

Neurodata Rehack: Generating new insights from existing neurophysiology data through secondary analysis

Allen Institute, October 3-5, 2022

Benjamin Dichter, Oliver Rübel, Ryan Ly, Satrajit Ghosh, Jerome Lecoq, Saskia de Vries, Carly Kiselycznyk, Kaitlyn Casimo, and Stephanie Albin

Contents

1 Executive Summary	3
2 Participants	4
3 Program	7
3.1 Open Neurodata Showcase	7
3.2 Neurodata Rehack	7
3.3 Detailed agenda	8
4 Projects	9
5 The Kavli NeuroData Discover Award	37
6 Exit Survey	38
7 Photos from the Event	49
Acknowledgements	52
Disclaimer	52



THE
KAVLI
FOUNDATION



DANDI



NEURODATA
WITHOUT BORDERS



CATALYST
NEURO



BERKELEY LAB
Lawrence Berkeley National Laboratory

2

1 Executive Summary

Overview: NeurodataReHack was a collaboration between NWB, DANDI, Allen Institute, and The Kavli Foundation with the goal to train participants to generate new insights from existing neurophysiology data through secondary analysis. The DANDI Archive now has 100+ publicly available neurophysiology datasets, standardized using NWB, including cutting edge data produced by the OpenScope program which has been designed and packaged to enable secondary analysis. This workshop trained participants in these tools for incorporation into their process of scientific discovery. In addition, participants learned how to leverage DANDI compute resources to employ a growing ecosystem of analysis and visualization tools that work natively with NWB. Instructional activities during the hackathon taught attendees about the open neurophysiology datasets available on the DANDI Archive, how to access and analyze data in the archive, and how to use the NWB standard to incorporate existing data into their scientific workflows. The Kavli Foundation organized a Neurodata Discovery Award proposal as an exclusive opportunity for attendees to apply for funding to continue projects conceived at the event.

Participants: There were 52 applicants, 28 of whom were accepted and invited, and all who were accepted attended the event. The attendees spanned 22 different institutions, 5 different countries, and were diverse in career stage, gender, and ethnicity.

Program: The program started with lectures on day one that allowed attendees to get started with their projects: searching DANDI, leveraging the DANDI Hub, and reading NWB files. On day 2 there were three lectures from invited speakers that demonstrated the usage of specific analysis tools with broad project applicability, demonstrating their usage in the NWB/DANDI ecosystem. The rest of the day was devoted to hacking. On day two there was also a group dinner at a local restaurant. The final day was devoted to hacking and concluded with individual short project presentations.

Outcomes: Attendee projects spanned different neural recording modalities, species, and scientific questions. Impressively, by the end of the 3-day workshop, many of the attendees had performed in-depth data analysis and generated figures to demonstrate preliminary results. Some attendees also incorporated the analysis tools introduced by guest speakers. All projects were summarized by the attendees in a shared project reporting document (see Sec. 4).

Health Management: This event was held while the world is still managing the COVID-19 pandemic. Allen Institute protocols were followed, including wearing masks in the lecture room and self-reporting illness. In addition, COVID tests were made available free-of-charge to workshop attendees. Attendees were encouraged to test themselves once on Monday and once on Wednesday morning.

Conclusion: The event was very well received by the participants (see Sec. 6) and was successful in training a cohort of scientists in the reuse of publicly available, standardized neurophysiology data and in demonstrating that such reanalysis can be done with the current ecosystem of software and available data (see Sec. 6).

3

Organizing Committee:

- Program Chairs:
 - **CatalystNeuro:** Benjamin Dichter
 - **Lawrence Berkeley National Laboratory (NWB):** Oliver Rübel, Ryan Ly
 - **MIT (DANDI):** Satrajit Ghosh
 - **Allen Institute:** Jerome Lecoq (OpenScope), Saskia de Vries (Neural Dynamics)
- **Allen Institute:** Carly Kiselycznyk, Kaitlyn Casimo
- **The Kavli Foundation:** Stephanie Albin

Event Websites:

<https://alleninstitute.org/what-we-do/brain-science/events-training/2022-neurodatarehack-hackathon/>
https://neurodatawithoutborders.github.io/nwb_hackathons/HCK14_2022_Seattle/

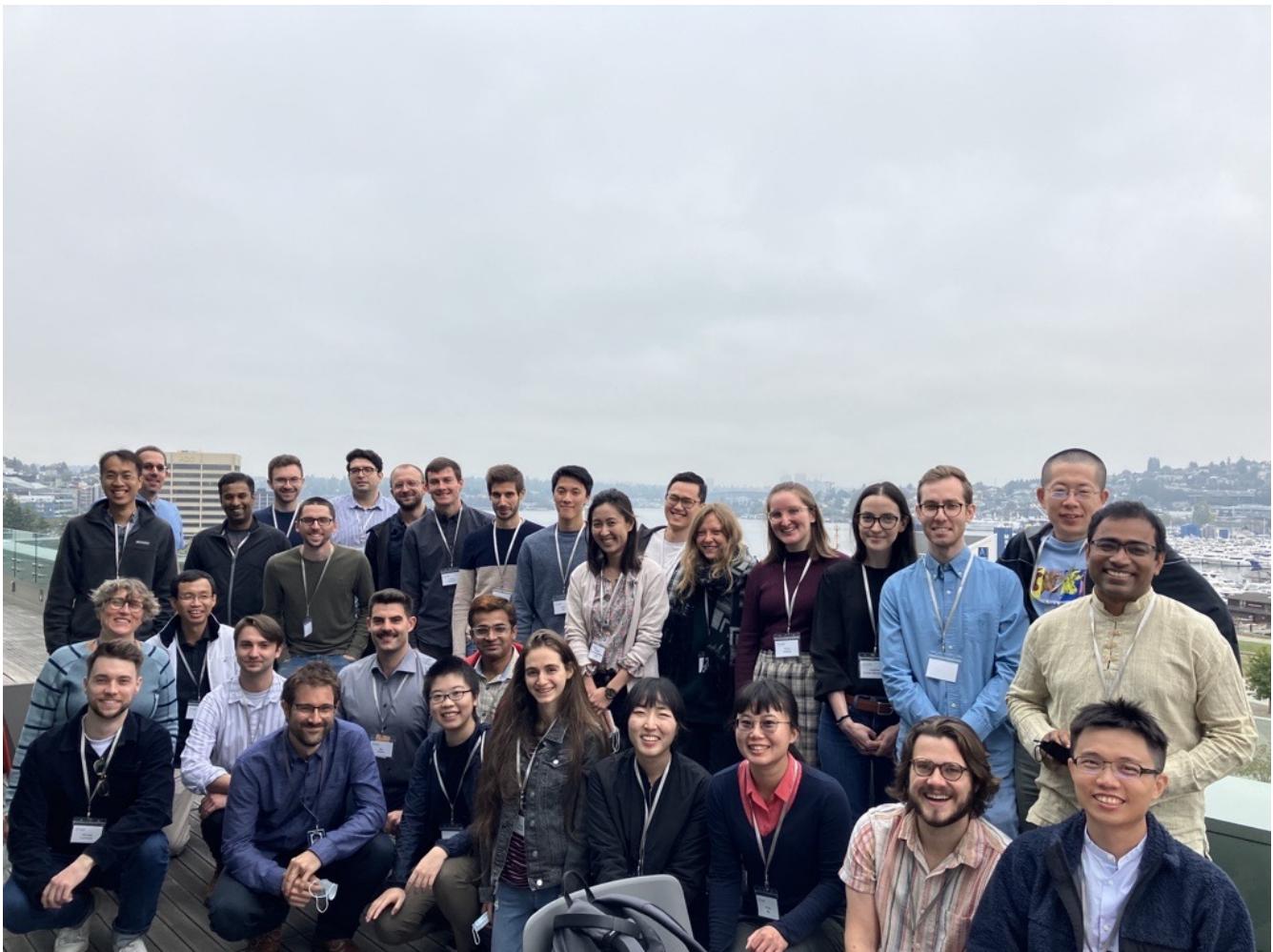


Figure 1: Participants of Neurodata Rehack 2022

2 Participants

The application process required applicants to answer questions about their background and propose projects that reuse existing data. 52 applications were received and reviewed by the entire application review committee: Benjamin Dichter, Ryan Ly, Jerome Lecoq, Satrajit Ghosh, and Saskia de Vries. Reviewers gave each of the applications a score of -1, 0, or 1 based on criteria such as appropriate background, relevance of project to the mission of the event, and feasibility of the project. The method resulted in a fairly clear cutoff, where 28 applicants received a positive score. This was quite lucky, as we were interested in admitting exactly 28 participants based on space constraints. We were also interested in potentially adjusting our admissions to maximize diversity, but we found that our accepted applicant pool was already quite diverse and no modifications were necessary.

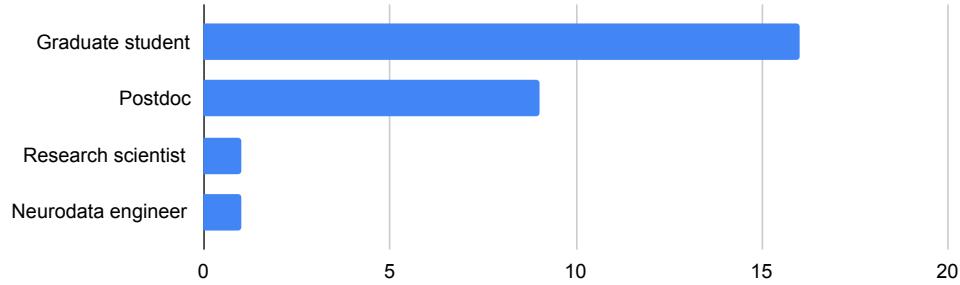


Figure 2: Current position of attendees.

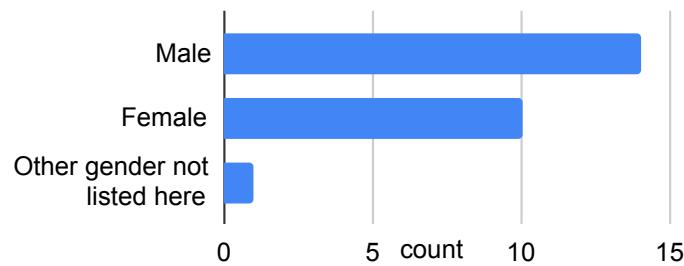


Figure 3: Gender of attendees.

