

## Report:

# 2022 NWB Users Days at HHMI Janelia

## 13th NWB Hackathon: July 25 - 27, 2022

Oliver Rübel, Benjamin Dichter, Ryan Ly

## Contents

<b>1</b>	<b>Executive Summary</b>	<b>3</b>
	<b>Participants</b>	<b>4</b>
	<b>Program</b>	<b>6</b>
	<b>Projects</b>	<b>7</b>
	<b>5 Exit Survey</b>	<b>9</b>
	5.1 Experience at the Workshop? . . . . .	9
	5.2 How can we improve future training events? . . . . .	12
	5.3 NWB Software . . . . .	14
	5.4 Your Data . . . . .	15
	<b>6 Photos from the Event</b>	<b>16</b>
	<b>Acknowledgements</b>	<b>19</b>
	<b>Disclaimer</b>	<b>19</b>



2

# 1 Executive Summary

**Overview:** The Neurodata Without Borders: Neurophysiology (NWB) project is an effort to standardize the description and storage of neurophysiology data and metadata. NWB enables data sharing and reuse and reduces the energy barrier to applying data analytics both within and across labs. The goal of the NWB User Days is to bring the experimental neurophysiology community together to train new users and further adoption and development of NWB and NWB software and tools. The 13th NWB Hackathon was an in-person User Days training event held at the HHMI Janelia Research Campus in Ashburn.

**Participants:** 28 participants registered for the event, of which 21 attended in-person, 3 participated virtually via Zoom, and 5 registrants could not attend. Participants came from 16 different institutions. Current positions of registrants were divided across Ph.D. Students and Post Doctoral Researchers (9), Researchers and Research Assistants (7), and Engineers (14).

**Program:** Building on the success and experience from previous NWB User Days, the program was designed to provide a balance between NWB and tool tutorials in the morning while providing dedicated time for hacking on projects in the afternoon. All presentations were optional and a second room for hacking was available to participants at all times. Housing was provided onsite by HHMI Janelia for all participants. Opportunities for social interaction and team discussions were available throughout the event during the open hacking sessions, during the provided group breakfast, lunch, and dinner, as well as during the speed-dating reception on Day 1 and after the event in Bob's bar.

**Outcomes:** During the event participants worked on projects related to: **1)** conversion of data to NWB, e.g., fly optical physiology data, Marmoset intracellular electrophysiology data, and Neuropixel extracellular electrophysiology data, **2)** development of extensions for integration of new data types and tools with NWB, e.g., *ndx-probe-interface* for probe descriptions, *ndx-pose* for pose estimation, or *ndx-beadl* for behavioral task descriptions and recordings and **3)** integration and enhancement of NWB-enabled tools, e.g., DataJoint, webnwb, SLEAP, and BEADL (see Sec. 4). The event helped introduce the BEADL and DataJoint Elements tools to NWB users and resulted in the integration of the tool SLEAP with NWB. As part of the breakout session on NWB Documentation on Day 3, participants provided feedback and suggestions to further enhance the NWB online documentation. As a result—and inspired by the success of the 1st NWB Docathon earlier in March 2022—the NWB team decided to host a follow-up virtual Docathon event in Sept 22 - 23, 2022.

**Health Management:** This event was the first in-person NWB training event since the beginning of the COVID-19 pandemic. Before and during the event participants followed the COVID procedures of HHMI Janelia. This included: registration in the Janelia COVID testing program; self-testing via antigen or PCR test 72 hours prior to travel; PCR test provided by Janelia on the first and last day of the event; and voluntary antigen tests provided by HHMI Janelia. Masks were not required during the event. Overall, the procedures were effective in preventing spread as part of the event. 2 participants switched to remote attendance due to testing prior to travel and 1 participant due to a positive result soon after arrival. No cases of spread amongst participants at the event have been reported.

**Conclusion:** The event was successful in on-boarding new NWB users and facilitating the integration of new tools with NWB. The event helped further connectedness among participants both socially and in the interconnectivity of the software they are developing. The event was overall well-received, with participants scoring their overall experience at the event at 4.8 out of 5 and the overall usefulness of the event with 4.5 out of 5 as part of the exit survey (see Sec. 5).

## Organizing Committee:

- Program Chairs:
  - Lawrence Berkeley National Laboratory: Oliver Rübel, Ryan Ly
  - CatalystNeuro: Benjamin Dichter
- Janelia: Janine Stevens and Alethea Vandamm
- The Kavli Foundation: Stephanie Albin

## NWB Event Website:

[https://neurodatawithoutborders.github.io/nwb\\_hackathons/HCK13\\_2022\\_Janelia](https://neurodatawithoutborders.github.io/nwb_hackathons/HCK13_2022_Janelia)

