# nwb4edu: an Online Textbook for Teaching and Learning with NWB Datasets

Ashley L. Juavinett [1] **and Victor Magdaleno-Garcia**[2]

1 UC San Diego, Neurobiology **2** UC San Diego, Cognitive Science

Summary

The nwb4edu online textbook provides an interactive set of lessons for educators, students, and self-guided learners who would like to use open neuroscience datasets for research and/or teaching. Specifically, it introduces users to Neurodata Without Borders (NWB), a relatively new data format for neurophysiology data (Rübel et al., 2022; Teeters et al., 2015) which can be accessed via DANDI (Distributed Archives for Neurophysiology Data Integration). The nwb4edu textbook contains tutorials within an interactive JupyterBook (Executable Books Project, 2021) that enable users to wrangle, analyze, and visualize NWB datasets. The datasets featured in this resource currently include rodent behavioral data, single-cell and extracellular electrophysiological recordings, and two-photon imaging experiments. Importantly, it also provides users with the necessary conceptual background to understand how the data were collected and how they can be analyzed. The broad understanding of how a dataset is structured and how to navigate a dataset can be applied to other data standards, while the in-depth understanding of NWB datasets can empower students and researchers to conduct their own research using these data.

Each lesson contains learning objectives, tasks, and a problem set so that students can test and extend their skills. Executable notebooks can be launched on a pre-configured JupyterHub, making these easily accessible to anyone with an internet connection. All of the materials are covered by a CC-BY-SA-4.0 and are available at http://nwb4edu.github.io.

In its current form, there are four complete lessons, which contain multiple notebooks. Each lesson contains specific learning objectives, such as:

- Identify and implement multiple ways to obtain a NWB dataset via DANDI
- Explore the metadata contained in a typical Dandiset (Economo et al., 2018a, 2018b) and explain the three common features of datasets: metadata, raw data, and processed data
- Implement spike sorting on extracellular recording data in mice (Campbell et al., 2021a, 2021b)
- Generate and interpret common visualizations of extracellular recording data
- Access and visualize single-cell electrophysiology and two-photon imaging data (S. E. de Vries et al., 2023; S. E. J. de Vries et al., 2020) from the Allen Institute

Furthermore, there are a growing set of single notebook lessons ("Greatest Hits in Neuroscience") in which users can re-create figures from published scientific papers (Churchland et al., 2012; Sargolini et al., 2006) using their NWB-formatted data. These lessons transform foundational papers in neuroscience into living lessons, breathing new life into old data and inviting all students in to conduct authentic research.

Statement of Need

There is an ongoing and enthusiastic movement in neuroscience to standardize data formats as a means of improving reproducibility and encouraging data reuse. NWB is one of these formats, and has proved itself to be an effective data standard to share, store, and build analysis tools for neuroscience data. The next generation of neuroscience researchers needs skills to work with data standards like NWB to perform data analyses, visualize neural data, and ultimately address pressing scientific questions. A standardized and accessible data format like NWB also opens possibilities for more easily integrating these cutting-edge datasets into the undergraduate or graduate classroom and can empower student-guided exploration.

While many researchers have acknowledged the expanding intersection of data science and neuroscience, there are very few educational resources to help students and educators grow these skill sets. There are a multitude of introductory programming resources and several notable modules for learning how to code alongside neural data, but fewer advanced resources for neural data science. To meet this need, the `nwb4edu` online workbook provides a set of exercises – for anyone from an advanced undergraduate to an early career researcher – who wishes to learn how to access, manipulate, and visualize NWB datasets. If we want open science and data standardization to be the future, we need to train our students accordingly and provide researchers with open access tools so that they can easily adapt data into their pipelines.

These materials are designed to be adapted as individual modules, so that instructors or self-guided learners may choose what best suits their learning needs. The project is entirely open access and can be tailored for different experience levels or modified to address specific learning objectives.

Instructional Design

The creation of this book was supported and motivated by a Kavli Foundation Seed Grant which supported the senior author (A.L.J.) to create a learner-facing resource for NWB datasets. Two of the lesson plans within, involving the Allen Institute data, were created and tested with undergraduate students in various neurobiology courses at UC San Diego. These were the first lessons in the new `nwb4edu` book. We then built two lessons focusing on native NWB format, with the goal of introducing students to the mechanics of a standardized data format as well as conducting very common neural data analyses and visualizations with the data. These native NWB lessons were used in a Neural Data Science course for undergraduate and graduate students at UC San Diego. Finally, in collaboration with the NWB team, we integrated two additional "Greatest Hits" lessons. These lessons had not yet not been used by any of the authors in a course setting.

