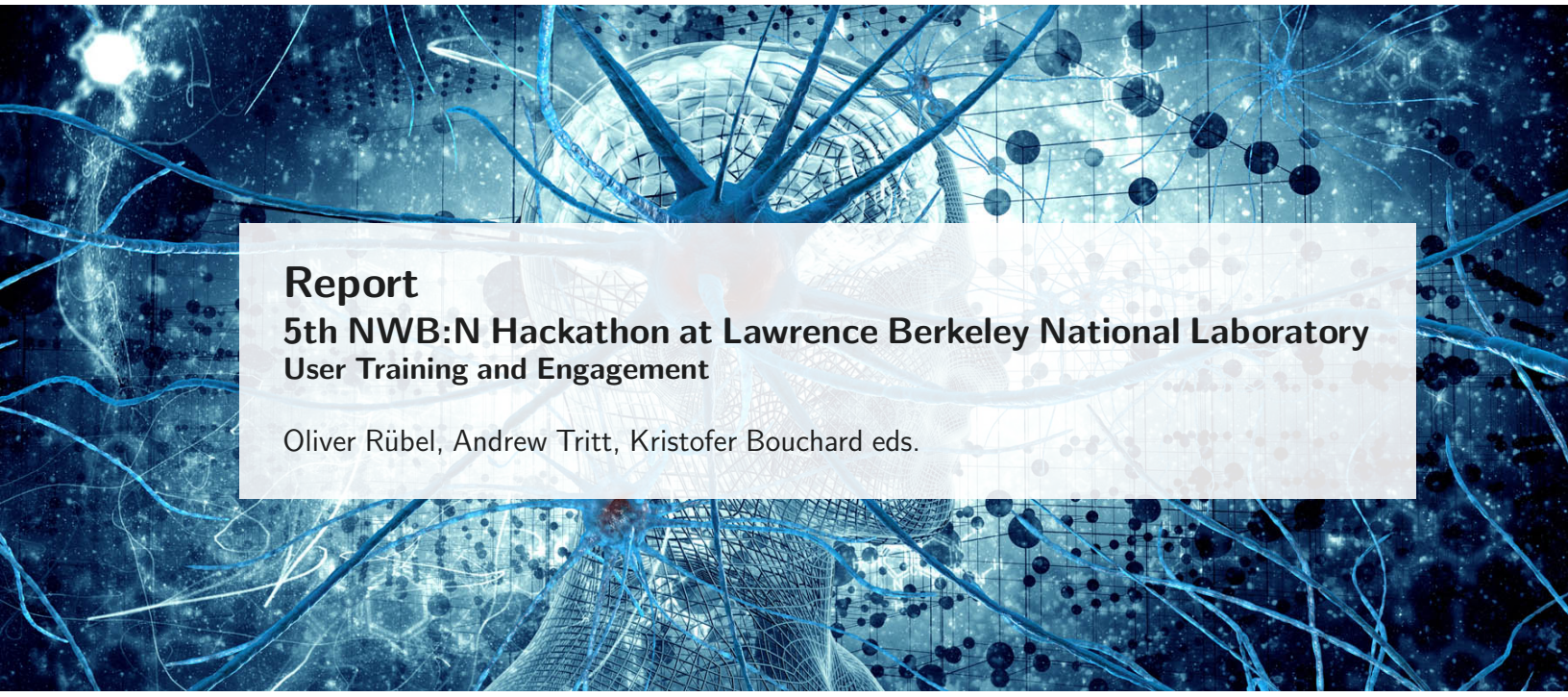


Report

5th NWB:N Hackathon at Lawrence Berkeley National Laboratory

User Training and Engagement

Oliver Rübel, Andrew Tritt, Kristofer Bouchard eds.