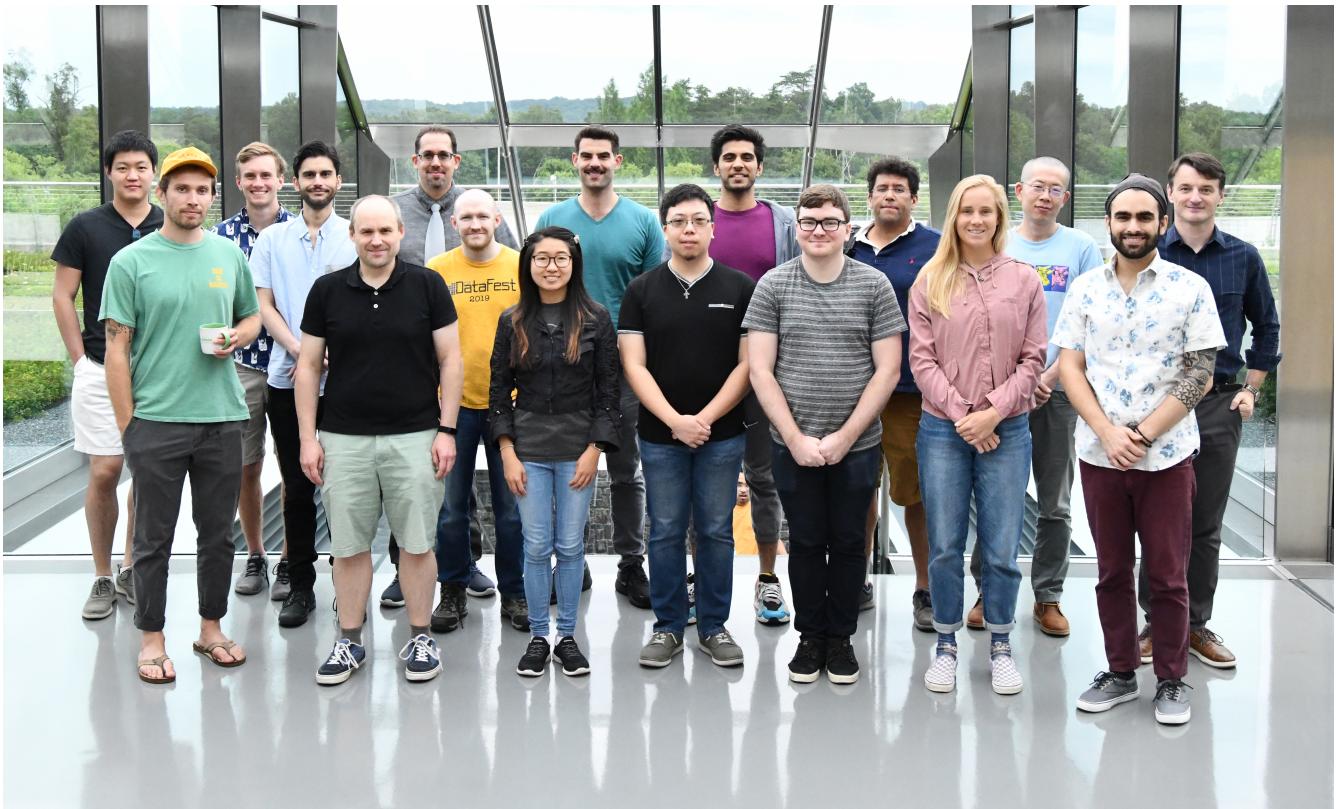


Figure 1: Participants of the NWB User Days 2022, Photo by Matt Staley, HHMI Janelia

## 2 Participants

In total 28 participants registered for the event. 21 participants attended the event in-person and 3 participated virtually via Zoom. 5 registrants did not participate due to cancellation of travel or noncompliance with HHMI Janelia COVID testing and reporting requirements.

Current positions of registrants were divided across Ph.D. Students and Post Doctoral Researchers (9), Researchers and Research Assistants (7), and Engineers (14) (see Figure 2). Participants came from 16 different institutions as shown in the summary of the number of participants per institution and across geographic regions in Table 1. 18 registrants (62.1%) had not previously attended an NWB event and 11 (37.9%) had attended at least one previous hackathon. Most registrants preferred either Python (23, 79.3%) or MATLAB (15, 51.7%) (see Figure 3). Finally, Figure 4 shows the data modalities used by registrants. 5 registrants (17.2%) requested additional travel support, with travel support being provided by the Kavli Foundation (4 registrants) and LBNL (1 registrant).

4

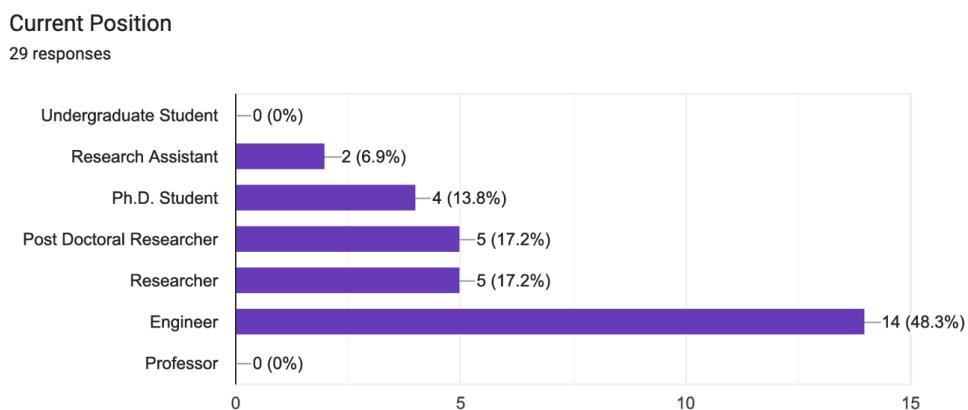


Figure 2: Current position of registrants.

Country	State	Institution	Count (in person)	Count (virtual)
U.S.A.	CA	Lawrence Berkeley National Laboratory	1	2
	CA	CatalystNeuro	1	
	IN	CatalystNeuro	1	
	VA	MBF Bioscience	1	
	CA	Brains@Play	1	
	CA	Salk Institute for Biological Sciences	2	
	CA	University of California, Los Angeles	1	
	NY	University of Rochester	1	
	MD	Johns Hopkins University	1	
	VA	Janelia Research Campus	3	
	CA	California Institute of Technology	1	
	NY	Columbia University	1	
	CA	Center17	1	
	CA	University of California, San Francisco	1	
	MO	Washington University in St. Louis	1	
	TX	DataJoint	2	
Germany		University Hospital Wuerzburg	1	1

Table 1: Number of participants per institution.

