Neuron

CellPress

NeuroView

Benefits of sharing neurophysiology data from the BRAIN Initiative Research Opportunities in Humans Consortium

Vasiliki Rahimzadeh,[1,12] Kathryn Maxson Jones,[1,2,12] Mary A. Majumder,[1] Michael J. Kahana,[3] Ueli Rutishauser,[4] Ziv M. Williams,[5] Sydney S. Cash,[6] Angelique C. Paulk,[6] Jie Zheng,[7] Michael S. Beauchamp,[8] Jennifer L. Collinger,[9] Nader Pouratian,[10] Amy L. McGuire,[1] Sameer A. Sheth,[11,*] and NIH Research Opportunities in Humans (ROH) Consortium

[1]Center for Medical Ethics and Health Policy, Baylor College of Medicine, Houston, TX 77030, USA
[2]Department of History, Purdue University, West Lafayette, IN 47907, USA
[3]Department of Psychology, University of Pennsylvania, Philadelphia, PA 19104, USA
[4]Department of Neurosurgery, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA
[5]Department of Neurosurgery, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114, USA
[6]Department of Neurology, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114, USA
[7]Department of Ophthalmology, Boston Children's Hospital, Boston, MA 02115, USA
[8]Department of Neurosurgery, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA
[9]Rehab Neural Engineering Labs, Department of Physical Medicine and Rehabilitation, University of Pittsburgh, Pittsburgh, PA 15219, USA
[10]Department of Neurological Surgery, UT Southwestern Medical Center, Dallas, TX 75390, USA
[11]Department of Neurosurgery, Baylor College of Medicine, Houston, TX 77030, USA
[12]These authors contributed equally
*Correspondence: sameer.sheth@bcm.edu
https://doi.org/10.1016/j.neuron.2023.09.029

Sharing human brain data can yield scientific benefits, but because of various disincentives, only a fraction of these data is currently shared. We profile three successful data-sharing experiences from the NIH BRAIN Initiative Research Opportunities in Humans (ROH) Consortium and demonstrate benefits to data producers and to users.

Introduction

Research funders, journals, and institutions have increased their expectations for FAIR (findable, accessible, interoperable, and reusable) neurophysiology data sharing.[1] Under the new US National Institutes of Health (NIH) Data Management and Sharing (DMS) Policy, research data generated using federal funds are now required to be deposited into designated archives "as soon as possible, and no later than the time of an associated publication or the end of the award/support period, whichever comes first" (NOT-OD-21-013). The NIH-led BRAIN Initiative enforces its own data-sharing policy, with similar terms of sharing (NOT-MH-19-010).

Sharing and reusing human neural data can inform new research directions, save money, drive innovation, enhance rigor, and minimize waste.[2] The case for sharing invasive human electrophysiology data, such as intracranial EEG (iEEG) and single-neuron recordings from the brain, is particularly compelling, owing to the rarity of these data and the resource-intensiveness required for their collection. Yet, investigators can feel disincentivized to share because of time, resources, and training required as well as concerns regarding lack of attribution.[1,2] Researchers sharing human brain data often are faced with similar disincentives, and thus, only a portion of these data is currently shared. Nevertheless, there have been many efforts to share human neuroimaging data over the past two decades, one of the most successful being the Human Connectome Project.[3] In this case, unanticipated issues arose that serve as key learning opportunities for invasive human electrophysiology, even as the benefits of sharing also have been clear, with at least 1,538 publications resulting from the data as of 2021.[1,3,4]

In this NeuroView, we describe scientific and social value propositions for the FAIR sharing of invasive human electrophysiology data and highlight the benefits of sharing for data generators. Our hope is to combat a pure compliance mindset of doing the minimum to satisfy data-sharing policies and avoiding penalties. We showcase examples of successful data sharing from three invasive human electrophysiology research groups within the BRAIN Initiative Research Opportunities in Humans (ROH) Consortium (herein referred to as the "ROH Consortium"). For each case, we describe and categorize strategies used to facilitate FAIR data sharing—for instance, the use of standardized file formats—and summarize how others have reused the data, such as in education (Table 1). Finally, we emphasize how sharing has benefited data generators and offer recommendations to help maximize these benefits for generators and users.