Image courtesy of www.nwb.org

Participants

- Oliver Ruebel (LBNL)
- Andrew Tritt (LBNL)
- Kristofer Bouchard (LBNL, NSE Lab)
- Max Dougherty (LBNL, NSE Lab)
- Tom Davidson (UCSF - Frank Lab)
- Karen Lavi (UCSF)
- Ali Mohebi (UCSF)
- Dylan Paiton (UCB)
- Evan Lyall (UCB)
- Jeff Teeters (UCB)
- Stephanie Albin (The Kavli Foundation)
- Kevin Brown (New York University)
- David Tingley (NYU)
- Sam McKenzie (NYUMC)
- Sebi Rolotti (Columbia University, Losonczy Lab)
- Jochen Weber (Columbia University, Zuckerman Mind Brain Behavior Institute)
- David Thibodeaux (Columbia University)
- Ben Dichter (Stanford)
- Aaron Milstein (Stanford University)
- Ivan Raikov (Stanford University)
- Kei Masuda (Stanford University, Giocomo Lab)
- Jason Bant (Stanford University)
- Adam Granger (Harvard Medical School)
- Duo Xu (Johns Hopkins University)
- Eric Finkel (Johns Hopkins School of Medicine, O'Connor Lab)
- Michael Grauer (Kitware)
- Jean-Christophe Fillion-Robin (Kitware)
- Michael Wulf (Cold Spring Harbor Laboratory, Kepecs Lab)
- Torben Ott (Cold Spring Harbor Laboratory)
- Nathan Clack (Vidrio Technologies)
- Lawrence Niu (Vidrio Technologies)
- Nicholas Cain (Allen Institute for Brain Science)

NEURODATA:N WITHOUT BORDERS

THE KAVLI FOUNDATION

BERKELEY LAB
Bringing Science Solutions to the World

Photo by Margie Wylie, Lawrence Berkeley National Laboratory (see the LBNL News Release)

Contents

Executive Summary **3**

Suggestions for Future Events 4

Suggestions for Specific Developments 4

Survey 5

Organization **6**

Program **7**

Breakdown of Event Times 7

Agenda: Day 1 7

Agenda: Day 2 8

User Projects **9**

Event Photos **13**

Acknowledgements **14**

Disclaimer **14**

Executive Summary

The Neurodata Without Borders: Neurophysiology (NWB:N) project is an effort to standardize the description and storage of neurophysiology data and metadata. NWB:N enables data sharing and reuse and reduces the energy-barrier to applying data analytics both within and across labs. This hackathon invited experts from the neuroscience community to explore adopting NWB:N for their data sharing and analysis needs and lab use cases. The goal of this event was to: **a)** train new users on NWB:N, **b)** promote adoption of NWB:N, **c)** work with users on programming projects, e.g, to integrate examples of their labs data into NWB:N, **d)** facilitate communication between users and developers and project teams, and **e)** engage with the community.

31 scientists from 13 major institutions and approximately 20 different labs/groups attended the event. In addition, we had attendance by one friendly canine ☺. The background in programming varied greatly among the participants, ranging from beginners to experts. Most participants had investigated NWB:N prior to the event but were generally in the early stages of exploration and adoption.

During the coding sessions, participants worked on their User Projects, which they had defined (and prepared) prior to the event. The user projects involved data from a broad range of modalities: **a)** electroyphysiology (voltage and current-clamp recordings, ECoG), **b)** optophysiology, **c)** static imaging modalities (2-photon imaging, fluorescent wide-field images, webcam, etc.,), **d)** behavioral data, **e)** processed data (e.g., from spike sorting), **f)** neural simulations (Neuron), **g)** stimulus data, and **h)** trial-based data. The user projects also included data from a number of different species, such as, mice, rats, and monkeys. See the User Projects section for a more in-depth overview of the user projects at the hackathon.

In order to provide a balance between user training and project-oriented work, we designed the agenda to consist of an even split of talks (3h) and tutorials (4h20min) and hacking on user projects (7h20min) (see the Program). Group discussions were held during lunch on the first day for introduction and review of user hacking projects and at the end of the second day for review of progress, feedback, and discussion of the event.

By the end of the hackathon, the vast majority of attendees had accomplished many of their goals, and several were well on their way to having NWB:N fulfill their labs needs. All attendees indicated an enthusiasm for the current software and were impressed with the (relative) ease of use given the generality. As an important step towards adoption, a few members (A. Milstein, S. Rolotti, J. Weber, M. Dougherty) indicated willingness to serve as 'tutors' for upcoming hackathon events: more effort is needed in the direction of training members of the community to become trainers themselves. There was also consensus on a few features that would greatly enhance the utilization and adoptability of NWB:N:

- **Extensions sharing:** The ability to share and use extensions across labs was recognized as a critical need. Formalizing and standardizing this process will be critical. This is already on the roadmap for NWB:N development as part of pending NIH proposals and was also part of the Extension Sharing project at the 4th hackathon at the Allen Institute .
- **Controlled Vocabularies and Ontologies:** The ability to associate controlled terms (e.g., enumerations) and complex ontologies with NWB:N data fields was recognized as a central need. This is already on the roadmap for NWB:N development as part of pending NIH proposals.
- **Tools for visualization and analysis:** Tools for exploration and visualization of NWB:N was recognized as a general need. Some development efforts are already underway (e.g., Kitware, Gepetto etc.) but more efforts and integrated support for this are needed.
- **Data search and query:** Advanced tools for search, discovery, and query of NWB:N data are needed to facilitate advanced data analytics. This is already on the roadmap for NWB:N development as part of pending NIH proposals.
- **Trial-based data:** Support for trial-based data is currently limited in NWB:N and needs to be expanded (see Issue #152)

These should be targets for future development efforts. See also the Suggestions for Specific Developments section for additional (lower-level) specific developments suggested by attendees at the hackathon.

Overall, the hackathon was a great success. Based on discussions at the hackathon the suggestion is to ideally have similar, user-oriented hacking events approximately every 6 month, at least at this early stage of development/adoption of NWB:N 2.

Suggestions for Future Events

- **The hackathon was very useful:** Nearly all participants indicated that they would be willing to attend another NWB:N hacking event in the future. All participants appeared to have made significant progress on their projects during the event.
- **Desire for even more time for project hacking:** Tom Davidson and other attendees indicated that a central value of the event is to get people in the same room. In general participants indicated a desire for more time for programming and working on projects.
- **2 days is a good length for the event:** The overall length of the meeting of 2 days appeared to have been perceived as good, in part because it provides a compromise between the busy schedules of participants and the value-add from a hacking event. In that regard, Nathan Clack and others suggested to have shorter/less talks to accommodate more project sessions. In addition, another option may be to have an optional third day just for project hacking and working groups.
- **Consider adding breakout sessions:** With regard to the issue of allowing for more hacking, it was suggested to have breakout sessions on particular topics (e.g., ephys, ophys, advanced I/O etc.) to enable attendees to focus on the topics that are of most value to them while at the same time allowing us to cover a broad range of topics.
- **Consider alternate name for the event:** Karen Lavi (UCSF) suggested that the term "Hackathon" may pose a hurdle for diversity. To attract more female attendees she suggested that the event should be renamed to be more open and welcoming. In the series of hackathons, this was really the first event that focused on user training and community engagement. As such, we may want use the term hackathon for developer events and create a new name for training and engagement events that better reflects the user/community focus. In general encouraging more diversity and participation of women and minorities will be useful.
- **Consider adding code/file review groups:** Jochen Weber (ZMBBI) suggested to add code/file review sessions in which participants would be paired with one (or more) participants from other lab, to review codes (generated at the event) in pairs or small subgroups to provide feedback, facilitate communication, and help prepare data (and codes) for sharing and typical use cases in other labs.
- **Satellite events:** The option of holding satellite tutorial, training, and community engagement events at major neuroscience conferences was also mentioned, specifically SfN and Cosyne.

Suggestions for Specific Developments

In addition to the more high-level direction for future development discussed in the Executive Summary we here describe additional, specific developments that were suggested by participants at the hackathon.

- **Create simple end-to-end tutorial:** Create a tutorial that shows the creation of an extension and its use for write and read on real data. https://github.com/NeurodataWithoutBorders/pynwb/issues/505
- **Improve support for trial data**: https://github.com/NeurodataWithoutBorders/nwb-schema/issues/152
- **Ways to browse through NWB:N files:** Improve introspection, search, query, and visualization capabilities.
- **Event data:** Specialize for ripple, downstate
- **Interfacing visualization and analysis tools with NWB:N:** E.g., Neuroscope (C/C++). Read NWB data for kilosort, mclust, neurosuite.
- **Cache the Specification:** Cache the specification (including extensions) always to the file. This suggestion also related to the desire for formal mechanisms for sharing extensions.
- **Communicate mission of NWB:N:** Should NWB:N be used only for sharing of data or day-to-day use? Answer: goal is day-to-day.
- **Multi-plane imaging:** Storing multi-plane imaging data is not optimally supported in NWB:N yet. In particular, ROIs sometimes span imagine planes in this context.
- **Neuron data:** Made a lot of progress already.
- **Enable modeling of data constraints for validation:** This could be either part of the schema language or part of a separate validation model to define additional data constraints. Currently this functionality is implemented by the API classes but should be formalized to ensure consistency of constraints cross languages/APIs.
- **Facilitate integration/linking to data in non-NWB:N files:** To ease adoption at labs some labs would like to link to custom external file formats (e.g., binary data) to avoid data copy and allow use of existing lab software tool-chains on existing data. In PyNWB this could be accomplished via custom I/O backends as well as via customization of the ObjectMapper API.