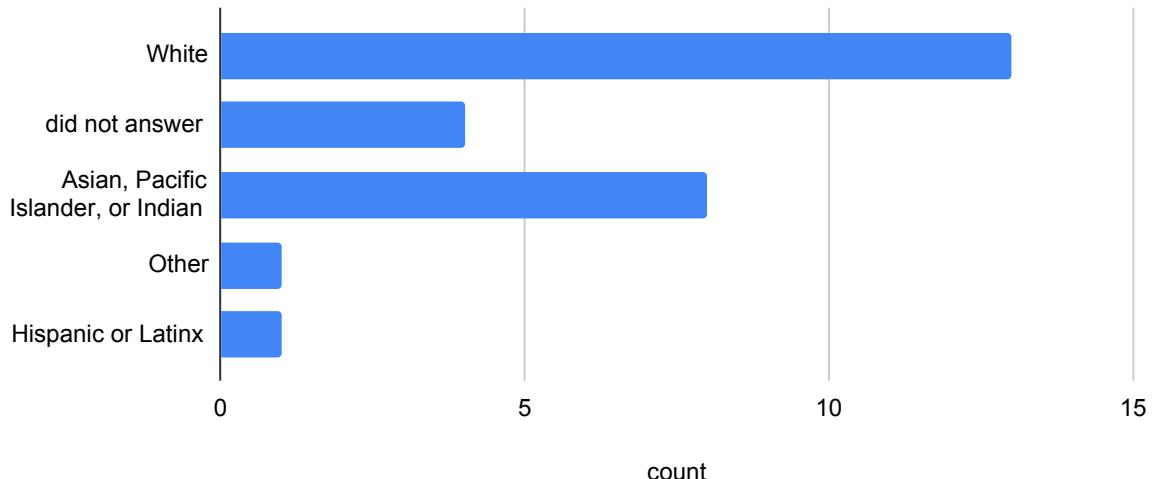


Figure 4: Ethnicity of attendees.

Country	State	Institution	Count
U.S.A.	DE	University of Delaware	1
	GA	Georgia Institute of Technology	1
	MD	The Johns Hopkins University	2
	NC	Duke University	1
	CA	The Salk Institute	1
	TN	Vanderbilt University	2
	MA	MIT	2
	CO	University of Colorado Boulder	1
	PA	St Jude Children's Research Hospital	1
	GA	Emory University	1
	CA	Stanford University	1
	CA	UC San Francisco	2
	IN	Purdue University	1
	TX	DataJoint	1
Germany		Max Planck Institute for Neurobiology of Behavior	1
Germany		Technical University of Munich	1
Germany		University of Tübingen	1
United Kingdom		The University of Edinburgh	2
Australia		The University of Sydney	1
Australia		South Australian Health and Medical Institute	1
Canada		University of Western Ontario	1
Canada		McGill University	1

Table 1: Number of participants per institution.

3 Program

3.1 Open Neurodata Showcase

The program was split into two events. The first was the Open Neurodata Showcase, a free and open virtual event that was designed to connect data contributors with neuroscientists interested in reusing their data. The program started with introductory talks for NWB, DANDI, and the Allen Institute OpenScope. This was followed by a panel discussion where experts discussed the challenges inherent in reuse of data. Next, 12 invited speakers who contributed data to DANDI gave lightning talks on their datasets. Finally, we moved to Gather to hold a virtual poster session so attendees could discuss the data with the contributors. The Open Neurodata Showcase was open to anyone; attendance was expected from Neurodata Rehack attendees. This event attracted about 60 attendees.

3.2 Neurodata Rehack

The following week, Neurodata Rehack was held at the Allen Institute. The first day consisted of talks training users in how to search and query across the DANDI Archive to find relevant datasets, how to leverage the computational resources of DANDI, and how to efficiently read NWB files in Python. The second day consisted of invited talks from three tool developers that would be of general interest for attendees and which might be useful to include in their projects. The final day was free hacking and final presentations.

3.3 Detailed agenda

DAY 0: Virtual Neurodata Showcase

Time (Pacific)	Session	Speakers
8:55-9:00am	Zoom webinar opens to participants	
9:00-9:10am	Introduction to the Open Neurodata Showcase	Ben Dichter
9:10-9:30am	Introduction to NWB: Scope, API, and tools	Oliver Ruebel
9:30-10:00am	Introduction to DANDI: Scope, web interface, downloading data, and DANDI Hub	Yaroslav Halchenko
10:00-10:30am	Introduction to OpenScope: Scope, available datasets, high-level tour of existing analysis notebooks	Jerome Lecoq
10:30-10:45am	Break	
10:45-11:30am	Roundtable discussion/panel on data reuse: pros and cons of reusing datasets, questions to ask, caveats	Bing Brunton, Mackenzie Mathis, Colleen Gillon
11:30am-12:00pm	Data blitz (1-2 minute presentations)	Data blitz presenters
12:00-1:30pm	Poster session: Held in Gather	All

DAY 1 (Monday)

Time (Pacific)	Session	Speakers
9:00-9:30	Breakfast, Meet and Greet	
9:30-10am	Introduction to event: Goals (NWB, DANDI, OpenScope), schedule, Kavli funding opportunities	Ben Dichter
10-10:30	Round table of project plans	All
10:30-11:45	Hack and discuss	
11:45-12:15	Advanced DANDI queries	Ben Dichter
12:15-1:00	Lunch	
1:00-2:00	DANDI Hub: ros3, NWB Widgets, analysis, example-notebooks repo	Satra Ghosh
2:20-3:00	OpenScope in-depth presentation of analysis notebooks	Jerome Lecoq
2:30-5:00	Hacking	

DAY 2 (Tuesday)

Time (Pacific)	Session	Speakers
9:00-9:30	Breakfast	
9:30-10:00	Check-ins	
10:00-11:00	Using SpikeInterface to process extracellular electrophysiology data in NWB	Alessio Buccino (virtual)
11:00-11:30	Hacking	
11:30-12:00	Tour of the building	
12:00-12:15	Group Photo	
12:15-1:00	Lunch	
1:00-2:00	Using SSM to perform bayesian inference on neural data in NWB	David Zoltowski
2:00-3:00	Modeling neural population dynamics and evaluating model fit with LFADS and NLB	Felix Pei
3:00-5:00	Hacking	
6:00-8:30	Meet in EVEN Hotel lobby for dinner at Cactus	

DAY 3 (Wednesday)

Time (Pacific)	Session	Speakers
9:00-9:30	Breakfast	
9:30-12	Hacking	
12:00-1:00	Lunch	
1:00-2:40	Hacking	All
2:40-3:00	NIH funding opportunities	Sandeep Kishore
3:00-4:00	5-minute presentations	All
4:00-5:00	Feedback for organizers roundtable	

4 Projects

Participants generally pursued independent projects, though a few did form teams. The projects spanned many different source datasets, and were diverse in scientific questions and analytic approaches. Some even incorporated multiple datasets. Several projects collaborated with David Zoltowski to incorporate the State Space Models (SSM) library that was presented as part of one of the tutorial talks during the event, which models neural data in latent variable generative models and allows for understanding underlying states of neural data. Amazingly, the final presentations revealed that many projects were able to achieve results within the limited timeframe of the event. Most presentations included figures designed to address specific scientific questions generated from secondary source data.

Participants of each project documented their own projects and progress in the shared [Project GoogleDoc](#) included below:

NeuroDataReHack Projects

Table of Contents

Project Title (Template)	5
Key Investigators	5
Project Description	5
Approach and Plan	5
Progress and Next Steps	5
Data	5
Materials	5
Background and References	5
Cross-species, cross-modality survey of bursting neurons	6
Decoding history dependent neural activity in primary and higher visual areas	7
Key Investigators	7
Project Description	7
Approach and Plan	7
Progress and Next Steps	7
Data	7
Materials	7
Background and References	7
Cross-Lab and Cross-Species co-clustering of cortical intracellular patch-clamp data	8
Key Investigators	8
Project Description	8
Approach and Plan	8
Progress and Next Steps	8
Data	8
Materials	8
Background and References	8
Using DANDI open datasets in transfer learning for decoding proprioception from neuronal calcium image using artificial neural network	9
Key Investigators	9
Project Description	9
Approach and Plan	9
Progress and Next Steps	9
Data	10
Materials	10
Background and References	10
Validating latent variable models in dandisets with rich behavioral data	11
Key Investigators	11
Project Description	11

10

Approach and Plan	11
Progress and Next Steps	11
Data	11
Materials	11
Background and References	12
Key Investigators	15
Project Description	15
Approach and Plan	15
Progress and Next Steps	15
Data	15
Materials	15
Background and References	15
Finding neuronal networks based on shared temporal activity across task conditions	16
Key Investigators	16
Project Description	16
Approach and Plan	16
Progress and Next Steps	16
Data	16
Materials	16
Background and References	17
Background and References	19
Exploring the network of excitatory neurons with regularized GLMs	20
Key Investigators	20
Project Description	20
Approach and Plan	20
Progress and Next Steps	20
Data: Dandiset 000041	20
Materials:	20
https://github.com/junipertcy/NeuroDataReHack	20
Background and References	20
Elucidating Neural Learning Rules from Calcium Imaging Data	21
Key Investigators	21
Project Description	21
Approach and Plan	21
Progress and Next Steps	21
Data	21
Materials	21
Background and References	21
Fitting state space models in SSM repo to NWB datasets	22
Key Investigators	22
Project Description	22
Approach and Plan	22
	2

Progress and Next Steps	22
Data	22
Materials	22
Background and References	22
Constructing time-probability-matrices from simultaneously recorded visual areas using HMMs	23
Key Investigators	23
Project Description	23
Approach and Plan	23
Progress and Next Steps	23
Data	23
Materials	23
Background and References	23
Key Investigators	24
Project Description	24
Approach and Plan	24
Progress and Next Steps	24
Data	24
Materials	24
Background and References	24
Spyglass: data analysis framework for reproducible neuroscience research	25
Key Investigators	25
Project Description	25
Approach and Plan	25
Progress and Next Steps	25
Data	25
Materials	25
Background and References	25
Contrast sensitivity across visual cortex	26
Key Investigators	26
Project Description	26
Approach and Plan	26
Progress and Next Steps	26
Data	26
Materials	26
Background and References	26
Ensemble computational models of neurons individually tuned with multimodal electrophysiology data	27
Probing excitatory computations underlying probabilistic learning	28
Key Investigators	28
Project Description	28
Approach and Plan	28
Progress and Next Steps	28

Data	28
Materials	28
Background and References	28

Cross-species, cross-modality survey of bursting neurons

Key Investigators

- Chenggang Chen, Johns Hopkins University

Project Description

- Comparing the proportion of burst spiking neurons in the sensory cortex of non-human primates (macaque and marmoset) and rodents (gerbil and mouse).
- Comparing the auditory and visual tuning properties of burst spiking neurons (BU), non-burst fast-spiking neurons (FS), and non-burst regular spiking neurons (RS)

Approach and Plan

- Extract the spike waveform and spike train from raw data, then use that information to classify all the units into three types: BU, FS, and RS. Characterize their tuning properties and running modulations.