#### What programming languages are you most comfortable using?

29 responses

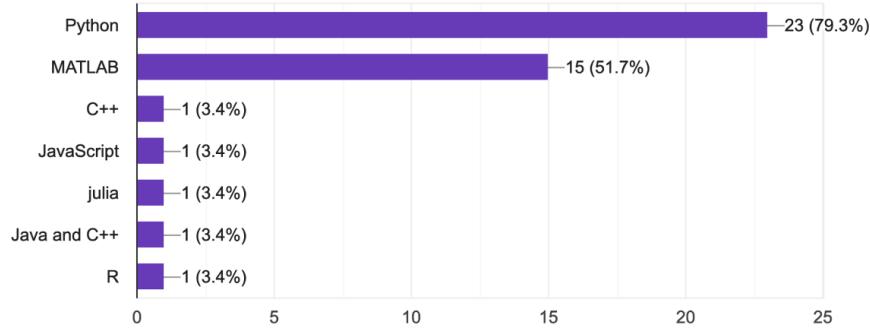


Figure 3: Programming languages preferred by registrants.

5

#### Which data modalities are you using?

29 responses

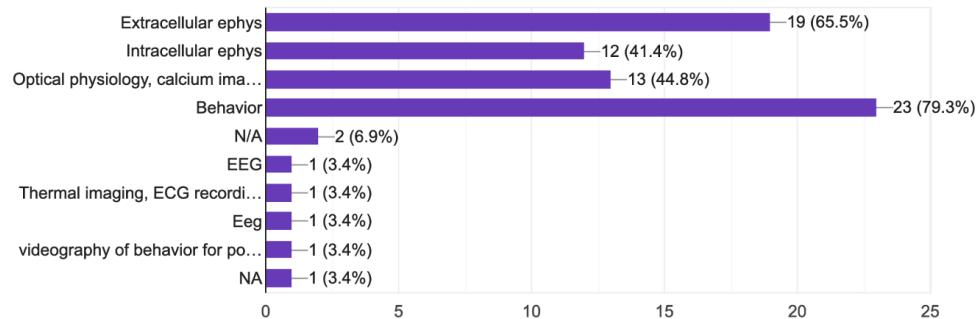


Figure 4: Data modalities used by registrants

### 3 Program

Color Legend				
Talks	Breaks	Hacking	Tutorials with hacking excercises (pick sessions that apply)	Group discussions

Day 0 (Sunday, July 24)				
6:00 PM	7:00 PM	1:00:00	Dinner (optional, Dining Room)	

Day 1 (Monday, July 25)				
Start Time	End Time	Duration	Axon-Dendrite	Spectrum
8:00 AM	9:00 AM	1:00:00	Breakfast (Dining Room)	Topic
9:00 AM	9:05 AM	0:05:00	Welcome to Janelia	
9:05 AM	9:20 AM	0:15:00	Welcome and overview of the hackathon	
9:20 AM	10:00 AM	0:40:00	Overview of NWB	
10:00 AM	11:00 AM	1:00:00	Introduction to extracellular electrophysiology	Ben Dichter, Ryan Ly
11:00 AM	12:00 PM	1:00:00	Introduction to optical physiology	Ben Dichter, Ryan Ly
12:00 PM	1:00 PM	1:00:00	Lunch (Dining Room)	
1:00 PM	2:00 PM	1:00:00	Introduction to intracellular electrophysiology	Oliver Ruebel
2:00 PM	2:45 PM	0:45:00	Overview of NWB conversion tools	Cody Baker
2:45 PM	3:15 PM	0:30:00	Hackathon project plans	
3:15 PM	3:30 PM	0:15:00	Refreshments break	
3:30 PM	6:00 PM	2:30:00	Hacking on projects	Hacking on projects
6:00 PM	7:00 PM	1:00:00	Reception / Ice Breaker (Reception)	
7:00 PM	8:30 PM	1:30:00	Dinner (Dining Room)	

Day 2 (Tuesday, July 26)				
Start Time	End Time	Duration	Axon-Dendrite	Spectrum
8:00 AM	9:00 AM	1:00:00	Breakfast (Dining Room)	Topic
9:00 AM	9:10 AM	0:10:00	Welcome to day 2	
9:10 AM	10:10 AM	1:00:00	Processing and analysis of NWB data	Oliver Ruebel
10:10 AM	10:40 AM	0:30:00	Advanced data I/O in PyNWB	Ben Dichter
10:40 AM	11:10 AM	0:30:00	Advanced data I/O in MatNWB	Oliver Ruebel
11:10 AM	12:00 PM	0:50:00	BEADL	Ben Dichter
12:00 PM	1:00 PM	1:00:00	Lunch (Dining Room)	
1:00 PM	1:10 PM	0:10:00	Group photo (meet at Reception)	Michael Wulf
1:10 PM	2:40 PM	1:30:00	Using the DANDI Archive	Cody Baker and Ben Dichter
2:40 PM	3:25 PM	0:45:00	DataJoint + NWB	Chris Brozowski
3:25 PM	6:00 PM	2:35:00	Hacking on projects	
6:00 PM	7:30 PM	1:30:00	Dinner (Dining Room)	Hacking on projects