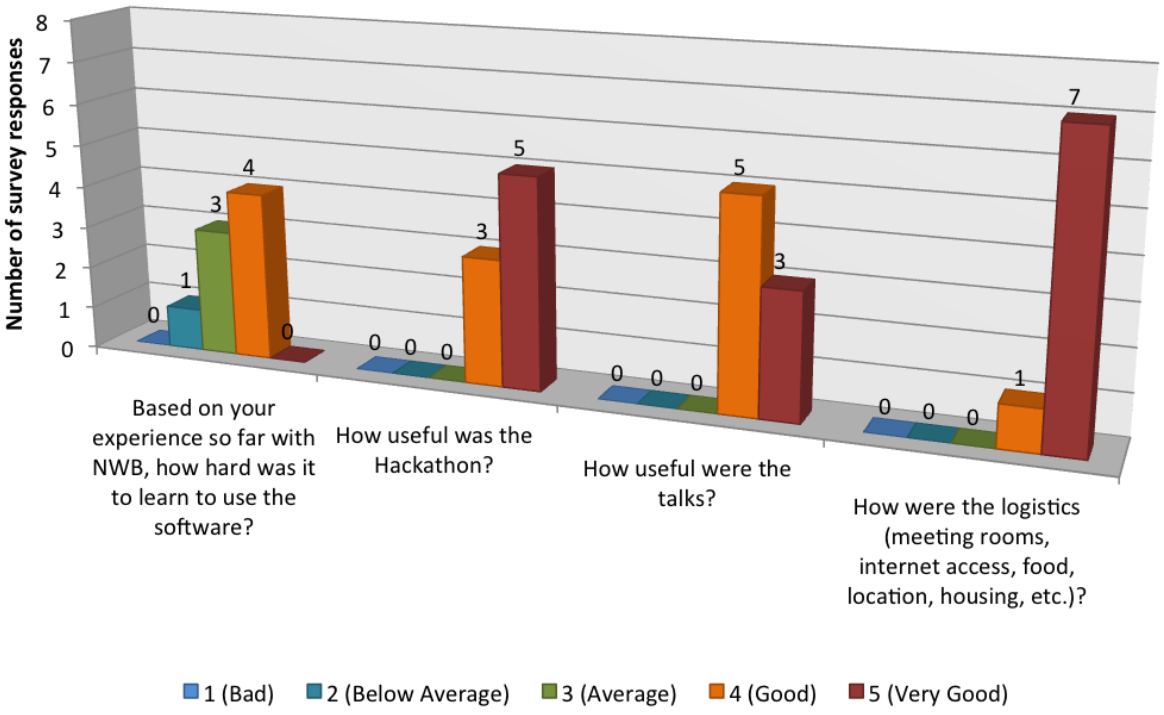
- **Iterative Write (PyNWB):** Works
- **Lazy load (PyNWB):** Works

Several Matlab-specific suggestions included:

- **Continue to support Matlab:** Attendees at the event estimated that ≈ 50% of target users are currently MATLAB users.
- **MatNWB: Consistent ordering of dimensions:** Dimensions in Matlab are ordered differently from Python. Make this consistent to follow the spec. https://github.com/NeurodataWithoutBorders/matnwb/issues/26
- **MatNWB: Support lazy data load:** https://github.com/NeurodataWithoutBorders/matnwb/issues/27

Survey

All participants were asked to participate in an exit survey. 8 participants (i.e., 26%) submitted responses to the survey. The following figure summarizes the results of the survey questions in which users were asked to rate the hackathon and software using a numeric scale ranging from 1 (worst) to 5 (best).



Organization

The meeting was organized by Oliver Ruebel (site/program chair), Andrew Tritt and Kristofer Bouchard with administrative support by Dionne Myers and additional logistics support by Stephanie Albin (Kavli Foundation). Travel support for the meeting was provided by the KAVLI Foundation and food and other meeting resources were sponsored by Lawrence Berkeley National Laboratory.