Progress and Next Steps

- Download datasets of various species from different archives
- Change the non-NWB format dataset to NWB format
- Extract the spike train from four NWB datasets using the same code
- Classify the neuronal population into three groups based on firing pattern and spike waveform
- Compute the auditory and visual tuning properties of three types of neurons
- In the future, I will further compare the difference among three types of neurons

Data

- Change to NWB format; lab data
- Change to NWB format; <https://crcns.org/data-sets/vc/pvc-5/about>
- Change to NWB format; https://gin.g-node.org/dianamaro/Amaro_et_al_2021_CurrBiol
- Already in NWB format; <https://dandiarchive.org/dandiset/000021>
- Already in NWB format; <https://dandiarchive.org/dandiset/000022>

14

Materials

- See below 'Background and References' for details

Background and References

- Mouse visual cortex, <https://www.nature.com/articles/s41586-020-03171-x>
- Macaque visual cortex, <https://www.sciencedirect.com/science/article/pii/S0042698914000200>
- Gerbil auditory cortex, <https://www.sciencedirect.com/science/article/pii/S0960982221008204>
- Marmoset auditory cortex, <https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3001642>
- Marmoset auditory cortex, <https://academic.oup.com/cercor/article/29/3/1199/4840634>

Decoding history dependent neural activity in primary and higher visual areas

Key Investigators

- Lan Luo, Duke University
Connect with me: <https://www.linkedin.com/in/lan-luo-q42/>

Project Description

- Investigate how does visual adaptation (history dependency of visual signals in the brain) transform the encoding of stimulus identity using Allen Institute Visual Behavior 2-Photon Imaging & Neuropixels recordings dataset

Approach and Plan

- Visual Behavior 2-Photon Imaging data viz with dimensionality reduction using different neural subpopulations
- Building linear and nonlinear decoder to decode visual input identity from neural activity
- Explore Visual Behavior Neuropixels data

Progress and Next Steps

- Decoding history dependent neural activity in visual areas

Data

- <https://portal.brain-map.org/explore/circuits/visual-behavior-2p>
- <http://portal.brain-map.org/explore/circuits/visual-behavior-neuropixels>

Materials

- <https://github.com/lanluo9/inter/blob/4cfdf89c713439b94803b5e078b1dff518a8834/results/poster/poster%20neurobio%20retreat%202021.pdf>
- <https://github.com/lanluo9/inter>

Background and References

Cross-Lab and Cross-Species co-clustering of cortical intracellular patch-clamp data

Key Investigators

- Sam Mestern, UWO

Project Description

With this project, we aim to derive novel insights regarding cortical neuron differences between species and recording conditions. We are also aiming to demonstrate the usefulness of a novel computational method for integrating electrophysiological data

Approach and Plan

- Extract intracellular single neuron features (using IPFX) from several datasets bridging across labs and species.
- From here, we will apply novel computational methods to integrate the datasets and facilitate co-clustering of similar species

Progress and Next Steps

- Downloaded several intracellular datasets from dandihub
- Extracted overall features from each dataset using IPFX's run_feature_collection
- Co-cluster datasets using extracted features

Data

- Tolias Patch-seq - <https://dandiarchive.org/dandiset/000008>
- AI Patch-seq - <https://dandiarchive.org/dandiset/000020>
- AI Patch-seq in human - 000023, 000228, 000142, 000209, 000288, 000109

Materials

- <https://github.com/AllenInstitute/ipfx>

Background and References

1. <https://www.nature.com/articles/s41586-020-2907-3>
2. [https://www.cell.com/cell/pdf/S0092-8674\(20\)31254-X.pdf](https://www.cell.com/cell/pdf/S0092-8674(20)31254-X.pdf)

Using DANDI open datasets in transfer learning for decoding proprioception from neuronal calcium image using artificial neural network

Key Investigators

- Seungbin Park

Project Description

Decoding proprioception is necessary for proprioceptive feedback to brain-machine interface to improve its movement performance. Artificial neural network is expected to be advantageous in revealing complex encoded proprioception from the two-photon calcium image. However, building massive datasets of the two-photon image and behavior recording is extremely challenging because animal experiments and behavior training require much time, effort, and sophisticated techniques. Moreover, it inevitably accompanies sacrificing numerous animals. Using DANDI open datasets can be a good solution for these problems. Transfer learning refers to the methodology to create high-performance learners using datasets from different domains that can be obtained more easily [1]. DANDI archive has already built various high-quality datasets so they can be exploited for transfer learning. I aim to use the dataset titled as 'A map of anticipatory activity in mouse motor cortex (DANDI ID: 000015). It includes two-photon images of the population activity of neurons related to behavior across a wide range of motor cortex [2]. The dataset is expected to be appropriate for the purpose in that proprioception and anticipatory timing are highly correlated [3]. The main goal is to improve the performance of the neural network trained with my own datasets of mouse limb positions and fluorescence traces extracted from two-photon images through transfer learning using the DANDI open datasets.

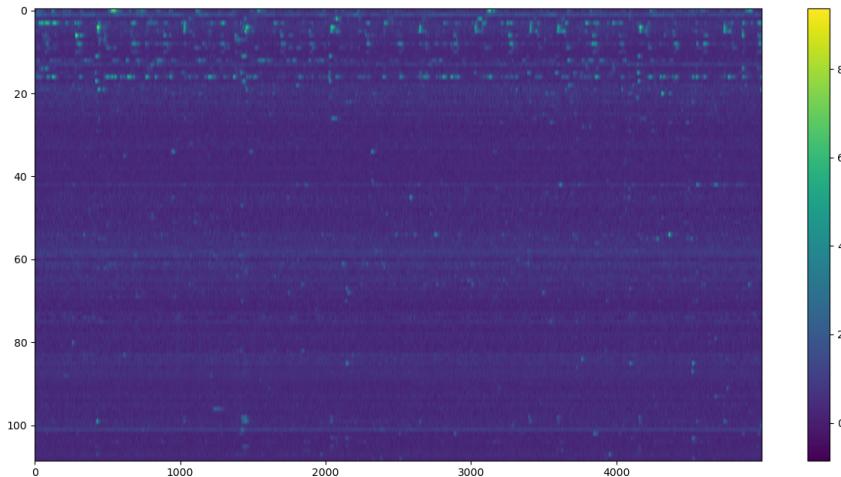
Approach and Plan

- 1. Load and explore the dataset.
- 2. Preprocess the DANDI dataset for training a neural network.
- 3. Train a neural network with the preprocessed DANDI dataset.
- 4. Train a neural network with my own dataset using pre-trained parameters from step #3.
- 5. Evaluate the performance.

17

Progress and Next Steps

- I am in step #1. I plan to follow the following steps in Approach and Plan section.



Data

- A Map of Anticipatory Activity in Mouse Motor Cortex (DANDI ID: 000015)
<https://dandiarchive.org/dandiset/000015?search=motor%20anticipatory&pos=1>

Materials

- DANDI example notebooks: <https://github.com/dandi/example-notebooks>
- PyNWB documentation: <https://pynwb.readthedocs.io/en/stable/>
- Suite2p documentation: <https://suite2p.readthedocs.io/en/latest/>
- Deeplabcut documentation: <http://www.mackenziemathislab.org/deeplabcut>

Background and References

3. [1] Weiss, Karl, Taghi M. Khoshgoftaar, and DingDing Wang. "A survey of transfer learning." *Journal of Big data* 3.1 (2016): 1-40.
4. [2] Chen, Tsai-Wen, et al. "A map of anticipatory activity in mouse motor cortex." *Neuron* 94.4 (2017): 866-879.
5. [3] Christina, Robert W. "Proprioception as a basis of anticipatory timing behavior." *Motor Control*. Academic Press, 1976. 187-199.

Validating latent variable models in dandisets with rich behavioral data

Key Investigators

Ben Lansdell, St Jude

Project Description

Tracking behavioral data from freely behaving animals, possibly in conjunction with neural recordings, is an exciting and growing direction in neuroscience. These datasets present rich opportunities for discovery, particularly by enabling the study of naturalistic behavior over long time spans. Unsupervised or semi-supervised learning methods that characterize neural activity and/or behavior are useful for summarizing and studying such datasets. I'm interested in understanding these methods, in datasets that have an interesting behavioral component, in addition to neural recordings. The first goal is to understand and implement in my own libraries the ability to read pose-tracking data from NWB datasets. The second goal is to use dandisets to investigate the utility of novel unsupervised/semi-supervised learning methods.

Approach and Plan

First goal: make NWB datasets, with ndx-pose data, readable in python library Ethome.

Second goal: recently there have been a number of interesting latent variable models, using deep generative models, that characterize neural activity, and that simultaneously model the relation between neural activity and behavior or task variables. One example is pi-VAE [1]. The authors claim this provides a more nuanced, yet still interpretable, characterization of the data, and could serve as an alternative to some of the standard methods in computational neuroscience. Is this true? In their AJILE12 dataset, Peterson et al [2] develops multiple linear regression to characterize the neural activity and its relation to behavioral and task related variables, using it to say which factors are most often encoded by activity in ECoG arrays. I plan to test this latent variable method, to see if the important variables revealed by this latent space analysis are the same as those revealed by the linear model.

Progress and Next Steps

19

Read behavior and ecog data from AJILE12. Next steps: format data into format expected by pi-VAE (list of spike data matrices for each trial/behavior), design similar analysis to Fig 6 of AJILE12 paper (Peterson et al 2021 [2]): recompute goodness of fit measures with/without different behavioral variables to judge the degree to which they're encoded in the data.

Data

Dandisets 55 and 231.