Acknowledgements

References

Campbell, M. G., Attinger, A., Ocko, S. A., Ganguli, S., & Giocomo, L. M. (2021a). *VR Data Neuropixel supporting "Distance-tuned neurons drive specialized path integration calculations in medial entorhinal cortex"* (Version 0.210819.0345). DANDI archive; DANDI Archive. https://doi.org/10.48324/dandi.000053/0.210819.0345

Campbell, M. G., Attinger, A., Ocko, S. A., Ganguli, S., & Giocomo, L. M. (2021b).

Distance-tuned neurons drive specialized path integration calculations in medial entorhinal cortex. *Cell Reports*, *36*(10), 109669. https://doi.org/10.1016/j.celrep.2021.109669

Churchland, M. M., Cunningham, J. P., Kaufman, M. T., Foster, J. D., Nuyujukian, P., Ryu, S. I., & Shenoy, K. V. (2012). Neural population dynamics during reaching. *Nature*, *487*(7405), 51–56. https://doi.org/10.1038/nature11129

Economo, M. N., Viswanathan, S., Tasic, B., Bas, E., Winnubst, J., Menon, V., Graybuck, L. T., Nguyen, T. N., Smith, K. A., Yao, Z., Wang, L., Gerfen, C. R., Chandrashekar, J., Zeng, H., Looger, L. L., & Svoboda, K. (2018a). *Mouse anterior lateral motor cortex (ALM) in delay response task*. DANDI archive; DANDI Archive. https://dandiarchive.org/dandiset/000006

Economo, M. N., Viswanathan, S., Tasic, B., Bas, E., Winnubst, J., Menon, V., Graybuck, L. T., Nguyen, T. N., Smith, K. A., Yao, Z., Wang, L., Gerfen, C. R., Chandrashekar, J., Zeng, H., Looger, L. L., & Svoboda, K. (2018b). Distinct descending motor cortex pathways and their roles in movement. *Nature*, *563*(7729), 79–84. https://doi.org/10.1038/s41586-018-0642-9

Executable Books Project. (2021). *Jupyter book* (Version v0.10.0). Zenodo. https://doi.org/10.5281/zenodo.4539666

Rübel, O., Tritt, A., Ly, R., Dichter, B. K., Ghosh, S., Niu, L., Baker, P., Soltesz, I., Ng, L., Svoboda, K., Frank, L., & Bouchard, K. E. (2022). The Neurodata Without Borders ecosystem for neurophysiological data science. *eLife*, *11*, e78362. https://doi.org/10.7554/eLife.78362

Sargolini, F., Fyhn, M., Hafting, T., McNaughton, B. L., Witter, M. P., Moser, M.-B., & Moser, E. I. (2006). Conjunctive Representation of Position, Direction, and Velocity in Entorhinal Cortex. *Science*, *312*(5774), 758–762. https://doi.org/10.1126/science.1125572

Teeters, J. L., Godfrey, K., Young, R., Dang, C., Friedsam, C., Wark, B., Asari, H., Peron, S., Li, N., Peyrache, A., Denisov, G., Siegle, J. H., Olsen, S. R., Martin, C., Chun, M., Tripathy, S., Blanche, T. J., Harris, K., Buzsáki, G., … Sommer, F. T. (2015). Neurodata Without Borders: Creating a Common Data Format for Neurophysiology. *Neuron*, *88*(4), 629–634. https://doi.org/10.1016/j.neuron.2015.10.025

Vries, S. E. de, Siegle, J. H., & Koch, C. (2023). Sharing neurophysiology data from the Allen Brain Observatory. *eLife*, *12*, e85550. https://doi.org/10.7554/eLife.85550

Vries, S. E. J. de, Lecoq, J. A., Buice, M. A., Groblewski, P. A., Ocker, G. K., Oliver, M., Feng, D., Cain, N., Ledochowitsch, P., Millman, D., Roll, K., Garrett, M., Keenan, T., Kuan, L., Mihalas, S., Olsen, S., Thompson, C., Wakeman, W., Waters, J., … Koch, C. (2020). A large-scale standardized physiological survey reveals functional organization of the mouse visual cortex. *Nature Neuroscience*, *23*(1), 138–151. https://doi.org/10.1038/s41593-019-0550-9