Program

- **Event Website:** https://neurodatawithoutborders.github.io/nwb_hackathons/HCK05_2018_Berkeley
- **Dates:** April 26-27, 2018
- **Location:** Lawrence Berkeley National Laboratory, Berkeley, CA, USA, Building 59 - Shyh Wang Hall
- **Slides:** The slides from the talks and tutorials presented at the event are available online at: https://neurodatawithoutborders.github.io/nwb_hackathons/HCK05_2018_Berkeley/#slides

Breakdown of Event Times

		Day 1	Day 2	Totals across both days
Talks		2:30:00	0:30:00	**3:00:00**
Tutorials with hacking exercises		1:30:00	2:50:00	**4:20:00**
Hacking on user projects		4:10:00	3:10:00	**7:20:00**
Breaks		1:50:00	2:00:00	**3:50:00**
Group Discussion		0:10:00	1:00:00	**1:10:00**
Totals per day		**10:10:00**	**9:30:00**	

Agenda: Day 1

Day 1: Thursday April, 26

Time	Duration	Topic	Speaker	Room	
7:30 - 8:00	0:30:00	Registration & Breakfast		59-4101	
8:00 - 8:10	0:10:00	Welcome to the NWB:N Hackathon @ LBNL	Oliver Ruebel	59-4102	
8:10 - 8:20	0:10:00	Neurodata Without Borders - Standardization of neurophysiology data ...	Kris Bouchard	59-4102	
8:20 - 8:40	0:20:00	Introduction to NWB:N – An ecosystem for standardizing neurophysiology data	Oliver Ruebel	59-4102	
8:40 - 9:40	0:30:00	NWB:N and PyNWB – A Python API for standardizing neurophysiology data	Andrew Tritt	59-4102	
9:40 - 10:10	0:30:00	Overview of MatNWB	Nathan Clack, Lawrence Niu	59-4102	
10:10 - 10:25	0:15:00	Break		59-4101, 59-4102	
10:25 - 11:10	1:30:00	NWB:N I/O – ophys	Nicholas Cain	59-4102	
10:10 - 11:55	1:30:00	NWB:N I/O – ephys	Ben Dichter	59-4102	
11:55 - 12:30	0:30:00	Tour of the National Energy Research Scientific Computing Center (Optional)		59-4102, NERSC	
11:55 - 1:00	1:05:00	Working Lunch with presentation on: Expectations and Overview of Hackathon Projects	Oliver Ruebel	59-4101, 59-4102	
1:00 - 3:00	2:00:00	Hacking on projects		59-4101, 59-4102, 59-4016	
3:00 - 3:10	0:10:00	Group Photo		Entrance 59	
3:10 - 3:30	0:20:00	Break		59-4101, 59-4102	
3:30 - 5:40	2:10:00	Hacking on projects		59-4101, 59-4102, 59-4016	
6:30	2:00:00	Group Dinner (Optional)		Triple Rock (1920 Shattuck Avenue)	

Agenda: Day 2

Day 2: Friday April, 27

Time	Duration	Topic	Speaker	Room	
8:30 - 9:30	0:30:00	Breakfast		59-4101	
9:00 - 9:10	0:10:00	Welcome to Day 2 of the NWB:N Hackathon @ LBNL	Oliver Ruebel	59-4102	
9:10 - 9:30	0:20:00	PyNWB Software Process and How to Contribute	Michael Grauer, Jean-Christophe Fillion-Robin	59-4102	
9:30 - 10:30	1:00:00	Writing Extensions	Ben Dichter	59-4102	
10:30 - 10:45	0:15:00	Break		59-4101, 59-4102	
10:45 - 11:45	1:00:00	PyNWB Containers – Front-end objects for NWB data	Andrew Tritt	59-4102	
11:55 - 1:00	1:15:00	Working Lunch with presentation on: Advanced Data I/O	Oliver Ruebel	59-4101, 59-4102	
1:00 - 1:30	0:30:00	Advanced Data I/O (Hacking exercises)	Oliver Ruebel	59-4102	
1:30 - 3:00	1:30:00	Hacking on projects		59-4101, 59-4102, 59-3049	
3:00 - 3:20	0:20:00	Break		59-4101, 59-4102	
3:20 - 5:00	1:40:00	Hacking on projects		59-4101, 59-4102, 59-3049	
5:00 - 6:00	1:00:00	Group Discussion (Review of progress and feedback)		59-4101	

User Projects

Prior to the event, each participant was asked by the organizers to formulate a specific project that she/he will work on. The participants created project pages on the hackathon GitHub repository. Participants then updated their pages during (and after) the event to document progress. The tables on the following pages provide an overview of the various user projects. The project headings include links to the specific project pages for further details.