Materials

Behavior analysis code: <https://github.com/benlansdell/ethome>

Background and References

- [1] Zhou and Wei 2020. Learning identifiable and interpretable latent models of high-dimensional neural activity using pi-VAE. <https://arxiv.org/pdf/2011.04798.pdf>
- [2] Peterson et al. 2021. Behavioral and Neural Variability of Naturalistic Arm Movements.
<https://www.eneuro.org/content/8/3/ENEURO.0007-21.2021.long#sec-18>

Sub-seconds Neural Emotion-Coding in Amygdala under Diverse Hippocampal Theta-Gamma States

Key Investigators

- Lu Zhang, Georgia Institute of Technology

Project Description

Leveraging my methods to capture hippocampal theta-gamma coupling states (Zhang et al., 2019), I found that “non-place” cells, traditionally being ignored, played a role in discriminating goals during spatial navigation (Zhang, et al, 2022). Following the application of my method within hippocampal circuit above, I will further investigate how hippocampal (HPC) theta-gamma states (TG states) affect neural emotion-coding in amygdala (AMY), a brain region highly interacting with hippocampus during emotion memory consolidation.

Approach and Plan

- Step 1: Categorizing HPC theta oscillations into slow-gamma, medium gamma and fast-gamma states using my previous computational methods integrating Morse wavelet and k means clustering.
- Step 2: Test whether the activities of amygdala cells are different across diverse HPC at both single unit (Firing rate) and population level (Bayesian decoding).
- Step 3: Test whether HPC-AMY interactions differs during different theta-gamma states, at LFP-LFP, unit-LFP, and unit-unit level.
- Step 4: Test whether the results in above steps varies across different behavioral states (learning period, REM sleep, and after memory consolidation)

Progress and Next Steps

- Downloading the data-set.
- Working on.

Data

- <https://dandisarchive.org/dandiset/000061/>
- Or <https://crcns.org/data-sets/hc/hc-14>

Materials

- Video: <https://jrnclub.org/research-films/sub-second-dynamics-theta-gamma-coupling>
- Code: <https://github.com/singerlabqt/IndividualThetaCluster>

Background and References

- **Background:** Oscillatory activity is often characterized based only on its frequency content, and interactions or nesting of one faster oscillation in slower such as gamma (30–150 Hz) nested in theta (6–12 Hz) in the hippocampus (HPC) (Buzsáki and Draguhn, 2004). However, current methods to

assess cross-frequency coupling averaging neural signals over long consecutive time periods, which obscure cycle-by-cycle sub-second dynamics that underlie cognitive computations (Kopell et al., 2014). To address that, I developed novel computational approaches combining signal processing and machine learning, to capture moment-to-moment changes in hippocampal theta-gamma coupling in rodents at single theta cycle timescale (Zhang et al., 2019). My methods provide new approaches to investigate the neural code in hippocampus or hippocampal interactions with other regions in spatial navigation, memory, and their alternations in aging and brain diseases (Zhang, et al, 2022). I plan to extend the application of my method to other brain region interacting with hippocampus, such as amygdala in this project.

- **References**

- Zhang, L., Prince, S.M., Paulson, A.L., Singer, A.C. (2022). [Goal discrimination in hippocampal non-place cells when place information is ambiguous](#). Proc. Natl. Acad. Sci. 119 (11), e2107337119.
- Zhang, L., Lee, J., Rozell, C., and Singer, A.C. (2019). [Sub-second dynamics of theta-gamma coupling in hippocampal CA1](#). Elife 8.
- Kopell, N.J., Gritton, H.J., Whittington, M.A., and Kramer, M.A. (2014). [Beyond the connectome: The dynome](#). Neuron 83, 1319–1328.
- Buzsaki, G. and Draguhn, A. (2004). [Neuronal oscillations in cortical networks](#). Science 304, 1926-1929.

Feature-based embeddings of non-stationary ECoG

Key Investigators

- Brendan Harris, USyd

Project Description

Describe the spatio-temporal dynamical structure of an EcoG recording using an existing pipeline for summarizing non-stationary neural data in a low-dimensional space of time-series features.

Approach and Plan

- Load the dataset into Julia (electrode traces and annotated joint positions) from the NWB file. Begin developing packages that wrap the DANDI and pynwb tools in Julia.
- Feed the dataset into the existing pipeline, then visualize the results (showing per-channel ECoG transitioning between dynamical regimes, and characterize the salient dynamical properties of each regime).

Progress and Next Steps

- Data are downloaded via the DANDI cli, and loaded into Julia. Automated downloads are not yet implemented.
- Identify a subject and time interval from the full dataset for a pilot analysis.
- Visualize the non-stationary feature-based embedding of the test data (e.g. animate the regions of feature space occupied by each EcoG channel against behavioral data such as the joint position or event labels, over time).

Data

<https://dandiarchive.org/dandiset/000055>

Materials

- <https://github.com/brendanjohnharris/Catch22.jl>
- <https://github.com/brendanjohnharris/ParameterInference.jl>

Background and References

- C. H. Lubba, S. S. Sethi, P. Knaute, S. R. Schultz, B. D. Fulcher, and N. S. Jones, “catch22: CAnonical Time-series CCharacteristics,” *Data Mining and Knowledge Discovery*, vol. 33, Art. no. 6, 2019.
- S. Güttler, H. Kantz, and E. Olbrich, “Reconstruction of the parameter spaces of dynamical systems,” *Physical Review E*, Art. no. 5, 2001.

Finding neuronal networks based on shared temporal activity across task conditions

Key Investigators

- Noga Mudrik (JHU)

Project Description

I propose a framework for finding interpretable functional neuronal networks presented across various non-stationary task conditions, and for assessing neural encoding uncertainty in the networks latent space. I choose to take inspiration from the GraFT algorithm [2] and to develop a multi-dimensional version of it for data from Poisson distribution. This method thus uses neural activity data recorded under different task conditions and finds a low dimensional meaningful representation of the neural activity in each trial, described by matrix factorization where one matrix describe the neuronal networks components and the other refer to their temporal activity. Each of these temporal activities' matrices can be viewed as a trajectory in the networks' space and trajectories associated with different trials of the same condition can jointly form a manifold. These network-space manifolds will then be further analyzed to assess the encoding uncertainty by considering the manifolds' internal structure and separability. In addition to identifying neuronal circuits, testing their activities under varied settings, and using them for uncertainty estimation, this data-driven graph-based framework will also be used to improve understanding of — 1) how different task components are encoded in the brain; 2) the role of inter-regional brain connectivity versus local brain oscillations in driving behavior; 3) abnormal brain activity under pathological or stress conditions; 4) neural variability between and within conditions; and 5) behavior robustness to neural damage.

Approach and Plan

- Find the most appropriate dataset for this task
- Data pre-processing and creating the tensors.
- Continue working on the python code for the multidimensional GraFT
- After finding the neuronal networks - study how their composition and activity differ over task condition and study the network level dynamics to assess encoding uncertainty.

Progress and Next Steps

- Start writing the multi dimensional graft code
- Choosing and downloading the initial dataset
- Data pre-processing
- Next step: apply the data to the multi-dimensional graft

Data

- <https://dandiarchive.org/dandiset/000127>
- (for future steps: <https://dandiarchive.org/dandiset/000028?search=NEUROPIXELS&pos=1>)

Materials

- Currently my code is in a private github repository (when I finish I will change it to public). The first version of the code (the python implementation to the graft method) is described and can be downloaded from here -
<https://pypi.org/project/GraFT-Python/>
-

Background and References

1. Raeed H Chowdhury, Joshua I Glaser, Lee E Miller (2020) Area 2 of primary somatosensory cortex encodes kinematics of the whole arm eLife 9:e48198. <https://doi.org/10.7554/eLife.48198>
2. A. S. Charles, et al., "GraFT: Graph Filtered Temporal Dictionary Learning for Functional Neural Imaging," bioRxiv, p. 2021.05.24.445514, May 2021, doi: 10.1101/2021.05.24.445514.

Human iPSC-derived neurons recapitulate phenotypic variation within the Human cortex

Key Investigators

- Michael Zabolocki, SAHMRI

Project Description

Human induced pluripotent stem cells (hiPSCs) offer a model which has the capacity to recapitulate the genetic underlying of the human brain across early neurodevelopment. Of these, organoid and monolayer models have repeatedly demonstrated that genetic signatures unique to human pre- and postnatal transitions can be recapitulated in-vitro. However, whether the functionality of these neurons compare to the human brain itself is unknown at the single-cell level. To address this, cortical hiPSC-derived neurons were then patch-clamped at late-stage time points (> 70 days) and compared to Allen Brain Institutes' human cortical brain biopsy acute brain slice patch-clamp database. We reveal that subpopulations of hiPSC derived neurons in organoid or monolayer models emerge to share similar functional properties with the Human cortex, independent of morphological differences. Following on from previous work, integrating additional Human cortical patch-clamp datasets available on Dandi will expand the significance of such findings. Within said subpopulations, patch-seq data available across both Dandi and Allen Brain datasets will be isolated to determine genetic differences in ion channel gene expressions.

Approach and Plan

- Load raw intracellular EPhys recordings (both voltage and current clamp recordings), collected from all Dandisets containing Human cortical intracellular EPhys.
- Intracellular Ephys feature extraction using a custom Python pipeline for voltage clamp and current clamp recordings.
- Quantify overlapping features between layers and regions with iPSC datasets, and calculate similarity scores
- Visual outputs and quantified metrics between subpopulations
- Extract ion channel expression data using available patch-seq data from 'similar' Human isolated subpopulations

Materials

https://github.com/mzabolocki/humanbiopsy_ipsc_ophys

<https://github.com/mzabolocki/BrainSpike>

Progress and Next Steps

- Data is downloaded via the DANDI, raw recordings extracted for all relevant Human cortical datasets. Downloads currently set to local.
- Identify relevant stimulus types (voltage clamp, current clamp long depolarisation, ramp, sag protocols) in the DANDI data.
- Compare features from Dandisets to existing.

Data

- Dandisets 000293 and 000297

Background and References

[https://www.cell.com/cell-stem-cell/pdfExtended/S1934-5909\(19\)30337-6](https://www.cell.com/cell-stem-cell/pdfExtended/S1934-5909(19)30337-6)

<https://www.nature.com/articles/s41593-021-00802-y>

<https://www.nature.com/articles/s41467-022-32115-4>

<https://www.nature.com/articles/s41593-021-00906-5>

Exploring the network of excitatory neurons with regularized GLMs

Key Investigators

- Tzu-Chi
- Yi-Yun
- Josefina

Project Description

This project explored statistical methods for extracting network structure from point process observations.