Day 3 (Wednesday, July 27)				
Start Time	End Time	Duration	Axon-Dendrite	Spectrum
8:00 AM	9:00 AM	1:00:00	Breakfast (Dining Room)	Topic
9:00 AM	9:10 AM	0:10:00	Welcome to day 3	
9:10 AM	10:00 AM	0:50:00	NWB extensions	Ryan Ly
10:00 AM	10:30 AM	0:30:00	NWB widgets	Ben Dichter
10:30 AM	12:00 PM	1:30:00	Hacking on projects	
12:00 PM	1:00 PM	1:00:00	Lunch (Dining Room)	Hacking on projects
1:00 PM	2:30 PM	1:30:00	Hacking on projects	
2:30 PM	3:00 PM	0:30:00	NWB documentation roundtable	Cody Baker
3:00 PM	3:15 PM	0:15:00	Refreshments break	Chris Brozowski
3:15 PM	4:00 PM	0:45:00	Hacking on projects	
4:00 PM	5:00 PM	1:00:00	Final project presentations	Hacking on projects
5:00 PM	6:00 PM	1:00:00	Hacking on projects	
6:00 PM	7:30 PM	1:30:00	Dinner (Dining Room)	Hacking on projects

## 4 Projects

During the event, participants were encouraged to apply the skills they were learning to implement their own projects and to document their progress in the shared [Project GoogleDoc](#). Below a brief summary of the projects reported by participants:

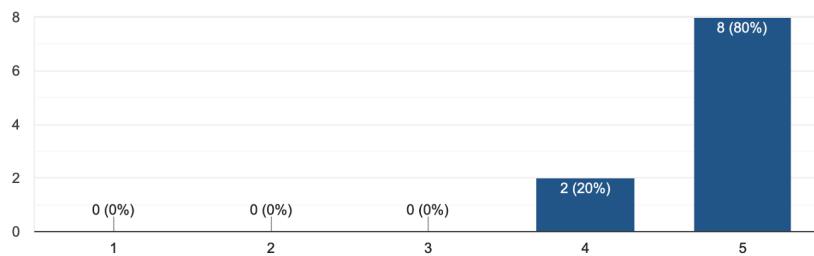
1. **Convert Marmoset monkey intracellular recording data to NWB:** Compared to extracellular and optical physiology, intracellular physiology could reveal the sub-threshold mechanisms underlying different coding principles in the brain. In vivo intracellular recording in awake animals is technically challenging. There are few publicly available intracellular recording datasets, and none of them is collected from the nonhuman primate brain. We have collected intracellular recording data from the auditory cortex of awake marmoset monkeys. To benefit the broad neuroscience community, we believe it will be helpful to release our datasets. We recorded, saved, and read those datasets using our customized MATLAB codes. Therefore, it is necessary to standardize our data before sharing it. *Key Investigators: Chenggang Chen (Johns Hopkins University), Lixia Gao (Johns Hopkins University), Xiaoqin Wang (Johns Hopkins University)*
2. **webnwb:** Building on our progress at the 2022 NWB-DANDI Developer Hackathon, refine our JavaScript API for creating and interacting with NWB files directly on the browser. *Key Investigators: Garrett Flynn (Brains@Play)*
3. **ndx-pose - Model Training Data:** The current NWB extension for pose estimation data is geared towards the results and doesn't cover model training data. We plan to further develop this extension to permit annotated data and model training parameters as applicable for both SLEAP and DeepLabCut. *Key Investigators: Chris Brozdowski (DataJoint), Liezl Maree (Salk Institute, Talmo Lab), Arlo Sheridan (Salk Institute, Talmo Lab)*
4. **NWB extension for deep brain stimulation data:** I am member of the collaborative research center (CRC) Retune (<https://retune.science/>), in which deep brain stimulation (DBS) is a core technique and is applied in both human and animal research. Standardization of DBS-related datasets, however, is basically non-existent. To address this need, our CRC drafted an extension to the BIDS standard focused at human research needs. Since we are using NWB for the standardization of all our animal research datasets, we would like to adopt this extension and implement it also as an extension to NWB. Ideally, this would eventually enable us to create a conversion tool that can convert DBS-extended BIDSified data into DBS-extended NWB files and vice versa. *Key Investigators: Dennis Segebarth (University Hospital Wuerzburg)*
5. **DataJoint Element DeepLabCut NWB export:** The DataJoint Elements project provides modality-specific open source pipelines. The extracellular electrophysiology Element has an NWB export tool that capitalizes on metadata present in Elements for lab, animal and session information. This project taps into metadata tools while also providing pose estimation data according to the relevant extension (ndx-pose). *Key Investigators: Chris Brozdowski (DataJoint)*
6. **Convert fly two-photon imaging and behavior data into NWB format:** The data is from experiments where we record Calcium dynamics with two-photon imaging in behaving flies navigating an immersive virtual reality. Thus each experiment consists of an image data file (tiff), which contains some metadata (acquired with scanimage) and a behavior data file recording the fly's movements and events in the virtual world (custom json format, generated with unity vr system). For each fly, we run multiple such experiments (trials). A dataset consists of the collection of data from multiple flies. *Key Investigators: Hannah Haberkern (Jayaraman lab, HHMI Janelia Research Campus)*
7. **DataJoint Element for BEADL:** The project designs a software module (DataJoint Element) for database storage/queries and automated ingest, synchronization, and analysis. The project will be hosted at <https://github.com/datajoint/element-beadl> *Key Investigators: Dimitri Yatsenko (DataJoint)*
8. **Converting pre-processed Neuropixels data into nwb format and doing visualization:** Converting pre-processed Neuropixels data into nwb format and doing some test visualization to confirm all the information is saved correctly. *Key Investigators: Chaoqun Yin (UCLA)*
9. **ndx-probe-interface:** Many researchers use SpikeInterface to sort spikes using different algorithms. One important feature that spike sorting algorithms require is the shape of the electrode and location of the electrodes themselves, however, this information is not saved to NWB files when a user would like to output to NWB from SpikeInterface. Additionally, to use SpikeInterface's Probe visualization software, one would need a ProbeInterface object with planar polygon information, so it is crucial that we save this data in an NWB extension to allow for reconstruction of this object. *Key Investigators: Geeling Chau (California Institute of Technology, Yue Lab)*