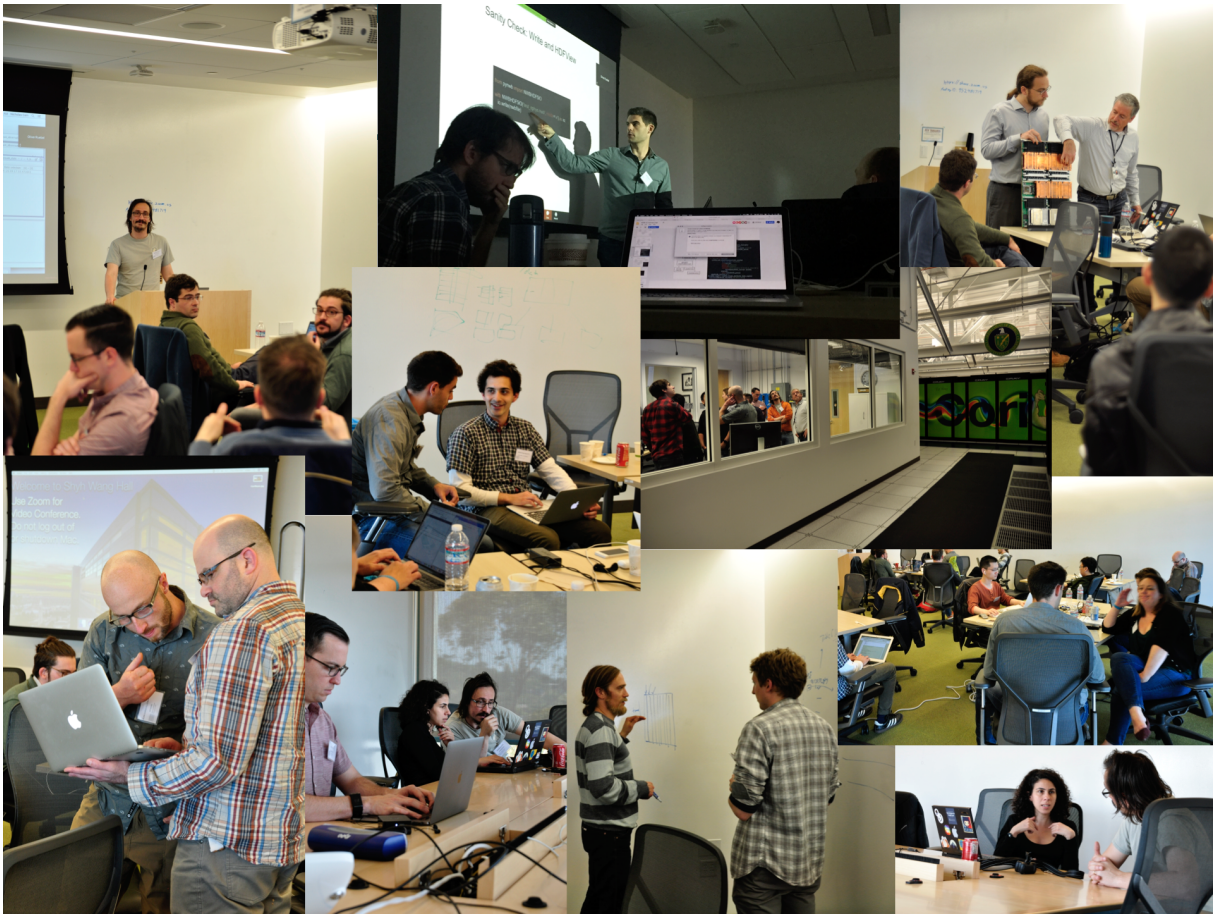
Title	Key Investigators	Key Institutions
NWB Camera Capture code This project will investigate the use of the NWB framework to compile data from multiple sources and formats into a single NWB file. This data includes intrinsic and fluorescent wide-field images from awake behaving mice, as well as behavioral webcam images to accompany it. We hope to construct an NWB file, import the data into the file, and then read it into a matlab workspace for later analysis.	Jochen Weber, David Nicholas	Zuckerman Mind Brain Behavior Institute, Columbia University
Conversion of Sabatini lab neurophysiology and imaging data to the NWB:N format The goal of this project is to convert the whole-cell electrophysiology data from the Sabatini lab to the NWB format. This data includes voltage-clamp and current-clamp recordings from neurons in acute brain slices. This data is typically analyzed for the presence or absence of synaptic responses or changes in membrane potential. However, we often extract other features such as response amplitude, latency, onset and offset kinetics, frequency of synaptic events or action potentials, and variability of responses in repeated trials. Therefore, a standard toolset to analyze this data in the NWB data format would also be useful. Finally, we often record electrophysiology data in tandem with 2-photon imaging, and would like to be able to incorporate imaging data into the NWB:N file format.	Adam Granger	Sabatini Lab, Harvard Medical School
Conversion of multimodal data between Losonczy Lab and NWB:N formats Our lab collects 2-photon imaging and extracellular electrophysiological data in awake behaving mice. Our overall goal is to increase the interoperability of our data structures and formats with those of NWB. We already have fairly well developed repositories for working with our data, including a public package for sequential imaging data (SIMA) as well as in-house lab analysis bundle (LAB) for associating and analyzing these various data formats. We therefore hope to allow for seamless import from and export to NWB using the APIs for these repos.	Sebi Rolotti	Losonczy Lab, Columbia University
Integrating Buzcode standard into NWB	David Tingley	Buzsaki Lab, New York University