Approach and Plan

- We analyzed spike trains from excitatory cortical neurons under REM sleep and model each spike train as a linear combination of the past activity of other neurons, up to 1 time bin in the past.

Progress and Next Steps

- Progress made:
 - Explored methods to infer functional connectivity between neurons based on firing history
 - Built a graphy display of network from the connectivity
- Future plan:
 - Assess goodness of fit
 - Explore non-linear models
 - Compare network structure across different states, awake, non-REM, REM, and across brain regions

Data: Dandiset 000041

Materials:

<https://github.com/junipercty/NeuroDataReHack>

<https://docs.google.com/presentation/d/1Y6b-s2QMO0UFuWa6hob9OCrinwIjyVw-Bg6OqtcGuKI/edit?usp=sharing>

28

Background and References

Watson, B.O., Levenstein, D., Greene, J.P., Gelinas, J.N. and Buzsáki, G., 2016. Network homeostasis and state dynamics of neocortical sleep. *Neuron*, 90(4), pp.839-852.

Truccolo, W., Eden, U.T., Fellows, M.R., Donoghue, J.P. and Brown, E.N., 2005. A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *Journal of neurophysiology*, 93(2), pp.1074-1089.

Elucidating Neural Learning Rules from Calcium Imaging Data

Key Investigators

- Felix Pei
- Alessandro Salatiello

Project Description

We want to analyze data from a mouse learning a task and see if any insight can be found regarding the learning rules used by the brain

Approach and Plan

- Download NWB data and gain familiarity with format
- Analyze changes in neural activity during training
- Compare with computational models using different learning rules.

Progress and Next Steps

- Continue improving computational models to better fit data and cover wider range of learning rules
- Extend calcium imaging analysis to single-trial

Data

- <https://dandiarchive.org/dandiset/000016/>

Materials

- <https://github.com/felixp8/neurodatarehack-2022/>

Background and References

6. Najafi F, Elsayed GF, Cao R, Pnevmatikakis E, Latham PE, Cunningham JP, Churchland AK. Excitatory and Inhibitory Subnetworks Are Equally Selective during Decision-Making and Emerge Simultaneously during Learning. *Neuron*. 2020 Jan 8;105(1):165-179.e8. doi: 10.1016/j.neuron.2019.09.045.
7. Murray JM. Local online learning in recurrent networks with random feedback. *eLife*. 2019. 8:e43299.

Fitting state space models in SSM repo to NWB datasets

Key Investigators

- David Zoltowski

Project Description

My goal is to fit state space models in the SSM code package to NWB datasets. I looked at two different human recording datasets.

Approach and Plan

I processed the data into trials and fit HMM / LDS / rSLDS models to the trial data.

Progress and Next Steps

I analyzed recordings from human PPC while a tetraplegic participant generated neural activity corresponded to attempted finger movements. I found some differences in dynamics across the different finger movements.

Data

- <https://dandiarchive.org/dandiset/000147?pos=8>

Materials

- <https://github.com/lindermanlab/ssm>

Background and References

8.

Constructing time-probability-matrices from simultaneously recorded visual areas using state space models

Key Investigators

- Brock Carlson
- Blake Mitchell
- David Zoltowski

Project Description

We would like to calculate Phi

Approach and Plan

- Describe the steps of your planned approach to reach the objectives.

Progress and Next Steps

- Describe the progress you have made on the project, e.g., which objectives you have achieved and how.
- Describe the next steps you are planning to take to complete the project.

Data

- Links to the dandiset(s) that you are using.

Materials

- Links to materials relevant to the project, e.g., code, videos.

Background and References

9. Use this space for information that may help people better understand your project, e.g., links to papers.

Project Title: Identification of theta states in the prefrontal cortex

Key Investigators

- John Stout

Project Description

Neural activity, paced at theta frequency (4-12Hz) in the medial prefrontal cortex, is linked to memory-guided decision making. I would like to identify when theta oscillations are present, then use those epochs to build models that predict behavioral/cognitive/neural states.

Approach and Plan

- First, identify approaches to extract “high” theta power states using extracellular recordings

Progress and Next Steps

Have spent a lot of time troubleshooting, but some progress has been made. Working on using a linear regression over log-transformed power spectra, then using a metric, like mean squared error, to identify when theta oscillations might deviate from the “typical” cortical oscillations based on the 1/f law

Data

[DANDI Archive](#) (000041)

Materials

Not ready yet!

Background and References

1. O'Neill, P. K., Gordon, J. A., & Sigurdsson, T. (2013). Theta oscillations in the medial prefrontal cortex are modulated by spatial working memory and synchronize with the hippocampus through its ventral subregion. *Journal of Neuroscience*, 33(35), 14211-14224.
2. Jones, M. W., & Wilson, M. A. (2005). Theta rhythms coordinate hippocampal–prefrontal interactions in a spatial memory task. *PLoS biology*, 3(12), e402.
3. Hallock, H. L., Wang, A., & Griffin, A. L. (2016). Ventral midline thalamus is critical for hippocampal–prefrontal synchrony and spatial working memory. *Journal of Neuroscience*, 36(32), 8372-8389.

Spyglass: data analysis framework for reproducible neuroscience research

Key Investigators

- Kyu Hyun Lee (Loren Frank lab, UCSF)

Project Description

Spyglass is a data analysis framework that brings together many open source tools, such as NWB, Datajoint, Spikeinterface, and others. Our lab has built reproducible analysis pipelines using these tools. Our goal is to make it possible to analyze any NWB file from DANDI with Spyglass. More specifically, we want to demonstrate that a neural decoding algorithm based on state-space models can be easily applied to data from other labs and can yield scientific insights.

Approach and Plan

One difficulty in achieving our goal is the heterogeneity among NWB files. To overcome this issue we plan to add a way to ingest the NWB file by augmenting information with an associated configuration yaml file.

Progress and Next Steps

- We have made progress toward achieving this goal, though it is not yet complete.
- We have demonstrated that the analysis tools we are using can be applied to data collected from other hippocampal labs by applying neural decoding to Dandiset 59.

Data

- <https://dandiarchive.org/dandiset/000059>

Materials

- <https://github.com/LorenFrankLab/spyglass>

Background and References

10. Use this space for information that may help people better understand your project, e.g., links to papers.

Contrast sensitivity across layers and cell types within primary visual cortex

Key Investigators

- Blake Mitchell
- Brock Carlson

Project Description

To analyze V1 responses (2-photon calcium imaging) to full field images with varying image contrast.

Approach and Plan

Describe the steps of your planned approach to reach the objectives.

Progress and Next Steps

- Describe the progress you have made on the project, e.g., which objectives you have achieved and how.
- Describe the next steps you are planning to take to complete the project.

Data

- Links to the dandiset(s) that you are using.

Materials

- Links to materials relevant to the project, e.g., code, videos.

Background and References

11. Use this space for information that may help people better understand your project, e.g., links to papers.

34

Ensemble computational models of neurons individually tuned with multimodal intracellular electrophysiology data

Key Investigators

- Krishna Pusluluri (Georgia State University)

35

Probing excitatory computations underlying probabilistic learning

Key Investigators

- Nuttida Rungratsameetaweemana (Salk Institute)

Project Description

It is not well understood how excitatory and inhibitory neurons within each cortical regions (as well as along the cortical hierarchy) work together to allow efficient information processing in dynamic environments.

Approach and Plan

- Constructing biologically plausible artificial neural network models.
- Developing a training paradigm to have the models perform cognitive tasks (such as probabilistic decision making) used in human and animal experiments.
- Investigating the neural computations that support successful learning and performance of the models on these tasks.
- Analyzing experimental data while the animals performed the same cognitive tasks.

Progress and Next Steps

- Developed recurrent neural network models to perform probabilistic learning tasks.
- Found that increased excitatory connections within neurons selective for likely sensory stimuli ('expected stimuli') support successful learning of statistical regularities.
- Next step is to look for how/where this dynamics is represented in the cortex.

Data

- 000037

Materials

- -

Background and References

- 12. Gillon et al., bioRxiv, 2021. Learning from unexpected events in the neocortical microcircuit

5 The Kavli NeuroData Discover Award

The Kavli Foundation issued a new award to generate new discoveries from reanalysis of data sets in the NWB format that is open only to attendees of the NeuroDataReHack workshop. The goals of this award are to:

1. Support innovative analyses of existing NWB-formatted datasets through conventional or novel analytic methods,
2. Promote studies and new approaches to analysis that will drive discoveries and accelerate the pace of fundamental research in neuroscience, and
3. Demonstrate that the secondary analysis of data can be used to examine questions beyond the scope of the original data.

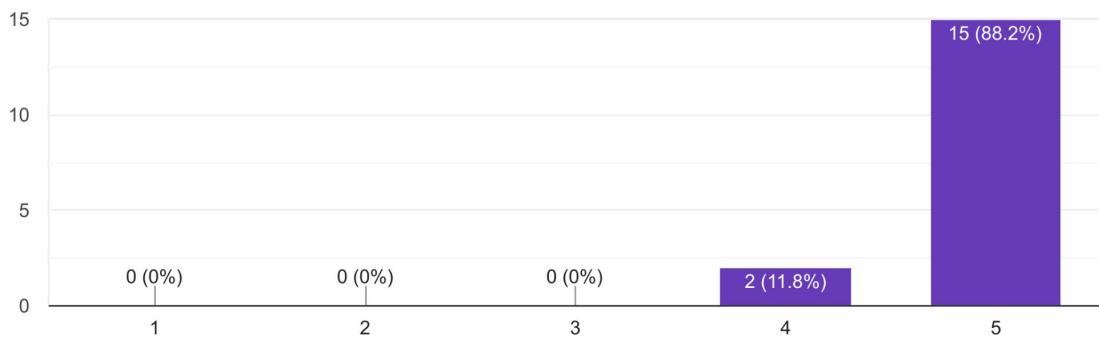
The award offers 50,000 USD for one year with a deadline of October 17th, shortly after the end of the NeuroDataReHack event. This funding opportunity received 14 applications, and they are currently under review.

6 Exit Survey

To help us improve this event in future years, attendees were asked to fill out an exit survey. We received 17 responses. For questions with a numeric score, 5 is the best and 1 the lowest score in the figures below.

