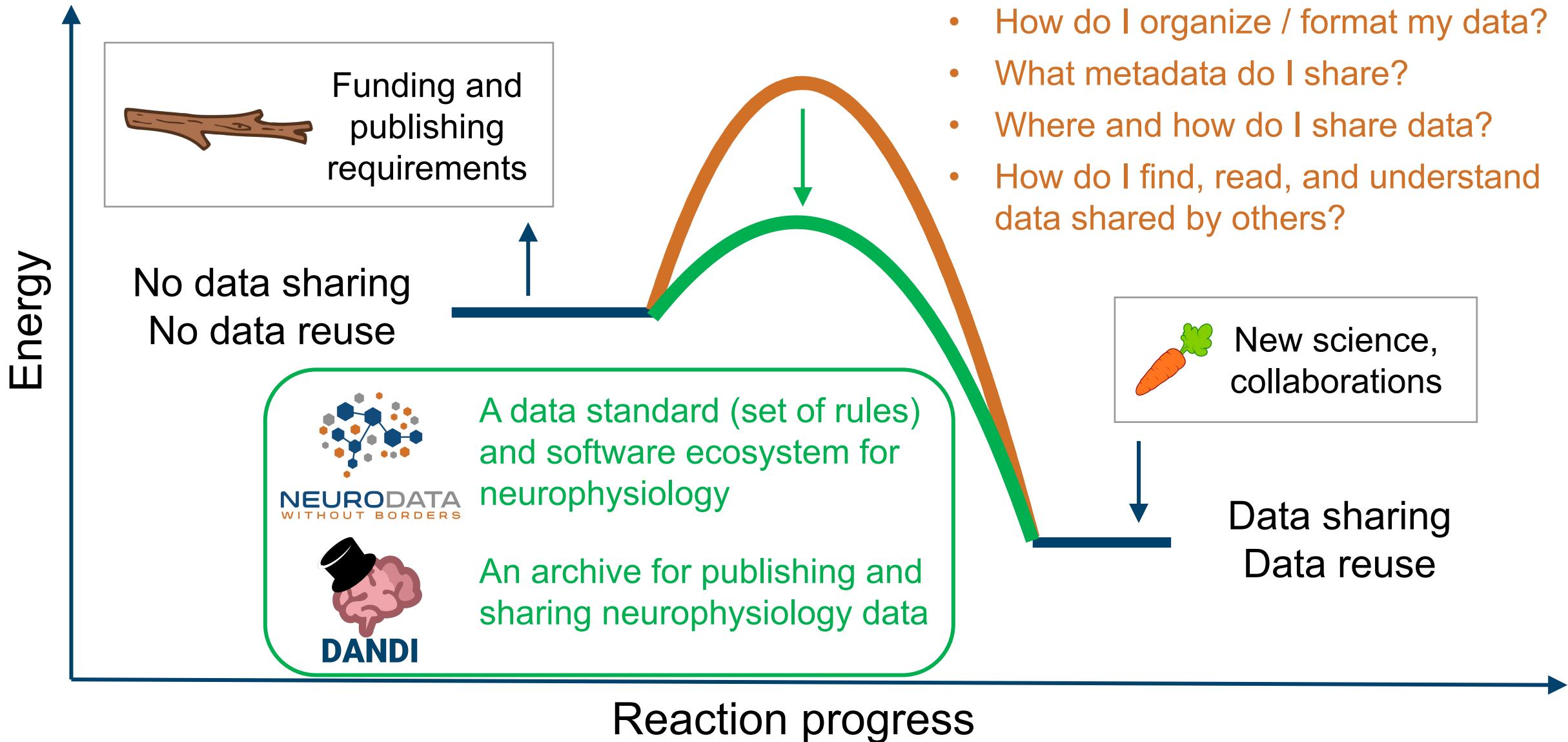
## Support



1R24MH117295

AWS Open Dataset  
Program

# The Energy Barrier of Data Sharing



# Upcoming Events

- **July 24 – 26:** [NWB User Days](#) (training workshop) @ HHMI Janelia (Ashburn, VA, USA)
- **July 27 – 29:** [NWB Developer Days](#) (developer hackathon) @ HHMI Janelia (Ashburn, VA, USA)
- **Sep. 5 – 8:** [NeuroDataReHack](#) (reuse hackathon) @ IBRO World Congress (Granada, Spain)

Go to <https://bit.ly/nwb-events> for a full list

## Resources

Get started with NWB: <https://nwb-overview.readthedocs.io>

Get started with DANDI: <https://www.dandiarchive.org/handbook/>

**Please give us feedback in this 2-minute survey**

<https://bit.ly/nwb-cosyne-survey>



# Legal



**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.