Title	Key Investigators	Key Institutions
A database of dopamine cell recordings in freely-moving rats Recordings of dopamine cells in behaving animals have been enormously influential in neuroscience, psychology, and even computer science. In head-restrained monkeys and mice performing passive tasks, dopamine cells appear to encode reward prediction errors (RPEs) - abrupt changes in the anticipation of future rewards. RPE signals are central to reinforcement learning algorithms, and are a core component of recent advances in artificial intelligence (e.g. AlphaGo). However, there is more and more data indicating that RPEs are an incomplete description of dopamine function, including recent results from our laboratory (Hamid et al. 2016, Nature Neuroscience). As part of our ongoing dopamine studies we have been recording from optogenetically-identified dopamine cells ($n > 33$) in the rat ventral tegmental area. These are - to our knowledge - the first recordings of identified dopamine cells in unrestrained animals performing distinct behavioral tasks, including a probabilistic reward ("bandit") task, a runway task, and a pavlovian approach task. Initial analyses suggest that the RPE idea is a good description of some aspects of dopamine cells firing, but not others. This unique, high-quality data set is certain to be of great interest to many researchers, especially computational neuroscientists who do not perform their own recordings. We wish to convert this dopamine data, with associated behavioral and anatomical metadata, into the NWB format as an important step towards freely-distributing these data to the community.	Ali Mohebi, Josh Berke	University of California San Francisco (UCSF)
Filter Read by Label We would like to define a standard way to associate data points in an NWB file with multiple user-defined categorical labels, and then be able to selectively read from a dataset just the data points that satisfy a boolean filter based on those labels. An example would be a dataset of UnitTimes containing spike trains recorded from an extracellular electrophysiology experiment. Each neuron in the dataset could be assigned a categorical label according to its 1) cell_type, 2) feature_selectivity (grid cell or head direction cell?), 3) active during sleep / wake?, etc. Then the user could read into memory the spikes from all cells that match a query for 'cell_type' == 'basket_cell', 'selectivity' == 'running_speed', 'behavioral_state' == 'sleep', e.g.	Aaron Milstein, Ivan Raikov, Ben Dichter	Standford University
O'Connor Lab's Crossmodal Attention Project Conversion of extracellular electrophysiology data from our crossmodal attention project to the NWB:N format	Eric Finkel	Johns Hopkins School of Medicine, O'Connor Lab
Calcium imaging of PFC neurons during Open Field paradigm The goal of this project is to assess the conversion of neuronal data of the Sohal lab to the NWB format. The data of this POC project includes in-vivo calcium imaging from PFC neurons in freely behaving mice in the open field aperture. A time series of the animal's position within the aperture is also included. Currently, our lab is using in house solutions for working with data in Matlab. Here we hope to allow seamless import and export to NWB using the available APIs.	Karen Lavi, Scott Wilke	Sohal lab, University of California San Francisco (UCSF)