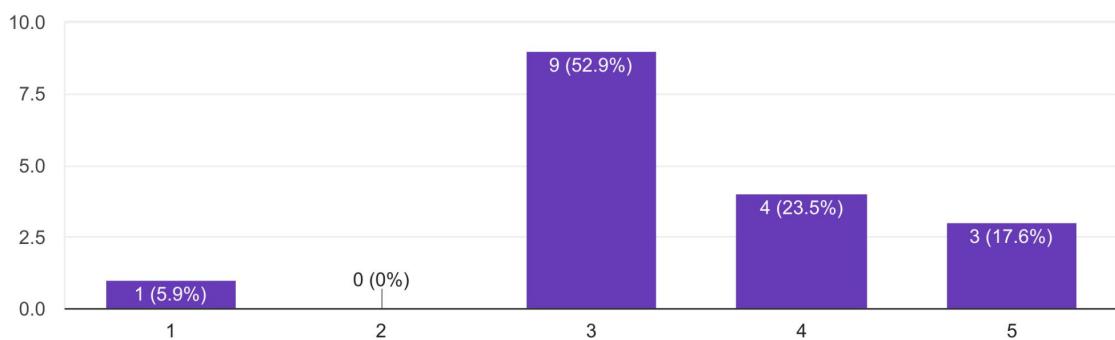
How was your experience at NeuroDataReHack?

17 responses



How easy was it to reuse neurophysiology data in the NWB format from the DANDI Archive?

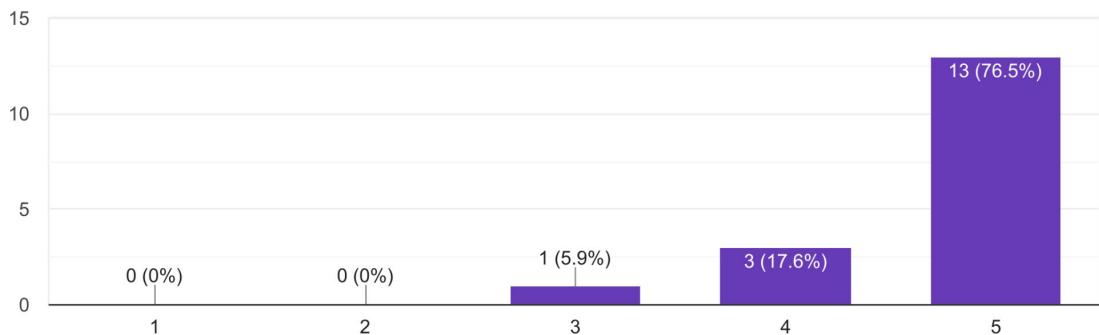
17 responses



38

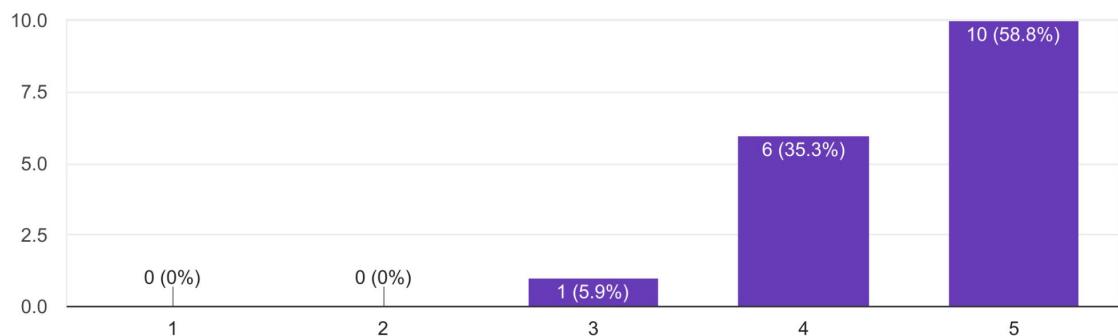
How easy was it get to help from the organizers and developers during the workshop?

17 responses



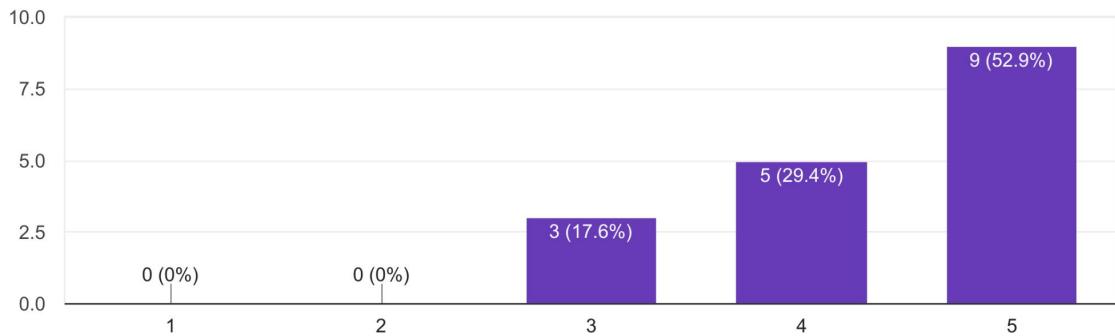
How useful were the training talks (advanced DANDI queries, DANDI Hub, reading files in NWB, reproducible analysis notebooks)?

17 responses



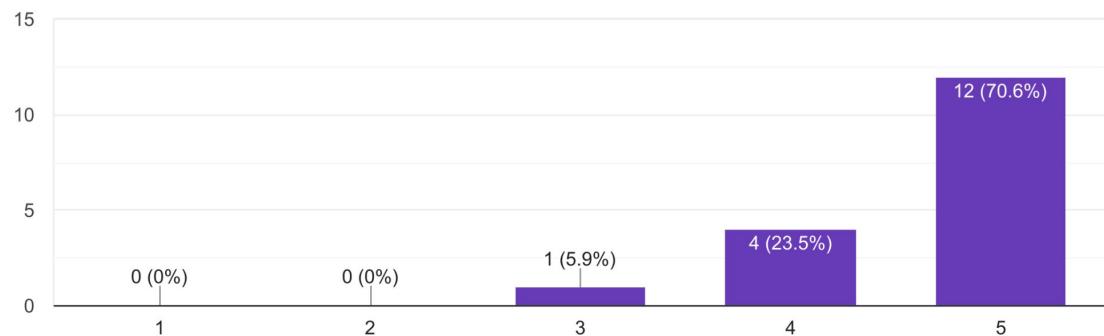
How useful were the breakout sessions (SpikeInterface, SSM, LFADS) overall?

17 responses



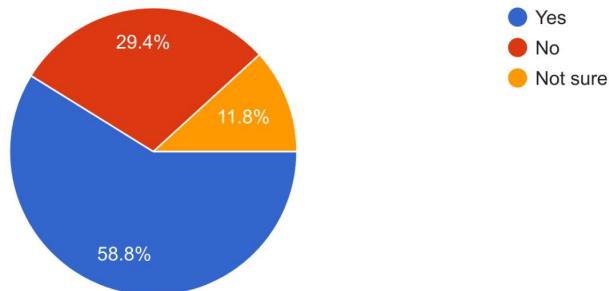
How useful was the hackathon for you to meet other scientists and developers?

17 responses



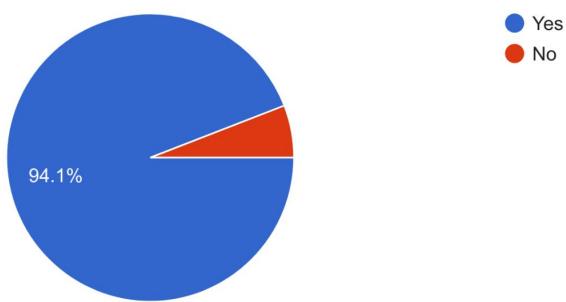
Do you plan to apply for the Kavli NeuroData Discovery grant?

17 responses



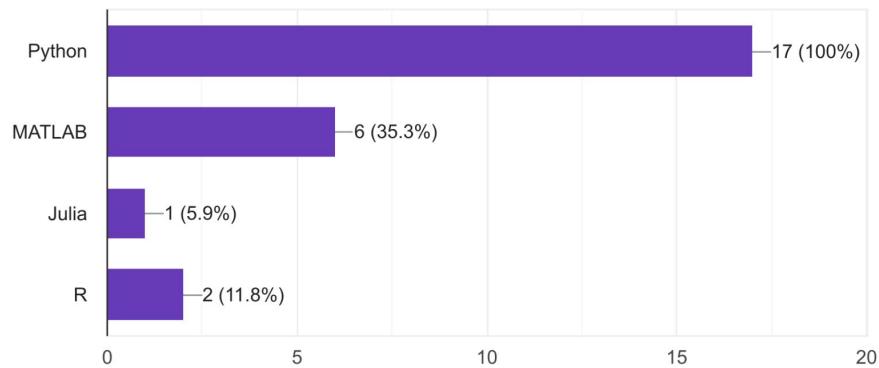
Do you now know how to read data from DANDI?

17 responses



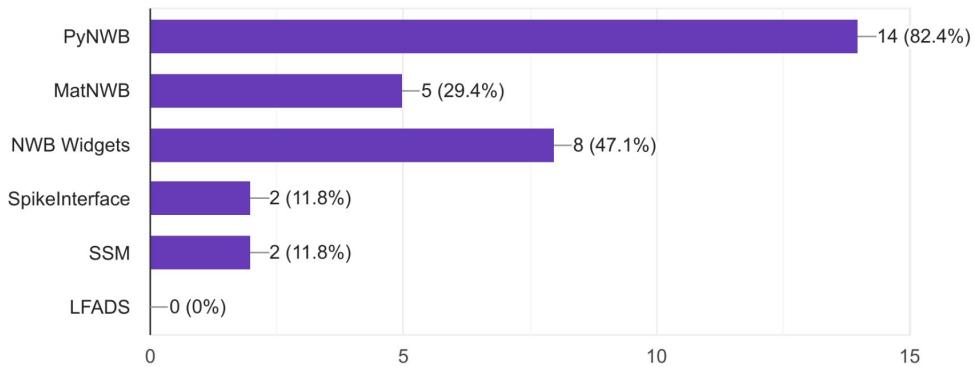
Which programming language(s) did you use to analyze your data?

