Cody C. Baker

cody.c.baker.phd@gmail.com cody-c-baker codycbakerphd.com
github.com/CodyCBakerPhD

Education

2015 – 2020 **Ph.D.** Applied and Computational Mathematics and Statistics at *University of Notre Dame*
Doctoral thesis: Second-order moments of activity in large neural network models

2011 – 2015 **B.S.** Applied Mathematics and Cognitive Science at *University of Evansville*
Senior thesis: Crime prediction models applied to the city of Evansville
Honors thesis: Goodness of fit metrics for joint probability distributions

Experience

2020 – 2024 **Research Software Engineer** for *CatalystNeuro*

2016 – 2020 **Research Assistant** for the Neural Computing Group at *University of Notre Dame*

2015 – 2017 **Teaching Assistant** for Applied Mathematical Methods at *University of Notre Dame*

Summer 2015 **Data Analyst** for *JRM Environmental Inc*

Technical Editor for *Penguin Random House*

2014 – 2015 **Lab Manager** for the Digital Humanities Lab at *University of Evansville*

Summer 2014 **Research Intern** for the BITLab at *Michigan State University*

Summer 2013 **Research Intern** for the Molecular Biosciences program at *University of Kansas*

Projects

CatalystNeuro

As a research software engineer, I was responsible for a wide range of administrative and technical tasks:

i developing and maintaining open-source software, including automated testing suites, documentation, tutorials, and demos
ii handling user interactions within the research community, such as timely responses to issues and feature requests
iii managing the group's cloud resources (Amazon Web Services), including storage, computing resources, and identity access management
iv offering technical support and consultations to current and future clients
v curated a grand total of 256 TB of high-value datasets to the NIH data archive on behalf of various research groups

The types of software I helped to create include the range of fullstack; everything from the backend APIs to frontend desktop and web applications. All software related in some way to the facilitation of terabyte-scale data management, analysis, and visualization for the field of neurophysiology.

2022 - 2024 **Lead developer** of *NeuroConv*
An automated data conversion tool capable of reading more than 40 distinct data formats used by neurophysiology experiment devices and writing to the NeurodataWithoutBorders (NWB) standard. Designed universal APIs for transparently handling each layer of complexity (tagging, grouping, metadata, temporal alignment, asset linking, buffering, chunking, and compression). Also implemented a cloud deployment system for running batched conversions via Amazon Web Services. Ran hands-on user training sessions at multiple conferences and workshops.

Projects (continued)

2023 – 2024 **Lead developer** of *NWB GUIDE*
A user-centric desktop application to guide experimentalists through the process of creating NWB files. Developed intuitive user interface with interactive validation, real-time suggestions for metadata improvements, and automatic data upload to the NIH archive. Tracked and documented dozens of hands-on user tests to refine the user experience.

2022 – 2024 **Lead developer** of *NWB Inspector*
A command line tool used by the NIH data archive to validate all data uploads as well as make suggestions for metadata improvements to enhance data findability and reuse.

2024 **S3 log parsing** for the DANDI Archive
Wrote custom command line tool for parsing and analyzing over 6 TB of S3 log files to track download activity and generate anonymized usage statistics for all assets on the archive.

UX consultant for *Neurosift*
Using experience from a related project, NWB Widgets, I provided feedback on the user interface design of a web application for browsing and streaming contents of NWB files. I also ran a number of performance benchmarking tests to identify bottlenecks in the streaming protocols caused by imperfect storage patterns and used the results to set better defaults for chunking and compression of large datasets consisting of both time series and image data.

Data schema designer for the Adesnik Lab at *University of California, Berkeley*
Coordinated with multiple top research groups pioneering novel neural photomstimulation methods capable of single-cell resolution. Designed a new data schema for the representation of temporal sequences with spatially distributed patterns and all associated metadata. Guided the submission of the formal extension proposal to the NWB standard.

Data publication for the Leifer Lab at *Princeton University*
Wrote custom code for interfacing with the output of a unique microscopy rig, including an electrically tunable lens capable of acquiring images at variable (but precisely measured) depths. Published a total of 10 TB of *C. elegans* microscopy data on the associated NIH data archive.

Data publication for the Visual Coding project at the *Allen Institute for Brain Science*
Wrote custom code to standardize the metadata for thousands of microscopy experiments and their associated visual stimuli; published 93.4 TB of mouse visual cortex microscopy data.

2022 **Data publication** for the Ahrens Lab at *Janelia Research Campus*
Wrote custom conversion code for a coupled virtual reality and microscopy environment; converted, uploaded, and published 23 TB of zebrafish microscopy data.