10. **Create an NWB File Manager GUI that explores and visualizes data across multiple NWB Files:** Working through multiple NWB files, and aggregating key data from them, so the user can lookup information in a Matlab GUI, and perform various comparisons. *Key Investigators: Himanshu Ahuja (Brain and Cognitive Science, University of Rochester)*
11. **Extend usage of NWB to include more of the ecosystem:** We recently switched to using NWB files to store some of our data in the lab. Having already done a first set of data conversions, the goal for this project is to learn more about the NWB ecosystem in order to check our current procedures, and leverage more of the available tools. *Key Investigators: Thomas Donoghue (Columbia University)*

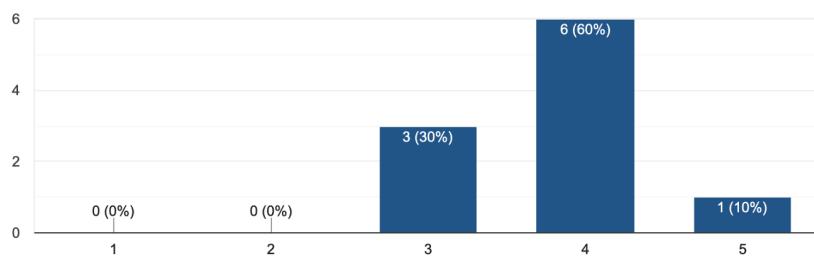
## 5 Exit Survey

### 5.1 Experience at the Workshop?

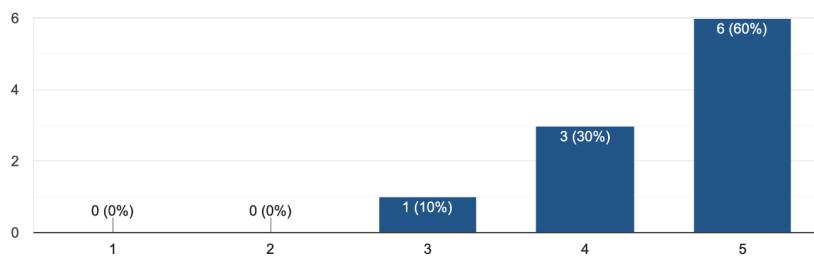
How was your experience at the NWB User Days Workshop?  
10 responses



How easy was it to learn to use the software?  
10 responses

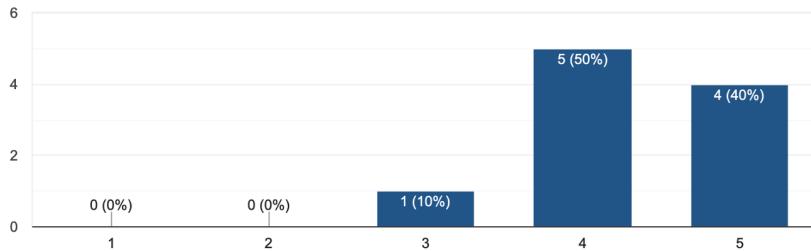


How useful was the workshop overall?  
10 responses



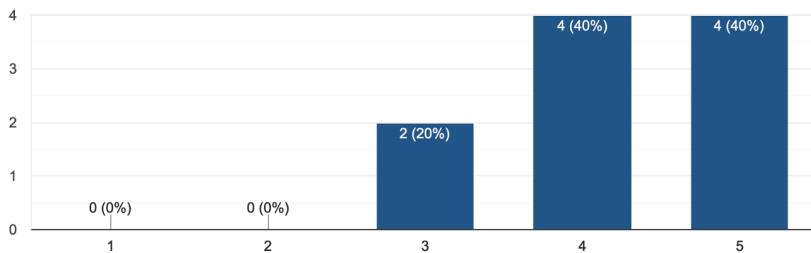
### How useful were the tutorials overall?

10 responses



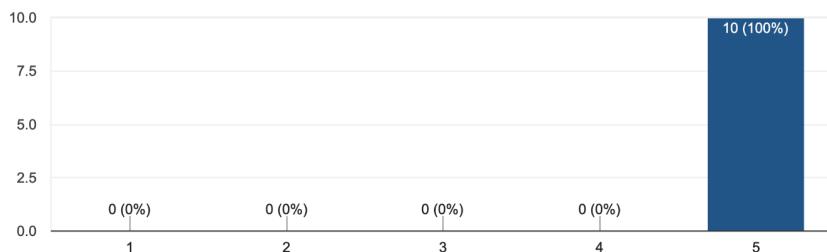
### How useful were the tool sessions? (NeuroConv, BEADL, DANDI, DataJoint, nwb-widgets)

10 responses



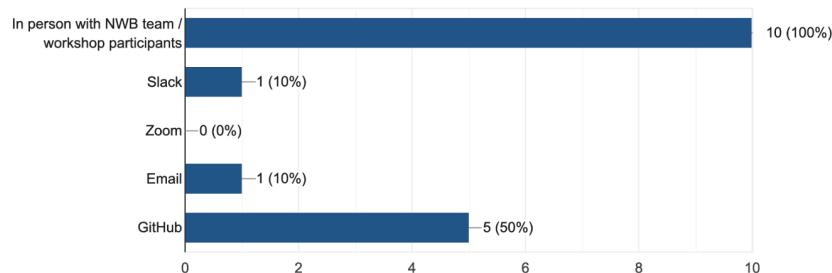
### How easy was it get help from the NWB team during the workshop?

