

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 project will generate three primary types of multimodal data from 50 to 80 mouse subjects:

- Electrophysiology data: Raw extracellular voltage recordings from 64-channel probes implanted in the mouse hippocampus, as well as processed spike time data. Estimated data volume: 100–1000 GB.
- Behavioral data: Raw video recordings of mice performing spatial navigation tasks, alongside processed positional and heading data extracted from video frames. Estimated data volume: 10–100 GB.
- Imaging data: Confocal microscopy 3-color image stacks of immunostained mouse brain tissue for verification of probe placement. Estimated data volume: 1–10 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.

All raw and processed data from the above modalities will be preserved and shared, including (a) unprocessed extracellular voltage traces, (b) spike-sorted data, (c) raw video files and extracted position/heading data, and (d) confocal imaging stacks. Sharing these data supports reproducibility, enables meta-analyses, and allows re-use for further studies of hippocampal function and methodology validation. Only data directly supporting published findings will be prioritized for sharing, but all data necessary for independent replication of results will be made publicly available.

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.

The following metadata and documentation will be made accessible:

Subject-level metadata (e.g., mouse strain, age, sex, experimental group, dates of recording)

Detailed experimental protocols (surgical procedures, behavioral task descriptions, imaging protocols)

Data acquisition parameters (sampling rates, hardware specifications, electrode/probe maps)

Data processing and analysis pipelines (including code and workflow descriptions)

Data dictionaries and README files describing file formats, variable definitions, and folder structures

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.

Specialized tools are needed to access and analyze scientific data:

- Electrophysiology data: Open-source tools such as Spike Interface, Open Ephys GUI, and MATLAB scripts for spike sorting and visualization. Links to code repositories and instructions will be provided.

- Behavioral data: Python-based toolboxes (e.g., DeepLabCut for pose estimation), and standard video playback software (e.g., VLC).
 - Imaging data: Open-source software such as FIJI/ImageJ for confocal image visualization and analysis.
- All custom code and analysis pipelines will be shared via GitHub repositories with documentation.

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 following data and metadata standards will be applied for interoperability:

- Electrophysiology data: Data will be stored in Neurodata Without Borders (NWB: N) format, an established standard for neurophysiology data.
- Behavioral data: Video files will be provided in standard MP4 format; positional/heading data in CSV or HDF5 files, following the NWB: N behavioral extension when possible.
- Imaging data: Confocal image stacks will be stored in OME-TIFF format, a widely adopted standard for microscopy data.
- Metadata: Metadata will be structured according to the NWB: N schema and the Minimum Information for Biological and Biomedical Investigations (MIBBI) guidelines where applicable.

Documentation will adhere to best practices for reproducibility and transparency in preclinical neuroscience research.

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.

- Electrophysiology and behavioral data will be deposited in the DANDI Archive (<https://dandiarchive.org>), which supports NWB: N and is suitable for large-scale neuroscience datasets.
- Imaging data will be archived in the Brain Image Library (BIL, <https://www.brainimagelibrary.org>) or the BioImage Archive (<https://www.ebi.ac.uk/bioimage-archive/>), both of which are appropriate for microscopy data.
- All associated code and documentation will be deposited in a public GitHub repository and linked to Zenodo for DOI assignment.

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.

All datasets will be assigned persistent unique identifiers (DOIs) upon deposition in the respective repositories. Metadata records will be indexed and searchable via repository platforms, and cross-referenced with related publications and project web pages.

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 made available no later than the time of publication of the main research findings or at the end of the project period, whichever comes first. Data will remain publicly accessible via the chosen repositories for a minimum of 10 years, in accordance with repository policies and NIH requirements.

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.

The data are derived exclusively from non-human animal (mouse) research and do not contain sensitive or personally identifiable information. No special restrictions on access, distribution, or reuse are anticipated. Users will be encouraged to cite the dataset DOI and adhere to the repository's terms of use and community standards for attribution.

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).

Access to all shared data will be open and unrestricted; no controlled access is required.

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. No human subjects are involved in this research.

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).

Compliance with this Data Management and Sharing Plan will be overseen by the Principal Investigator (PI) in coordination with the laboratory's Data Steward. The PI will be responsible for ensuring timely deposition of data and documentation prior to publication or project closeout. Quarterly reviews will be conducted by Data Steward to monitor progress, verify data integrity, and confirm adherence to repository requirements. The institution's Office of Research Compliance will provide additional oversight as needed.