17 responses



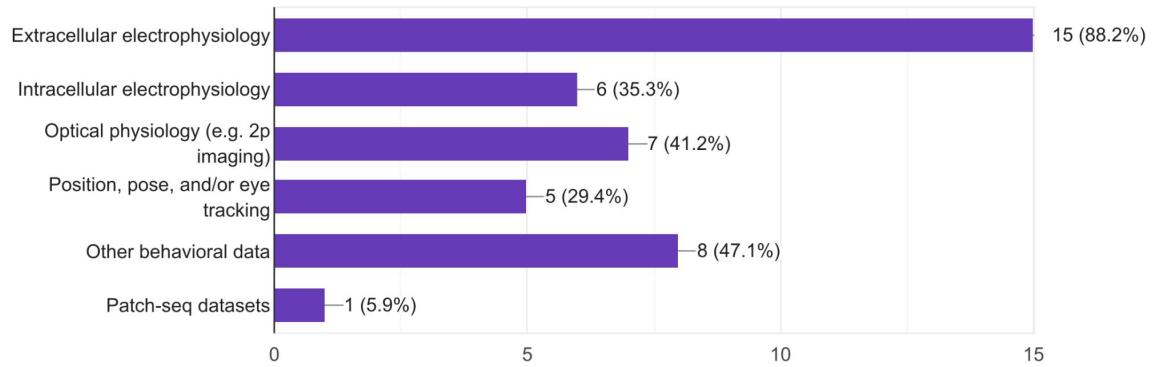
Which software tools did you use to analyze your data?

17 responses



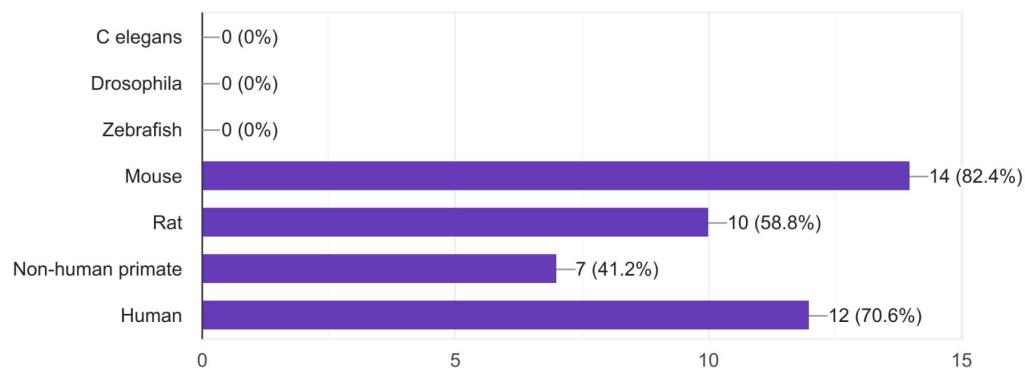
What types of data are you interested in reusing from DANDI?

17 responses



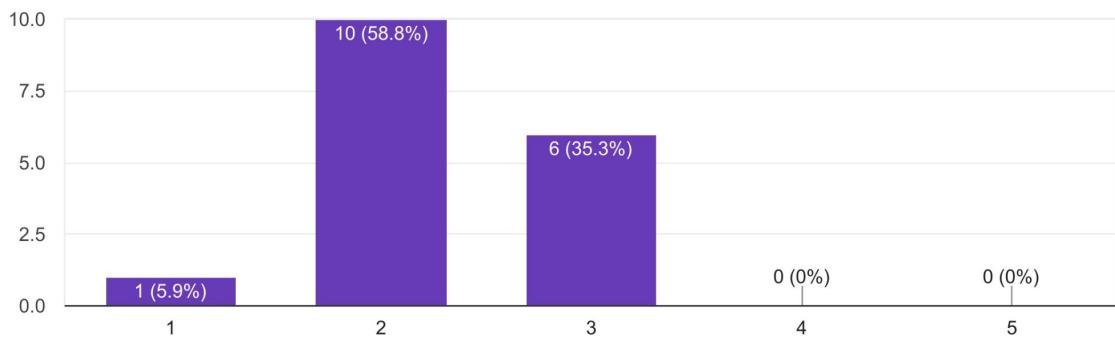
From what animal models are you interested in reusing data on DANDI?

17 responses



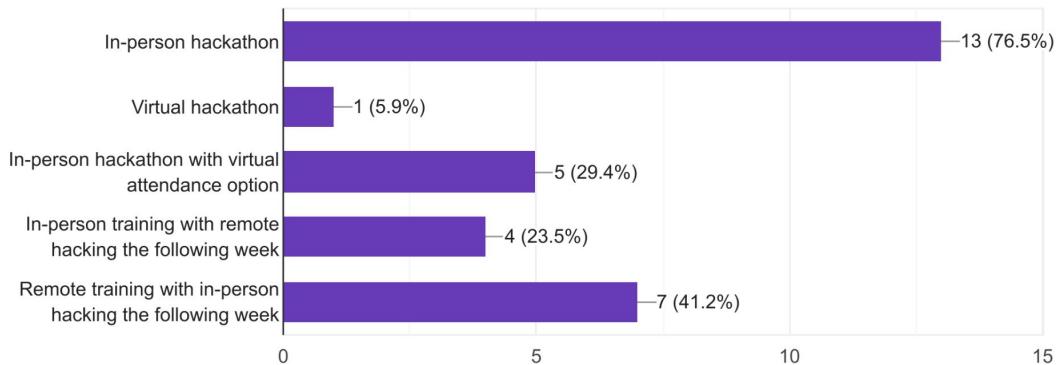
How was the length of the hackathon? (3 = just right)

17 responses



What hackathon formats would you prefer in the future?

17 responses



What did we do well at the hackathon?

- talks were helpful, environment was positive and fun
- You did a great job of providing enough time for me to truly dig into my project.
- Providing plenty of opportunities to hack and meet/work with other individuals
- Open agenda with plenty of room for hacking.
- Hotel and diner is great. Allen institute is also great.
- great talks and tutorials, very good atmosphere, the organizers put a lot of effots and wre very available
- Overall the hackathon was excellent. I found that having the developers in the same room as those implementing the software ecosystem to be very advantageous. In addition, numerous project leaders who have used the NWB ecosystem for publications. Lastly, the additional lectures on extra softwares which integrate NWB formats was beneficial and provided interesting modalities to analyse the data. The cohort selection was skilled and varied, but enough so to spark interesting collaborations, insights and discussions.
- I sensed and enjoyed the sense of community – meeting individuals with a similar mindset regarding open data/open science.
- Bring like-minded researchers together to work on projects (supported by common NWB-formatted data sets)
- Everything was great especially the lectures!

What could we improve on at future hackathons?

- making it longer, since talks interrupted hacking for most of the first two days
- For the electrophysiology data, it would be useful to have an example of how to plot stimulus triggered LFP and spiking responses presented sooner in the hackathon.
- Less time spent in talks (ssm/hidden markov), more time spent in inquiries. Assume that we know nothing on DANDI inquiry as there was a common difficulty in figuring out how to extract data from DANDIsets and work with the NWB format
- Breakfast should be slightly earlier to start the event at 9am.
- Arrange some outdoor activities at the Sunday night so that people could know each other well before the workshop.
- It was a great workshop, thanks for organizing it. I just feel that more networking time and opportunities to form collaborations between the participants would be helpful. For instance - some small team tasks or "research speed date" would be interesting and helpful. Also - I think it is better to open the slack channel before the workshop and to share the notebooks and slides before it so we can take notes with the slides during the presentations
- Possibly have a competition to finish projects across a few different themes/datasets
- Even though the hackathon was a large success, there are several ideas which I would suggest could be improved. Firstly, the event could benefit from given groups a 'task' to work on with specific dataset rather than self-guided (or existing) projects. The time was limiting, I personally felt that by the 3rd day I was really just beginning to experience the additional software tools available to analyse neurophysiology data in the NWB format. Lastly, more lectures would be beneficial, not just explicitly on NWB-based software packages. Lectures on different electrophysiology analyses would potentially enable 'neurodatabhakers' to begin experimenting with different datasets out of their original comfort zones (e.g. optical physiology) if they had grounding knowledge in the area. Yet again, well done on the event.
- More guidance about what's expected during the hackathons. E.g. working in teams expectations, form of final presentation
- encourage people to work in groups (of 2-3); more organized timeline and milestone (so that attendees know in more detail what to learn before the talk)
- A slightly longer workshop would be great and it will helpful if there is some shared document where all attendees can put some info about themselves and their plans on.

What training/hacking events on open neurophysiology data/software would you like to see us organize in the future?

- more neurodatareahacks would be cool
- I would love to see examples of how to convert my neurophysiology data into NWB format.
- I would love to see examples of how to convert my neurophysiology data into NWB format.
- This kind of event but with an additional virtual component with training to read and re-analyze NWB data.
- Someone like this is great. Collaboration is still very limited. One idea is to assign several tasks and people will be grouped by their common interest in the specific task.
- Dveloping our own widgets for NWB files; networking events
- I would like to learn more about the different analysis tools available for extracellular data, and to be kept up to date with any developments surrounding these. This may include oscillation detection, spike-sorting (correct curation and quality controls), PAC etc. These are just some ideas.
- Nwb extensions/tool building!
- it would be nice if you could organize an event that focus on a *specific type* of neurophysiology data, for example, event that focus on secondary analysis of extracellular physiology data
- Examples on how to link across different datasets.

In your opinion, what would make it easier for others to reuse neurophysiology data? What would incentivize others to reuse neurophysiology data?

- increasing awareness of (and continuing to improve) pipelines where you can just plug a dandiset in and analyze it
- It would be helpful to have specific information about what abbreviations and shorthand terms are used throughout the dandiset. It would be great if more specific information about how the nwb files within a dandiset are formatted were on the front page of the dataset in the archive. For example, in the 000021 dandiset, it would have been helpful to have a description of what the "locations" abbreviations stand for, and what locations were recorded in which files.
- More details (experimental design, behaviors et al.) should be included easily for introduction of the data.
- NWB standard is a great start. But to work with this data, the field is going to need a ton of broadcasted resources, like pretty much spoon feeding. Down to the nuts and bolts. Some good media broadcasting of youtube videos (showing NWB handling/ DANDI inquiry) would be a great resource.
- If the intended goal is for all to use NWB, it needs to be made very easy to learn because people get into their own data handling procedures and there are technically easier ways to upload your data to comply with open-data standards (figshare).
- A great incentive is to make it clear that once you know how to work with NWB and DANDI data once, you have access to hundreds of datasets. The selling point of analyzing data for preliminary purposes is also great - animal lives matter.
- More metadata, more backgrounds, more raw data (spike train and spike waveforms, not just values in their figure or averaged firing rate). Organizing more workshop like this and having more RFA/RFP.
- the data are great but good documentation will be helpful
- Have more detailed metadata. And make the metadata at the group level and at the individual subject level (when there is inter subject variability in what data is available) accessible through the webapp. You could possibly have a drop down menu, which when clicked, shows metadata for various objects with a nested structure (kind of like how IDEs display various in memory objects). And the webapp could only stream this metadata on the fly when someone requests for it.
- Another important aspect is to make sure there is not a huge fragmentation in standards across scientific communities. Have an open communication with different standards bodies and open data archives, to arrive at a somewhat agreeable, unified framework. This would be very helpful when someone is trying to reuse data, and so they will not need to learn a new data format for every new dataset they find interesting.