10 responses



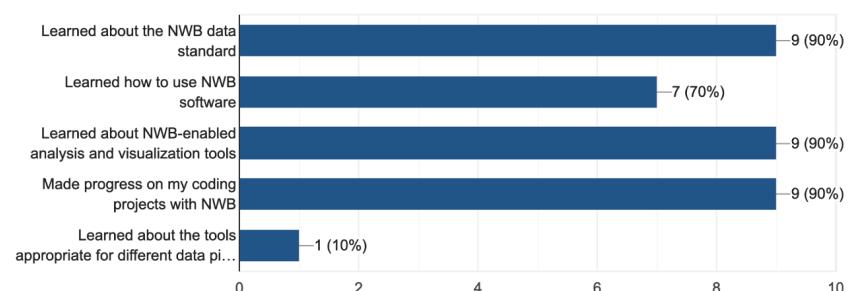
### Which communication channels were most useful for you during your hacking projects?

10 responses



### During the workshop I

10 responses



## What were the most helpful parts of the workshop?

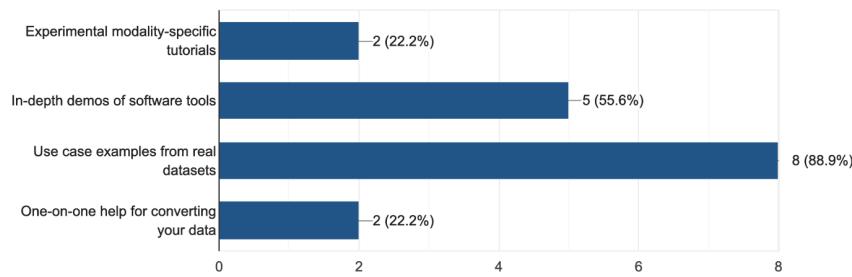
- Hacking and discussing issues
- The lecture about the topic that was used in my project.
- Ad hoc interactions with everyone
- Availability and helpfulness of the mentors!
- Running and debugging the codes with real-time, on-site help
- Direct access to the NWB team
- Being able to interact with developers and experts and learn some of the "meta-documentation", which really helped to contextualize and describe what things are where and how the whole ecosystem works together.
- Being around the other participants especially those knowledgeable about NWB was incredibly helpful. Without this I would have found it difficult to understand the full NWB ecosystem. Especially what level of detail is needed for different data pipelines. For example, I would have asked myself do I need to understand the Schema or just higher level operations such as the NWB conversion tools? Participating in the workshop allowed me to understand which tools are appropriate for different use cases. I feel it would have been difficult and overwhelming if I had just tried to learn this on my own. Additionally, being able to meet other very knowledgeable participants was incredibly helpful. I now have contacts that I can reach out to if I need help.

## Testimonials

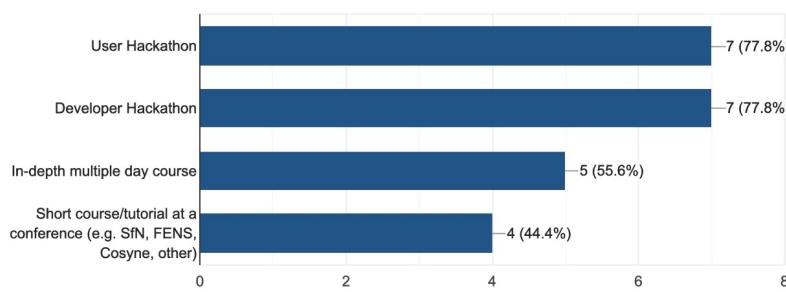
*"The NWB ecosystem is a well thought out and beautifully developed set of tools designed to standardize neuroscience data in order to facilitate transparency and collaborations. The standardization makes it easy for other scientists to understand what occurred during an experiment. This makes it easier for data sharing within a lab as well as for other collaborations and replications. I believe that the NWB program will be the standard in the field of neuroscience. The use of the hdf5 file format is brilliant allowing users to organize data and important metadata from different data streams in a logical directory like structure. Additionally, the ability to load subsets of data allows users to only load what is needed out of the inherently large neuroscience datasets. I hope that other neuroscientists invest the time to learn how to utilize the NWB tools and incorporate them into all data collection streams. I believe that this will result in better organization as well as the ability to extract as much information from the rich datasets being generated recently." Greg I. Telian PhD UC San Francisco*

## 5.2 How can we improve future training events?

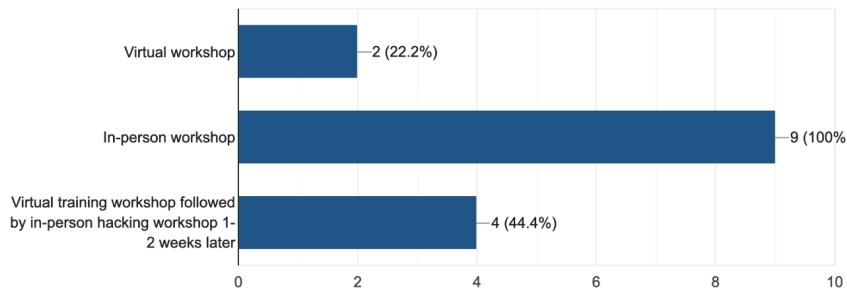
What types of sessions would you like to see in future NWB workshops?  
9 responses



What types of future training and/or development events would you be interested in participating in?  
9 responses