Members of the ROH Consortium research groups contributed these case studies. We aimed to capture known examples of reuse for data generated by each group; however, they are not exhaustive, and we expect the benefits of FAIR sharing to expand as more data are shared and our fundamental understanding of the human brain unfolds.

CelPress

Neuron
NeuroView

Table 1. Major characteristics of three successful data-sharing processes developed for invasive human electrophysiology within the NIH Research Opportunities in Humans (ROH) Consortium

	Restoring Active Memory (RAM) project: Michael Kahana Laboratory, University of Pennsylvania, and partners	Memory Intracranial Neural Dynamic (MIND) project: Ueli Rutishauser Laboratory, Cedars-Sinai Medical Center, and partners	High-resolution single-unit recordings of cortical neurons dataset: Laboratories at Harvard Medical School and Massachusetts General Hospital
Population	patients with refractory epilepsy	patients with refractory epilepsy	participants undergoing clinical intraoperative physiological recordings
Data type(s)	• electrocorticographic recordings • patient demographics • individual electrode contact atlas location and coordinates for localization • FreeSurfer files • D-cortical surface renderings • seizure onset zones • interictal spiking • behavioral event data for ten different memory tasks • open- and closed-loop brain stimulation tasks	• Electrophysiological recordings (single neurons and local field potentials) • patient demographics • behavior • task stimulus • task event timestamps • electrode locations	• high-resolution laminar recordings
Format(s)	BIDS	NWB	• SpikeGLX • OpenEphys
Archive(s)	University-hosted webpage	DANDI DABI	Dryad
Strategies for data sharing	• standardized file formats	• standardized file formats • outreach to unique users • publication and use of open-source software	• standardized file formats • publication and use of open-source software
Selected ways data have been reused	• education, training • analytical tool development	• outreach, education, and training • analytical tool development • scientific discovery (research)	• analytical tool development • raw data export and analysis
Benefits for data generator	• early-stage investigator training and career advancement • reanalysis of existing data • facile dataset curation	• resolution of technical issues • enhanced lab productivity • identification of collaboration opportunities • facilitates reanalysis of existing data • use in teaching	• stimulated interest in intra-species comparison of brain • multidisciplinary collaboration

Penn Restoring Active Memory project: Michael Kahana laboratory, University of Pennsylvania, and partners

Michael Kahana's laboratory at the University of Pennsylvania investigates the neural basis of human memory. Since 2010, the group has shared data from their multi-center brain recording studies of human memory through a university-hosted webpage, which links data and, often, analysis code to host data files and published papers. In 2014, Kahana's team completed several public, large-scale data releases. Data were released across eight academic medical centers participating in the Restoring Active Memory (RAM) project, with funding from the BRAIN Initiative (https://memory.psych.upenn.edu/RAM).

Process of sharing the data

Data release from the RAM project included annotated data from more than 400 neurosurgical patients undergoing intracranial electrode recording for seizure mapping. These data included electrocorticographic (ECoG) recordings from >1,700 experimental sessions, mostly involving memory experiments and/or brain stimulation. The shared metadata include demographic information, individual electrode contact atlas location and coordinates for localization, FreeSurfer files, cortical surface renderings, seizure onset zone, interictal spiking, experiment design documents, session notes and behavioral event data for multiple memory tasks, and open- and closed-loop brain stimulation tasks.

All investigators on the RAM team obtained informed consent from patients to share their de-identified data. Kahana's group also released a set of Python tools (CMLreaders) to facilitate access to data collected using multiple recording systems. Kahana's group has begun converting the RAM data for release in Brain Imaging Data Structure (BIDS) format (https://bids.neuroimaging.io/), which was originally developed for MRI as part of OpenNeuro (https://openneuro.org/), one of the BRAIN data archives. In 2019, BIDS was extended to other data types, including EEG and iEEG.[5]

Converting the RAM data into the iEEG-BIDS format required several weeks' effort by an experienced data scientist. Kahana's group also has recently uploaded over 7,000 h of scalp EEG and memory tasks in BIDS to OpenNeuro. The considerable investment of time and expertise required to convert data to the BIDS standard has been discussed in other contexts.[6] Yet, the Kahana laboratory's approach to sharing employing this format has led to demonstrable advantages, as noted below.

