

Data Management and Sharing Plan

Element 1: Data Type:

A. Types and amount of scientific data expected to be generated in the project:

Summarize the types and estimated amount of scientific data expected to be generated in the project.

The study will generate multimodal data from 50 to 80 mouse subjects, combining electrophysiological, behavioral, and imaging modalities to investigate hippocampal function. The expected data types and estimated volumes are as follows:

Electrophysiology data: Neural recordings will be collected using 64-channel extracellular electrode probes targeted to the mouse hippocampus. Data will include both raw extracellular voltage recordings and processed spike time data. The estimated data volume to be shared is 100-1000 GB.

Behavioral data: Spatial navigation will be recorded concurrently with electrophysiological activity. These data include raw video frames of freely moving mice within an arena, along with processed positional and heading data for each frame. The estimated data volume is 10-100 GB.

Imaging data: Confocal microscopy will be used to capture 3-color image stacks of immunostained mouse brain tissue, verifying probe placement in the hippocampus. The estimated data volume is 110 GB.

B. Scientific data that will be preserved and shared, and the rationale for doing so:

Describe which scientific data from the project will be preserved and shared and provide the rationale for this decision.

We will share all data generated during the project period except any data generated during protocol optimization and/or training of lab personnel that do not meet NIH's definition of scientific data (i.e., data that are not of "sufficient quality to validate or replicate research findings").

C. Metadata, other relevant data, and associated documentation:

Briefly list the metadata, other relevant data, and any associated documentation (e.g., study protocols and data collection instruments) that will be made accessible to facilitate interpretation of the scientific data.

To ensure proper interpretation, reuse, and reproducibility of the scientific data, comprehensive metadata and documentation will be made available alongside each dataset. General metadata for all data types will include details such as animal strain, sex, date of birth, age at the time of the experiment, laboratory and institution information, and drug treatment status, including concentration and route of administration. For electrophysiology data, additional metadata will capture electrode number and position, reference electrode type and location, equipment models, and recording parameters such as sampling rate and filter cut-offs. Behavioral data will be accompanied by metadata on equipment models (e.g., arena and camera), along with camera specifications including frame rate and resolution. Imaging data will include information on the targeted brain region, magnification, resolution, and color channels used. In addition to these metadata elements, associated documentation such as study protocols, data collection instruments, and standard operating procedures will be provided to support transparency and enable effective data reuse.

Element 2: Related Tools, Software and/or Code:

State whether specialized tools, software, and/or code are needed to access or manipulate shared scientific data, and if so, provide the name(s) of the needed tool(s) and software and specify how they can be accessed.

All data files will be readable using either the PyNWB or MatNWB application programming interfaces (APIs), which require use of the Python or MATLAB programming languages. The APIs can be freely downloaded by following instructions on the Neurodata Without Borders website. Python is also freely available for download online. MATLAB requires a paid license from its owner, MathWorks. We will use Python to write custom analysis and/or visualization code that processes data from the format in which it is shared (see below), and we will share any such custom code on GitHub. The location of this code on [Code Sharing Repository X] will be provided as additional metadata in the repository we have selected for data sharing (see below).

Element 3: Standards:

State what common data standards will be applied to the scientific data and associated metadata to enable interoperability of datasets and resources and provide the name(s) of the data standards that will be applied and describe how these data standards will be applied to the scientific data generated by the research proposed in this project. If applicable, indicate that no consensus standards exist.

The Neurodata Without Borders (NWB) data standard will be applied to all electrophysiological, behavioral, and immunohistochemical data designated for sharing. For each experimental session, all raw and processed electrophysiological, behavioral, and imaging data and metadata associated with the session will be written hierarchically into a single .nwb file.

Element 4: Data Preservation, Access, and Associated Timelines:

A. Repository where scientific data and metadata will be archived:

Provide the name of the repository(ies) where scientific data and metadata arising from the project will be archived.

All electrophysiological, behavioral, and immunohistochemical data and associated metadata will be archived in Distributed Archives for Neurophysiology Data Integration (DANDI), a data sharing platform specialized for hosting data conforming to the NWB data standard.

B. How scientific data will be findable and identifiable:

Describe how the scientific data will be findable and identifiable, i.e., via a persistent unique identifier or other standard indexing tools.

Upon submission to DANDI, each shared dataset will be assigned a persistent unique identifier, which will be provided in any publication(s) corresponding to the dataset.

C. When and how long the scientific data will be made available:

Describe when the scientific data will be made available to other users (i.e., no later than the time of an associated publication or end of the performance period, whichever comes first) and for how long data will be available.

Scientific data will be uploaded to DANDI upon acceptance of the first manuscript in which they are included or at the end of the performance period, whichever comes first. Shared scientific data will be available to other users indefinitely.

Element 5: Access, Distribution, or Reuse Considerations:

A. Factors affecting subsequent access, distribution, or reuse of scientific data:

NIH expects that in drafting Plans, researchers maximize the appropriate sharing of scientific data. Describe and justify any applicable factors or data use limitations affecting subsequent access, distribution, or reuse of scientific data related to informed consent, privacy and confidentiality protections, and any other considerations that may limit the extent of data sharing.

There are no factors affecting subsequent access, distribution, or reuse of scientific data generated by this project.

B. Whether access to scientific data will be controlled:

State whether access to the scientific data will be controlled (i.e., made available by a data repository only after approval).

Scientific data generated by this project will not be controlled (i.e., they will be openly accessible).

C. Protections for privacy, rights, and confidentiality of human research participants:

If generating scientific data derived from humans, describe how the privacy, rights, and confidentiality of human research participants will be protected (e.g., through de-identification, Certificates of Confidentiality, and other protective measures).

Not applicable; this project does not involve human research participants.

Element 6: Oversight of Data Management and Sharing:

Describe how compliance with this Plan will be monitored and managed, frequency of oversight, and by whom at your institution (e.g., titles, roles).

The PI will have direct responsibility for oversight of data management and sharing for the project. The PI will establish and conduct quarterly meetings with lab members to identify all data that has been generated during the past quarter and which data must be shared to comply with the approved Plan. During preparation of research manuscripts and/or 6 months in advance of the end of the project, the PI will oversee the preparation of datasets to be submitted to DANDI.