Title	Key Investigators	Key Institutions
Conversion of Pesaran lab electrophysiology data to the NWB:N format The goal of this project is to convert local field potential and spike time data from the Pesaran Lab to the NWB format. This data includes behavioral data from awake, behaving Rhesus Macaques. This data is typically analyzed by aligning spike times or local field potential data to behavioral events of interest, such as the initiation of a saccade. Therefore, it would be useful to have a clean mechanism for associating data from combinations of arbitrary behavioral conditions (e.g. a saccade to the left during an instructed delay task two-alternative forced choice task), user-defined data conditions (e.g. spike time data from cells tuned to saccade direction from electrodes in the lateral intraparietal sulcus) at user-specified events (e.g. the beginnning of an optogenetic stimulation sequence). As our lab continues to extend recording modalities (e.g. to 2-photon calcium imaging), it would also be useful to have the ability to seamlessly integrate new data formats and types. Our lab has previously developed in house solutions for working with the above data in Matlab. Here we hope to allow seamless import and export to NWB using the available APIs.	Kevin Brown, Bijan Pesaran	Pesaran lab, New York University
Converting e-phys/behavioral data into NWB:N data format (Kepecs lab) This project aims to convert electrophysiological, imaging and behavioral data into the nwb data format. Data was obtained from different projects in Adam Kepecs' lab at CSHL. We made progress converting voltage timeseries and spike events intwo NWB:B. For our trial-based behavioral data, there seems to be no appropriate neurodata type available at the moment. We discussed with Andrew that there could be a table-like neurodata type for organizing trial data, such as categories, doubles and logicals. And additional Trial neurodata type could be used to specify in particular the start times or epochs of each trial. More generally, our trials are defined as a sequence of states, i.e. could be described using state machines, as Kris pointed out.	Michael Wulf, Torben Ott, Adam Kepecs	Kepecs lab, Cold Spring Harbor Laboratory
Conversion of Adesnik lab calcium imaging data to the NWB:N format The goal of this project is to convert the whole-cell electrophysiology data from the Sabatini lab to the NWB format. This data includes voltage-clamp and current-clamp recordings from neurons in acute brain slices. This data is typically analyzed for the presence or absence of synaptic responses or changes in membrane potential. However, we often extract other features such as response amplitude, latency, onset and offset kinetics, frequency of synaptic events or action potentials, and variability of responses in repeated trials. Therefore, a standard toolset to analyze this data in the NWB data format would also be useful. Finally, we often record electrophysiology data in tandem with 2-photon imaging, and would like to be able to incorporate imaging data into the NWB:N file format.	Evan Lyall	Adesnik Lab, University of California Berkeley (UCB)

Title	Key Investigators	Key Institutions
User Training and Use Cases The goal of this project is 1) to conduct new user training and 2) to interact with attendees on their use cases to identify needs for NWB:N.	Oliver Ruebel, Andrew Tritt, Ben Dichter, Nicholas Cain, Kristofer Bouchard, Michael Grauer, Jean-Christophe Fillion-Robin, Nathan Clack, Lawrence Niu	Lawrence Berkeley National Laboratory, Stanford University, Allen Institute for Brain Science, Kitware, Vidrio
Convert Bouchard Lab Data	Max Dougherty	Neural Systems and Engineering Lab, Lawrence Berkeley National Laboratory
Community engagement	Stephanie Albin	The Kavli Foundation
Community engagement. Frank Lab Data.	Tom Davidson	Frank Lab, University of California San Francisco (UCSF)
	Dylan Paiton	Univerity of California Berkeley (UCB)
	Jeff Teeters	Univerity of California Berkeley (UCB)
	Sam McKenzie	NYUMC
	Kei Masuda	Giocomo Lab, Stanford University
	Jason Bant	Stanford University
	Duo Xu	Johns Hopkins University

Event Photos



Acknowledgements

We would like to thank Nicholas Cain, Ben Dichter, Michael Grauer, Jean-Christophe Fillion Robin, Nathan Clack, Lawrence Niu, Kristofer Bouchard, Andrew Tritt and Oliver Ruebel for their contributions as presenters and facilitators at the event. Special thanks also to Stephanie Albin (Kavli Foundation) for logistics support and for her participation at the hackathon. We thank Dionne Myers (LBNL) for local logistics support. We thank Margie Wylie-Petruzziello for photography and assistance during the event as well as Linda Vu and Kathy Kincade for LBNL communications support. We thank David Paul, Elizabeth J. Bautista, Cary Whitney and all the other NERSC staff that have helped with the amazing tour of the NERSC supercomputers. We would like to thank all the participants for the great enthusiasm and commitment and for making the hackathon a great success! We thank Lawrence Berkeley National Laboratory and the Kavli Foundation for financial support of the hackathon.

Disclaimer