How others have reused the data and the benefits to data generators

The RAM project exemplifies how sharing of invasive human electrophysiological data may be achieved on a large scale, despite the high costs of collection and the analytical complexity of the data. More than a dozen peer-reviewed papers on the electrophysiology of human memory using publicly available data curated by the Kahana lab have been published. Many authors of these papers did not have any affiliation with the RAM research effort. Moreover, the research reported in these papers likely would not have been pursued by the investigators on the RAM team. Thus, here the act of data sharing inspired new research.

Through more than a decade of sharing, no one in the Kahana group has ever been "scooped" by another group working with shared data. Instead, data sharing has motivated other groups to pursue novel analyses of existing datasets. Additionally, the large size and volume of the datasets shared have helped the Kahana group overcome curation challenges associated with sharing other datasets.

Memory Intracranial Neural Dynamic project: Ueli Rutishauser laboratory, Cedars-Sinai Medical Center, and partners

Our second case study involves Ueli Rutishauser's laboratory at Cedars-Sinai Medical Center (https://www.cedars-sinai.edu/research/labs/rutishauser.html), which leads the Memory Intracranial Neural Dynamic (MIND) project that also includes several academic partners. Specifically, the Rutishauser group studies mechanisms of learning, memory, and decision making, placing particular focus on the human brain at the single-neuron level. Generated data include the activities of neurons recorded using depth electrodes in patients with intractable epilepsy.

Process of sharing the data

Members of the MIND project on human episodic memory use the Neurodata Without Borders (NWB) format to structure their data.[7] NWB data files are hosted on Distributed Archives for Neurophysiology Data Integration (DANDI), another BRAIN Initiative-funded archive (https://www.dandiarchive.org/). Alongside recorded electrophysiological data, NWB files include metadata related to data acquisition, task stimuli, event timestamps, participant demographics, behavioral responses, and Montreal Neurosurgical Institute coordinates of recording electrodes. The result is a comprehensive NWB file containing all data and metadata needed to analyze and reuse patient data by protocol type.

How others have reused the data and the benefits to data generators

Several major reuse cases of MIND project data have included activities related to education, training, and tool development. For example, in 2022, datasets from the group were used at the Allen Institute's NeuroDataReHack Hackathon (https://alleninstitute.org/events/2022-neurodatarehack-hackathon/), including activity data recorded from 1,863 neurons in the medial temporal lobe across 59 human subjects with intractable epilepsy. The International Neuroinformatics Coordinating Facility (INCF)/MATLAB Community Toolbox Training Project used another Rutishauser lab-generated dataset to teach users how to employ the NWB Application Programming Interface. The standardized datasets also have doubled as an educational resource for unique users. For example, children ages 8 to 15 years reviewed the data using NWB-based graphical user interfaces and helped translate the original publication of the dataset into a version for *Frontiers for Young Minds*.[8] The interactive, web-based viewing, made possible by DANDI, enabled readers with limited programming experience to visualize the structured data and understand key findings from the parent study.

Additionally, conducting data releases has been a valuable experience for those within the laboratory. While preparing single-neuron data in the NWB format, numerous subtle inconsistencies and technical issues in data-to-be-shared have been discovered, often with the

help of the automatic validators provided by NWB and DANDI. These errors were subsequently resolved and properly documented. Sometimes, these issues were not apparent in initial analyses. Thus, without the rigor of standardization for sharing, errors would have remained undetected. Second, generating standardized datasets has promoted continuity of laboratory productivity and data preservation. New laboratory members, interns, rotation students, and students conducting class projects have been easily able to reuse datasets generated by others. Third, the increased visibility and accessibility conveyed by standardized, publicly available datasets have led to other research groups approaching the laboratory to initiate collaborations. As a result, the Rutishauser team has discovered other scientists with similar interests in diverse fields.