What type(s) of workshops would you prefer in the future?  
9 responses



### How can we do better?

- Start with a lecture on the core data types that most people use (or take a survey and guesstimate what your attendees might be using). I liked having the lecture about my topic closer to the end so I could ask pre-formed questions, but I also feel like I wasted a day being confused and in general just starting making progress later in the workshop than I would have liked. Maybe having parallel lectures about different topics with parts 1 & 2 so everyone can get an introduction and follow-up session on their specific interest.
- Maybe offer some starter notebooks to check out prior to attending to limit the time spent on becoming familiar with the basics
- No idea so far
- I would suggest a hackathon-specific slack channel for more casual announcements
- I believe making it easier for new users to understand which NWB tools best applies to them would be super helpful. At first it was intimidating as a new user to understand where I should focus my efforts. Additionally, I think PyNWB examples with small but relevant datasets would be super helpful. I believe being able to see what a typical PyNWB script looks like will give new users the ability to write

better code and understand the general process. Another challenge is understanding how a specific data stream (e.g. particular animal behavior) fits within the NWB data standard. There are examples but many of them are within examples for particular functions. I think having a page with various examples and related functions would be really helpful.

**What types of sessions would you like to see in future NWB workshops?**

- For such a short workshop, maybe daily project goals and presentations or something of the sort to encourage attendees to start early and go fast!
- Some brief technical backgrounds on different techniques. For example, show some example data from DANDI.
- A demonstration of the conversion process, featuring verbalized thought process - "this is how I approach X". And a session on how to organize multimodal files.
- I enjoyed all the sessions presented. I would like to see a more intermediate level tutorial about making an NWB file from multiple data sources. For example, present a dataset that has say Spiking data, behavioral data such as run speed, lick times, pose estimation and show how they all fit together within the NWB standard as well as which tools would be best suited. Perhaps showing how NWB conversion tools and more manual entries using PyNWB can be used together.

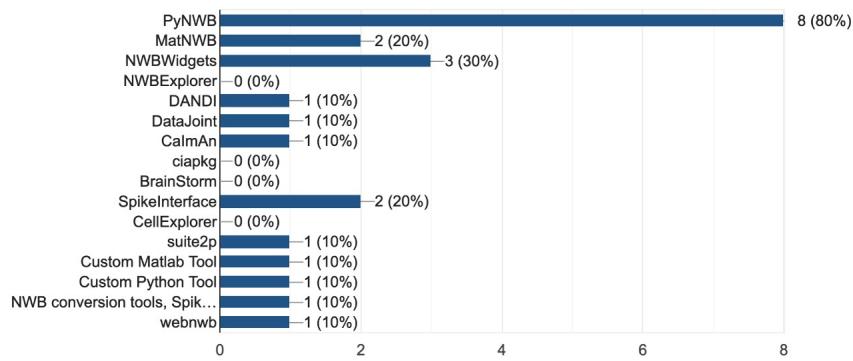
**What types of future training and/or development events would you be interested in participating in?**

- Scientific question driven data-reanalysis is really useful. 14th hackathon at Allen should be very interesting.
- I would be interested in participating in hackathons and related meet-ups that are focused on working on tools in the NWB ecosystem.
- I'm interested in additional workshops and online seminars.

### 5.3 NWB Software

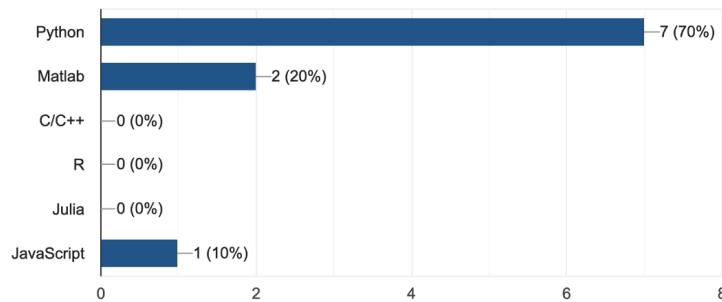
Which software tools are you using to interact with your data?

10 responses



Which programming languages are you using to interact with your data?

10 responses



What existing software (data processing/analysis/management) and data acquisition systems would you like to see integrated with NWB?

- No idea so far since I used custom software. Maybe ScanImage should be included.
- A hdf5 file explorer would be nice.

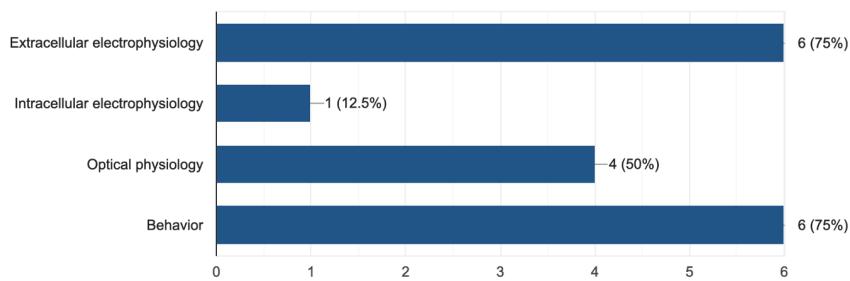
What tools would make it easier for you to adopt NWB?

- Simple demo codes with very detailed notes.

## 5.4 Your Data

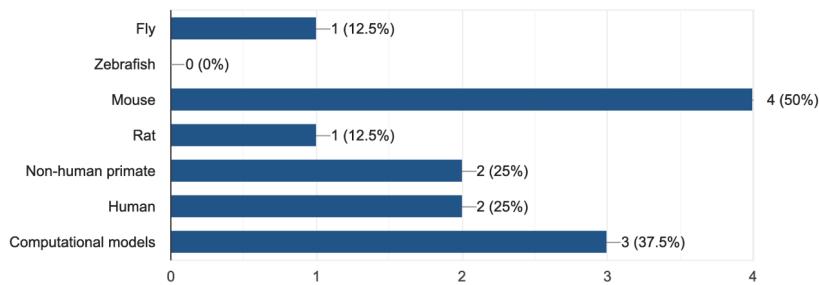
Which data modalities are you using

8 responses



Which subjects are you using in your research?