- I think at current, a downfall with the NWB ecosystem appears to be the metadata storage. An example of this would be the inconsistencies between intracellular electrophysiology data in the format structure. For example, some DANDIs will promote 'voltage clamp' acquisition yet looking through the raw data these can not be found. Since the course I have spoken to several owners of these DANDIs whom have informed me that voltage clamp recordings do not exist in these (Howard Derek) but can be made available. Hence, this appears to be a confusion with the NWB data upload.
- In addition, there are key metadata information which would be critical for electrophysiologists repurposing the intracellular patch-clamp data to find (e.g. sampling rates, perfusion temperatures, solution concentrations, stimulus protocols). In some cases this is provided, but not all. For Patch-seq transcriptomic datasets this may also include the quality controls (if applied) and gene count cutoffs. I understand that this is a difficult task and at current and NWB significantly improves on the lack of standardisation surrounding neurophysiology data-sharing.
- For the Human data sets, I found that many were missing the location and layer in the brain that the patch-clamp recordings were made. This is essential to asking important questions surrounding the electrophysiological properties of different brain regions and subsequent layers. Patient information for the resections (If ethics approves) would also be very beneficial to asking further questions surrounding the relationship between the patient and their electrophysiology. If this is something to be discussed with the owner of the DANDIs, a possible email option should be made available for further questions. I found myself emailing a repository owner for such information but could not easily find this.
- I believe with further 'vetting' of the DANDIset uploads by owner this will improve the metadata standardisation. Making this information to find as simple as possible without users needing to look elsewhere will improve the usage of the NWB ecosystem for neurophysiology data repurposing.
- More detailed tags on the Dandi Archive, so that users can locate the data sets with confidence, and need not open the file and explore them with NWBWidgets to check if the data set indeed contains relevant information.
- Easy-to-follow video tutorials with example notebooks.

Testimonials

"The NeuroDataReHack workshop undoubtedly helped me grow and develop as a scientist. I gained access to an entire suite of tools and learned several new analytical techniques. The few days spent at the Allen Institute contained some of the most formative moments of my graduate career thus far." Brock Carlson, Vanderbilt University, Graduate Program in Psychology.

"Data handling and storing procedures can vary significantly across labs and even between individuals within a lab. This variability makes reproduction and data exploration difficult and prone to error. NWB provides a solution to this issue by implementing data standardization practices. Once you have the capability to access one NWB-formatted dataset, the door is opened to so many other datasets! My appreciation for NWB really comes from my time-spent at the NeuroDataReHack event. I would recommend this event to anybody with practical coding experience and would recommend that any or all neuroscientists attend or watch a mini event held by the NeuroDataReHack team." John Stout, University of Delaware

47

Any additional questions, comments or suggestions?

- The hardest part of this event was figuring out the NWB format and actually accessing the data. I would argue it actually took away from the event a little because the first day or so was spent trying to figure out the programming logic.
- A super simple solution that keeps your audience diverse is to have a mandatory virtual event where simple NWB and DANDI operations are covered in python and code is provided. By recording this lecture/event, participants can do the backend work before actually arriving for the 3-day event! I also liked the Day 0 event, but I'm not sure that the discussion panel was needed. The short presentations were nice, and the opportunity to meet individuals at the virtual poster was also nice!
- I am looking for the 2nd NeuroData Rehack.
- It was a great workshop, thanks for organizing it! I would be happy if there are follow up events to keep in touch with the organizers and participants
- I hope there will be a next event. I will be encouraging many more people from my institute to apply. I also think that it would be great (if it exists) to be emailed monthly regarding on any updates to the NWB ecosystem and DANDI archives and new datasets uploaded. Thank you all yet again for having me at such a great event!

- Thank you so much for organizing the hackathon! :-)
- Thank you the organizers so much again for putting this together!

7 Photos from the Event

Day 0: Open Neurodata Showcase Virtual Poster Session

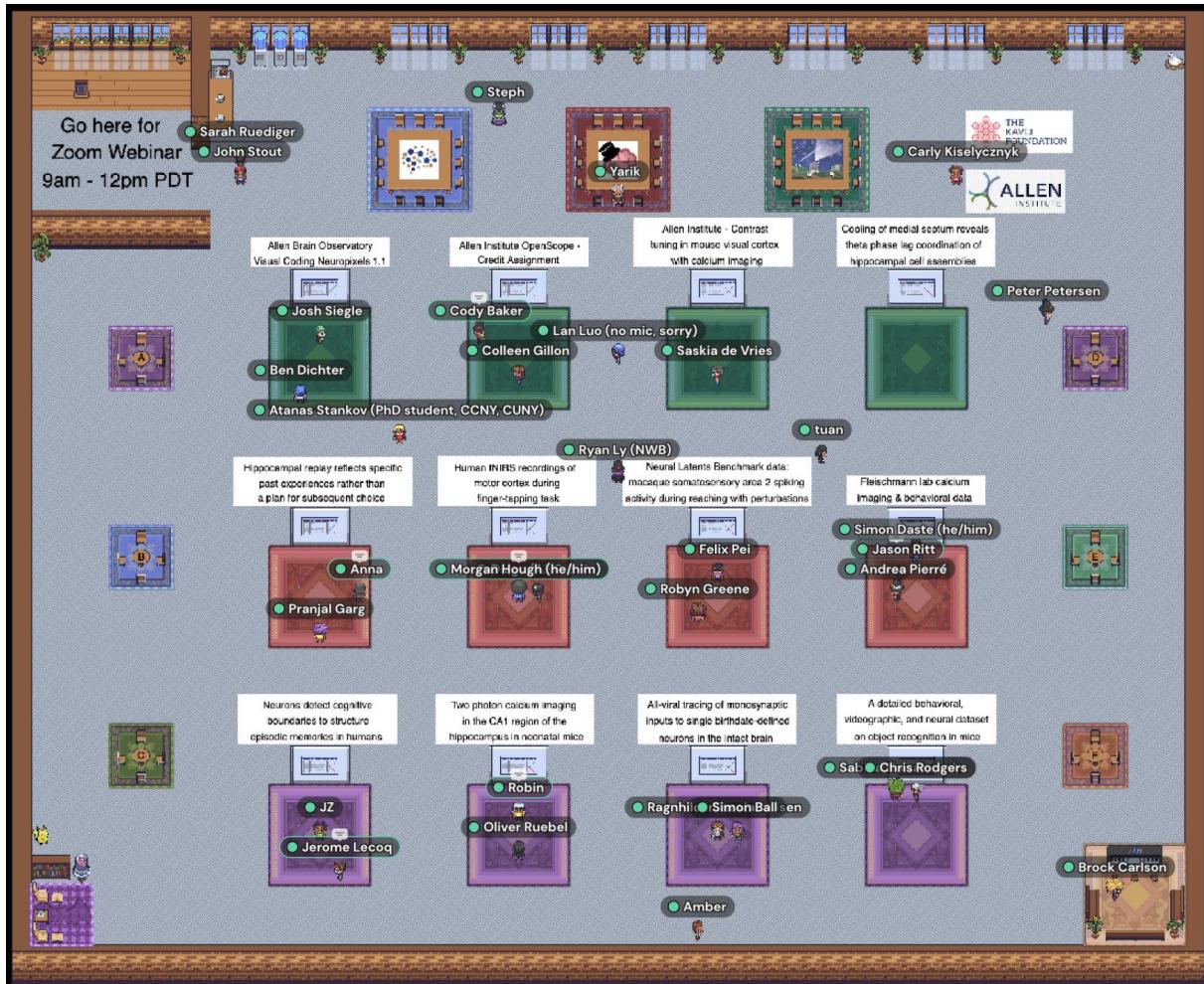
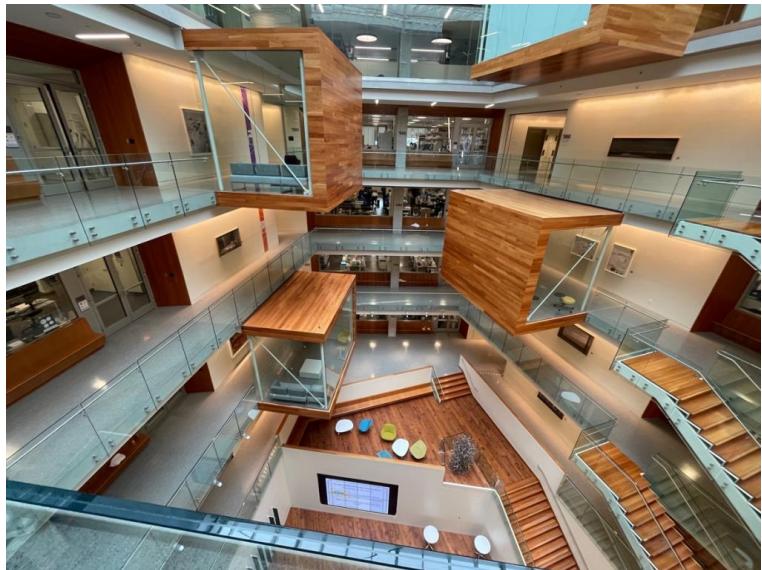


Figure 5: Photos from Day 1, Oct 3, 2022.

Day 1

Arrive at the Institute and getting to know each other



Catering by Hungry and Harried



50



Figure 6: Photos from Day 2, Oct 4, 2022.

Day 2

Hacking!



Tour of the Allen Institute



Group dinner at Cactus



51

Figure 7: Photos from Day 3, Oct 5, 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). Thanks to the Allen Institute for welcoming us into their space, organizing local logistics, and coordinating catering. We also want to thank the Kavli Foundation for the Discovery Award RFP.

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.