High-resolution single-unit recordings of cortical neurons dataset: Laboratories at Harvard Medical School and Massachusetts General Hospital

A collaborative group at Massachusetts General Hospital (MGH), which includes the laboratories of Ziv Williams, Sydney Cash, and Angelique Paulk at Harvard Medical School (https://zivwilliams.mgh. harvard.edu/), is our final case study. This group has been focusing on using single-unit recordings of cortical neurons to study human cognitive processes at a cellular level.[9] Unlike prior projects focused on generating large datasets, this project generates smaller datasets through the implementation of Neuropixels (https:// www.neuropixels.org/), ultra-high density, fully integrated linear silicon microprobes. This technology enables acute, high-resolution laminar recordings from cortical neurons in participants undergoing clinical intraoperative neurophysiological recordings. As this approach is still nascent and includes a small number of participants (n < 30), open access to and dissemination of the data have been important priorities for the development of the field.

Process of sharing the data

The MGH team has developed a pipeline for de-identifying data, making it freely available on Dryad (https://datadryad. org/stash/dataset/doi:10.5061/dryad.d254

7d840) and DANDI (https://dandiarchive. org/dandiset/000397). The team members also place constraints on how their data may be used, consistent with the ethical guidelines for invasive human neurophysiology recently issued by the ROH Consortium.[10] As described in further detail elsewhere,[9] the raw data are re-saved in a way that ensures the surgical date cannot be traced back to individual participants. The metafiles also are edited to ensure that no dates are saved, and file names are re-coded such that there are no patient identifiers.

Finally, associated imaging scans are defaced using a manual software enabling the removal of participants' facial features on MRI while preserving their anatomical brains. Once shared online, detailed descriptions are linked to the recorded data as well as to the open-source software and codes used to process the recordings. For example, direct links are provided to the SpikeGLX, Open Ephys, and Probe map export functions, which can allow the raw data to be easily explored. Finally, access to the motion correction software used to optimize signal registration is provided (https:// github.com/evarol/dredge). These codes are paired with the recording data in a standardized (.bin) file format, which can be read by various programming languages as metadata text files on Dryad and DANDI.

How others have reused the data and the benefits to data generators

Based on total downloads/hits on Dryad and DANDI from unique users and via direct invitations received to collaborate on new projects, it has become clear that the collaborative group pioneering work on high-resolution, single-unit recordings has catalyzed activity around the shared datasets, including for comparative analyses. These new techniques and approaches used to address human cortical recordings have generated interest from outside groups performing animal studies using similar high-density microprobes, for example. In addition, others have connected with the group to explore the cell-type identities of cortical cells and compare the extracellular action potential shapes of cortical neurons across animals and humans. Overall, the group anticipates that freely opening access to data

using FAIR practices will continue to enhance not only the development of collaborative opportunities but also encourage feedback from others on how to optimize recording and analytic techniques.

Recommendations

The case studies from the BRAIN Initiative ROH Consortium presented demonstrate that shared, intracranial human electrophysiology data can be (and are) reused and that the process of sharing FAIR data has tangible benefits for generators alongside users (Table 1). Here, we highlight three benefits of sharing for data generators. To promote sharing in ways that maximize these benefits for data generators and users, we also provide three recommendations.

First, *workflows for standardized data generation promote continuity of laboratory productivity and data preservation*. Fewer data are lost during laboratory turnovers, and members are able to work more efficiently. The enormous efforts expended by the ROH groups to transform data into standardized formats also demonstrate the critical need for new software tools that ease the burdens of generating data and making it more useful by collating metadata, reformatting existing data, uploading it to archives, or facilitating secondary exploration. Nevertheless, the benefits of having individuals and teams in place with the expertise and experience necessary for preparing and managing data should not be underestimated. Thus, when preparing grant applications, we recommend that *researchers include dedicated resources in their budgets for data-sharing training and labor*, including for data scientists and programmers who can help manage data-sharing requirements. The frequency of uploading shared data will vary with the pace and volume of each project, but we recommend a minimum frequency of approximately every 6 months to ensure consistency of effort. The NIH also provides general guidelines for DMS budgeting (https://sharing. nih.gov/data-management-and-sharing-policy/planning-and-budgeting-for-data-management-and-sharing/budgeting-for-data-management-sharing).