8 responses



## 6 Photos from the Event

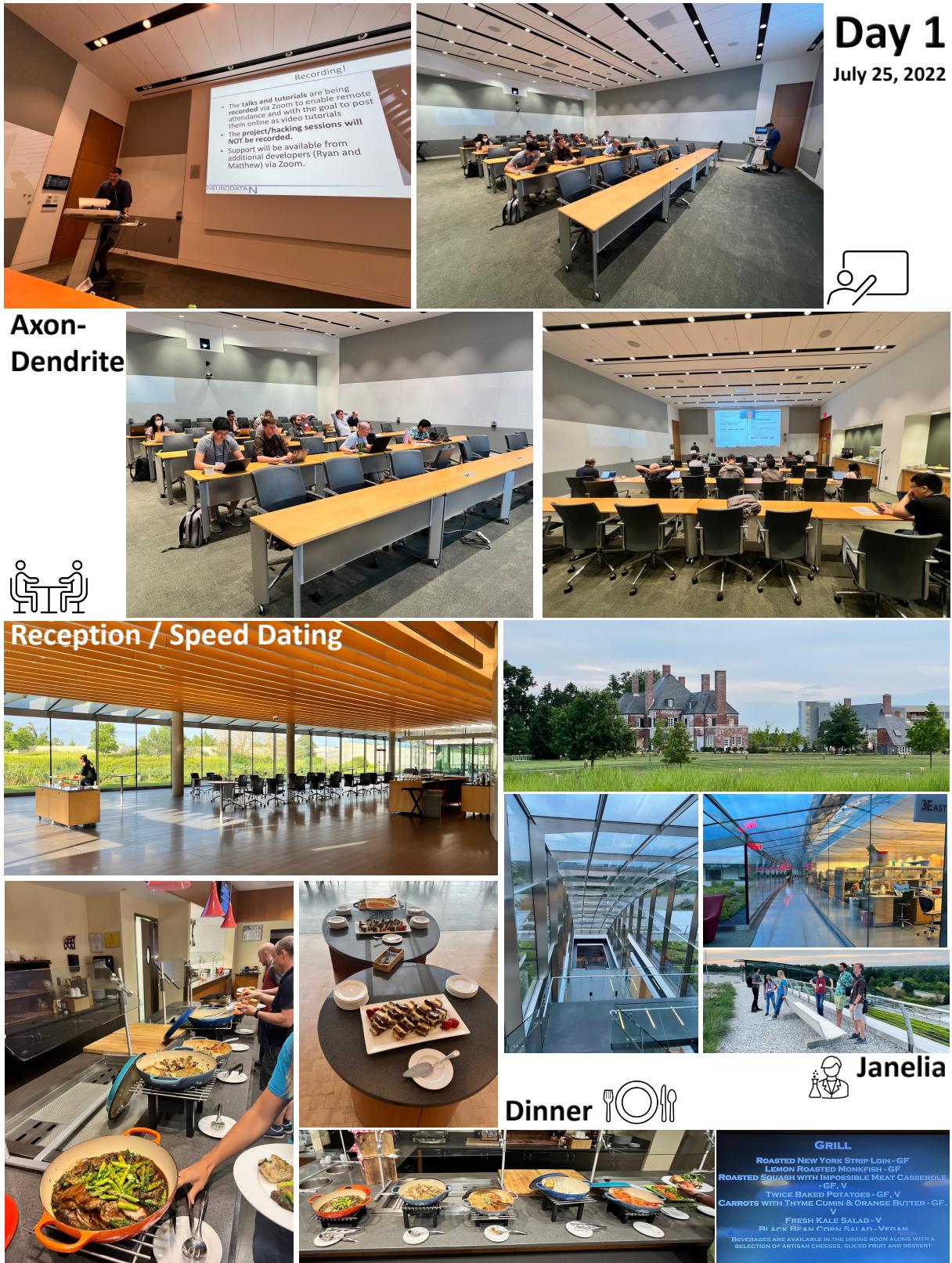


Figure 5: Photos from Day 1, July 25, 2022.

# Day 2

July 26, 2022



Figure 6: Photos from Day 2, July 26, 2022.

# Day 3

July 27, 2022



18

Figure 7: Photos from Day 3, July 27, 2022.

## Acknowledgements

We would like to thank the presenters and hosts of the breakout sessions for their contributions to the workshop (see Sec. 3). We would like to thank all participants for the great enthusiasm and for making the event a great success! Organization and operations of the event and NWB research activities were supported by the National Institute of Neurological Disorders and Stroke of the National Institutes of Health under Award Number U24NS120057 to O. Rübel (LBNL) and B. Dichter (CatalystNeuro). Housing, rooms, catering, and on-site organization was supported by HHMI Janelia. Additional travel support for participants was provided by the Kavli Foundation. Community projects on NWB are also supported by the Kavli Foundation and Simons Foundation (see also <https://www.nwb.org/projects/>).

## Disclaimer

This document and related content were prepared as an account of or to expedite work sponsored at least in part by the United States Government. While we strive to provide 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. Use of the Laboratory or University's name for endorsements is prohibited.

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. Neither Berkeley Lab nor its employees are agents of the US Government.

Links to external pages and documents included in the documents do not constitute an endorsement of the content or company and we are not responsible for the content of such links.