Plan Overview

A Data Management Plan created using DMP Tool

Title: Building an integrated framework to understand neurodevelopmental disorders

Creator: Stefan Lemke

Affiliation: University of North Carolina at Chapel Hill (UNC-CH) (unc.edu)

Funder: National Institutes of Health (nih.gov)

Funding opportunity number: PA-20-188

Grant: https://grants.nih.gov/grants/guide/pa-files/PA-20-188.html

Template: NIH-Default DMSP

Project abstract:

Genetically distinct neurodevelopmental disorders (NDDs) are typically studied in isolation. However, many share clinical features including learning disability, behavioral inflexibility, and impaired motor skills. Understanding NDDs in a shared context, rather than as unique conditions, would reveal whether these impairments are associated with common or unique neural dysfunctions. Such insights are crucial to design therapeutic approaches as the prevalence of NDDs is rising and disease-modifying therapies remain rare.

Start date: 09-01-2024

End date: 08-31-2029

Last modified: 07-08-2024

Copyright information:

Building an integrated framework to understand neurodevelopmental disorders

Data Type

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.*

Describe data in general terms that address the type and amount/size of scientific data expected to be collected and used in the project (e.g., 256-channel EEG data and fMRI images from ~50 research participants). Descriptions may indicate the data modality (e.g., imaging, genomic, mobile, survey), level of aggregation (e.g., individual, aggregated, summarized), and/or the degree of data processing that has occurred (i.e., how raw or processed the data will be)

This project will produce data of the following formats: videos of animal behavior (5TB), electrophysiology data from the brain (40TB), histology imaging (4TB), and computational analysis results (1TB). This data will be collected from 10 mouse models of neurodevelopmental disorders and wildtype littermates. We estimate a total of 50TB of data.

We expect to generate the following data formats: .AVI video files, IMEC format raw electrophysiology data, .CZI Carl Zeiss microscopic image files, and .MAT MATLAB files.

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.*

For behavior data, the 3D position of relevant body parts derived from analysis of video will be preserved and shared. For electrophysiology, spike times, firing rates, and z-scored activity for individual neurons surrounding the behavioral task under study will be preserved and shared. All histological and analysis files will be preserved and shared. Metadata related to the trial structure and animal behavior during the experiment will be clearly listed within the processed electrophysiology data files. I will use Wasabi for long-term storage and DANDI for long-term data archiving and sharing, of raw video and electrophysiology files totaling ~50TB.

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.

All methods and protocols and specialized equipment associated with data collection and analysis will be described in detail in the published documentation and preserved and shared. In addition, all metadata related to the experimental animal and trial structure (name, date, weight, genetic strain, trial type, labels and times of experimental variables) and behavioral labels will be clearly listed and included within the shared computational analysis files.

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.

Matlab and/or Python software is required to access and manipulate our shared scientific time series data. Electrophysiology experiments require Matlab. Image analysis software, such as ImageJ can be used to access and manipulate our histology imaging data. Analysis code used to produce the documented results will be preserved and shared.

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

For our combined behavior and electrophysiology datasets, no consensus data storage standards exist. Data acquired by all participants will include a master metadata file (excel document) that lists all experimental animals by name, sex type, date of birth, and their associated group labels. Each experimental animal name is associated with a data folder containing subfolders for behavior/video, electrophysiology, and histology or imaging data. Within each experimental animal data folder is another metadata excel document that lists and describes the experiment-specific details (date, weight, surgery notes, experimental protocol and behavior notes) for each experiment for that animal name. To assist in the reuse of the datasets by naive investigators, all brain and behavior time series data will be interpolated to a common time series (1ms bins, from 5 seconds before to 10 seconds after trial onset). To enable data sharing in non-proprietary formats, data will also be made available as .csv files.

Data Preservation, Access, and Associated Timelines

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 dataset(s) that can be shared will be deposited in the DANDI Archive.

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.

To find experiments and data of interest, a naive investigator will start by opening the master metadata excel document that lists all experimental animals by name, sex type, date of birth, and their associated experiment group labels. Each experimental animal name is associated with a data folder containing subfolders for behavior/video, electrophysiology, and histology data. Within each experimental animal data folder is another metadata excel document that lists and describes the experiment-specific details (date, weight, surgery notes, experimental protocol and behavior notes) for each experiment for that animal name. For all experiments, time series data (video, electrophysiology) is interpolated and stored with common time series binning and labeling guidelines. Moreover, the shared analysis code will recapitulate the results and therefore point the investigator to the location of the relevant data. Additionally, DANDI assigns persistent identifiers. Inclusion of these identifiers in publications will enable the identification of datasets used in publications.

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 time of an associated publication or end of the performance period, whichever comes first) and for how long data will be available.

Scientific data and metadata will be made available immediately upon publication (including pre-prints) or at the end of the performance period, and will be made available indefinitely, as funding permits, and a minimum of 10 years after the funding period.

Access, Distribution, or Reuse Considerations

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 anticipated factors or limitations that will affect the access, distribution or reuse of the scientific data generated by the proposal.

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 our scientific data repositories will be freely available and not controlled.

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).

N/A

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 plan will be monitored and managed by the principal investigator Stefan Lemke and primary mentor Adam Hantman. Oversight of the plan will occur throughout the study, as well as at specific project milestones (onset, data collection, organization and analysis, presentation, data archiving, etc.).

Planned Research Outputs

Dataset - "Electrophysiological, Behavioral, and Histological Dataset"

Planned research output details

Title	Type	Anticipated release date	Initial access level	Intended repository(ies)	Anticipated file size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
Electrophysiological, Behavioral, and Histological ...	Dataset	Unspecified	Open	DANDI		None specified	None specified	No	No