Second, *preparing invasive human electrophysiology data for sharing can reveal errors in data or code, creating*

opportunities to revise these errors. This is in large part because standardized file formats and associated archives (e.g., NWB and DANDI) add checks for internal consistency. Sharing open-source acquisition and analysis code also allows users to plot simple analyses as entry points into datasets, inspiring more sophisticated analyses later. Thus, *sharing data using standardized formats with rich metadata and, where possible, sharing open-source acquisition and analysis code* is as important as sharing parent datasets. Even widely used data formats (i.e., NWB and BIDS) presently have little interoperability with one another, which is another challenge that requires further work. Yet, labs still should consider (even if not obligated to do so by funders) releasing fully documented and standardized versions of datasets as soon as possible after generation and after data have been appropriately protected (i.e., removal of personal identifiers). The NIH does not require that data be shared until the first publication of a dataset or the end of a funding period (whichever comes first). Yet, a common concern about "scooping" remains wherein a secondary user of data publishes key findings from a dataset before the producer of that dataset has the opportunity. Our small number of case studies presents no evidence that this has occurred, likely in part because considerable analysis of iEEG data is required to reach standards for publication. Nevertheless, this remains a valid concern, and investigators in the field will need to feel like the benefits of sharing outweigh the risks in order for this behavior to become self-sustaining.

Third, *scientists who share data for reuse or reanalysis in educational settings can increase visibility, whether for datasets or for themselves and other laboratory members*. Laboratories organized around preparing FAIR data can train current and future students on how to reuse them and, in turn, on how to reuse others' data. Thus, when preparing shared datasets, we recommend that researchers also *consider employing them for educational purposes, for instance, as training datasets or in hackathons*, or in outreach efforts to unique users such as young learners. Such work has a cascading effect, raising awareness for research and promoting further opportunities. Re-

searchers who share their data with the Allen Institute for Brain Science are automatically eligible for its open data science symposia, for example, catalyzing activity around shared data resources. The INCF also sponsors free online courses for data reuse, with periodic updates in line with new computational tools (https://www.incf.org/about/what-we-do). In sum, FAIR data sharing can expand the neuroscience workforce and open opportunities for collaboration. This amplification of both expertise and resources increases return on investment in data generation, especially in the longer term.

Finally, while the cases presented in this NeuroView emphasize the feasibility and benefits of data sharing for invasive human electrophysiology, participants must always be appropriately consented and data reused according to their informed choices. Thus, investigators should provide as much transparency as possible in terms of where and how the data will be shared, the relevant benefits and associated risks of sharing, and who will likely have access to data that are shared, consistent with the most up-to-date ethical considerations and recommendations.[10]

Conclusions and future directions

Data sharing in neuroscience can have many benefits. While only a fraction of the human neural data generated so far have been shared, this situation is likely to change in the coming years, as the NIH-wide DMS policy increases sharing across all fields. Here, we profiled successful data-sharing processes for invasive human electrophysiology from three teams in the BRAIN Initiative ROH Consortium. We highlighted three tangible benefits of sharing FAIR data for scientists generating these data. Finally, to further promote FAIR sharing, we provided three recommendations to help maximize these benefits.

CONSORTIA

The members of the NIH Research Opportunities in Humans (ROH) Consortium are Ralph Adolphs, Richard A. Andersen, Gordon Baltuch, Peter Brunner, Sydney S. Cash, Edward Chang, Jennifer L. Collinger, Nathan Crone, Evelina Fedorenko, Itzhak Fried, Josh Gold, Jaimie Hender-

son, Leigh Hochberg, Matthew Howard, Michael J. Kahana, John Magnotti, Adam Mamelak, Nader Pouratian, Robert Mark Richardson, Ueli Rutishauser, Gerwin Schalk, Charlie Schroeder, Krishna Shenoy, Sameer A. Sheth, Nanthia Suthana, Nitin Tandon, Ziv M. Williams, and Jonathan Wolpaw.