Data publication for the International Brain Lab
Wrote conversion code for interfacing with a custom metadata database; converted, uploaded, and published 66 TB of synchronized electrophysiology, audio, and behavioral event data.

2021 **Data publication** for the Buzsaki Lab at *New York University*
Wrote custom conversion code for a dozen previous datasets; converted, uploaded, and published 39 TB of mouse electrophysiology and behavioral data.

2021-2024 **Supervised data publications**
Instructed multiple members of the CatalystNeuro team in the process of converting, uploading, and publishing the data for research groups. Projects supervised included groups such as Datta (Harvard), Jazayeri (Massachusetts Institute of Technology), Tye (Salk Institute), Clandinin (Stanford), and the MICrONS project with the Tolias lab (Baylor University). Data types spanned the breadth of electrophysiology, microscopy, and behavior (including audio/video).

Publications

Journal Articles

1. C. J. Gillon et al., "Open Data In Neurophysiology: Advancements, Solutions & Challenges," *ArXiv*, arXiv–2407, 2024.
2. J. Magland, J. Soules, C. Baker, and B. Dichter, "Neurosift: DANDI exploration and NWB visualization in the browser," *Journal of Open Source Software*, vol. 9, no. 97, p. 6590, 2024.
3. M. Hawrylycz et al., "A guide to the BRAIN Initiative Cell Census Network data ecosystem," *PLoS biology*, vol. 21, no. 6, e3002133, 2023.
4. C. Baker, E. Froudarakis, D. Yatsenko, A. S. Tolias, and R. Rosenbaum, "Inference of synaptic connectivity and external variability in neural microcircuits," *Journal of computational neuroscience*, vol. 48, pp. 123–147, 2020.
5. C. Baker, V. Zhu, and R. Rosenbaum, "Nonlinear stimulus representations in neural circuits with approximate excitatory-inhibitory balance," *PLoS computational biology*, vol. 16, no. 9, e1008192, 2020.
6. C. Baker, C. Ebsch, I. Lampl, and R. Rosenbaum, "Correlated states in balanced neuronal networks," *Physical Review E*, vol. 99, no. 5, p. 052 414, 2019.

Conference Presentations

1. C. Baker, V. Zhu, and R. Rosenbaum, "NWB GUIDE: Simplifying the conversion of neurophysiology data to NWB format," in *Society for Neuroscience (SfN)*, 2024.
2. C. Baker, S. Weigl, H. Mayorquin, and B. Dichter, "NeuroConv: Automated conversion of neurophysiology data to NWB format," in *Society for Neuroscience (SfN)*, 2023.
3. C. Baker, V. Zhu, and R. Rosenbaum, "Nonlinear computations in semi-balanced networks," in *Computational and Systems Neuroscience (COSYNE)*, contributed talk (top 4% of abstracts), 2020.
4. C. Baker and R. Rosenbaum, "Inferring connectivity and latent input covariance from spike train correlations," in *Computational and Systems Neuroscience (COSYNE)*, poster (top 43% of abstracts), 2018.

Workshop Sessions

1. **Baker, Cody** and G. Viejo, "Artificial intellgence, machine learning, computing, and visualization in neuroinformatics," in *Open Data in Neuroscience (ODIN)*, Massachusetts Institute of Technology, 2023.

Skills

Coding	Python, MATLAB, R, LaTeX
DevOps	Git, GitHub Actions, Docker, unit & integration testing, Storybook, Chromatic, Puppeteer, Codecov, pre-commit linting, user & quality testing
Libraries	NumPy, SciPy, pandas, sklearn, OpenCV, FFmpeg, timeit, airspeed velocity, Matplotlib, Plotly, Flask, PyInstaller, Pydantic, Boto3
Frameworks	AWS (IAM, S3, EC2, Batch, DynamoDB), Electron, REDIS, REST, GraphQL, Jupyter, Sphinx
Data formats	HDF5, Zarr, JSON, YAML, XML, Markdown, RST, Parquet, TSV, XLSX, MP4, AVI, TIFF, PNG, JPEG, MP3, WAV, PKL, MAT
Administration	GitHub, AWS, PyPI, Condaforge, Docker Hub, ReadTheDocs
WebDev	HTML, CSS, JavaScript, TypeScript

Skills (continued)

Interpersonal — Research, mentoring, training, client consultations

Interests

Cooking | Running | Ice Hockey | Audio Analysis | Gaming