AUTHOR CONTRIBUTIONS

V.R., K.M.J., and M.M. prepared the manuscript for submission and drafted, edited, and finalized the introduction, recommendations, and conclusions and future directions. With M.J.K., U.R., J.Z., Z.M.W., S.S.C., and A.C.P., they also compiled, edited, and finalized the case studies. V.R. and K.M.J. created the table. All other authors significantly contributed to the content and perspectives and/or edited the manuscript. All authors approved the final version of the manuscript.

DECLARATION OF INTERESTS

N.P. is a consultant for Abbott and Sensoria Therapeutics. S.A.S. is a consultant for Boston Scientific, Zimmer Biomet, Neuropace, Sensoria Therapeutics, Koh Young, and Varian Medical. He is co-founder of Motif Neurotech.

REFERENCES

1. White, T., Blok, E., and Calhoun, V.D. (2022). Data sharing and privacy issues in neuroimaging research: Opportunities, obstacles, challenges, and monsters under the bed. Hum. Brain Mapp. *43*, 278–291. https://doi.org/10.1002/hbm.25120.

2. Hendriks, S., Ramos, K.M., and Grady, C. (2022). Survey of Investigators About Sharing Human Research Data in the Neurosciences. Neurology *99*, e1314–e1325. https://doi.org/10.1212/WNL.0000000000200886.

3. Elam, J.S., Glasser, M.F., Harms, M.P., Sotiropoulos, S.N., Andersson, J.L.R., Burgess, G.C., Curtiss, S.W., Oostenveld, R., Larson-Prior, L.J., Schoffelen, J.-M., et al. (2021). The Human Connectome Project: A

CellPress

retrospective. Neuroimage *244*, 118543. https://doi.org/10.1016/j.neuroimage.2021.118543.

4. Poline, J.-B., Kennedy, D.N., Sommer, F.T., Ascoli, G.A., Van Essen, D.C., Ferguson, A.R., Grethe, J.S., Hawrylycz, M.J., Thompson, P.M., Poldrack, R.A., et al. (2022). Is Neuroscience FAIR? A Call for Collaborative Standardisation of Neuroscience Data. Neuroinformatics *20*, 507–512. https://doi.org/10.1007/s12021-021-09557-0.

5. Holdgraf, C., Appelhoff, S., Bickel, S., Bouchard, K., D'Ambrosio, S., David, O., Devinsky, O., Dichter, B., Flinker, A., Foster, B.L., et al. (2019). iEEG-BIDS, extending the Brain Imaging Data Structure specification to human intracranial electrophysiology. Sci.

Data *6*, 102. https://doi.org/10.1038/s41597-019-0105-7.

6. Bush, K.A., Calvert, M.L., and Kilts, C.D. (2022). Lessons learned: A neuroimaging research center's transition to open and reproducible science. Front. Big Data *5*, 988084. https://doi.org/10.3389/fdata.2022.988084.

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

8. Zheng, J., Chen, G., Kreiman, G., and Rutishauser, U. (2023). Little Memory Editors Living Inside Your Brain. Front. Young Minds

11, 968856. https://doi.org/10.3389/frym.2023.968856.

9. Paulk, A.C., Kfir, Y., Khanna, A.R., Mustroph, M.L., Trautmann, E.M., Soper, D.J., Stavisky, S.D., Welkenhuysen, M., Dutta, B., Shenoy, K.V., et al. (2022). Large-scale neural recordings with single neuron resolution using Neuropixels probes in human cortex. Nat. Neurosci. *25*, 252–263. https://doi.org/10.1038/s41593-021-00997-0.

10. Feinsinger, A., Pouratian, N., Ebadi, H., Adolphs, R., Andersen, R., Beauchamp, M.S., Chang, E.F., Crone, N.E., Collinger, J.L., Fried, I., et al. (2022). Ethical commitments, principles, and practices guiding intracranial neuroscientific research in humans. Neuron *110*, 188–194. https://doi.org/10.1016/j.neuron.2021.11.011.