DATA MANAGEMENT AND SHARING PLAN

An example from an application proposing to collect single cell genomic data from mice and humans.

If any of the proposed research in the application involves the generation of scientific data, this application is subject to the NIH Policy for Data Management and Sharing and requires submission of a Data Management and Sharing Plan. If the proposed research in the application will generate large-scale genomic data, the Genomic Data Sharing Policy also applies and should be addressed in this Plan. Refer to the detailed instructions in the application guide for developing this plan as well as to additional guidance on sharing.nih.gov. The Plan is recommended not to exceed two pages. Text in italics should be deleted (but this has not been done in the sample below). There is no "form page" for the Data Management and Sharing Plan. The DMS Plan may be provided in the *format* shown below.

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.

As detailed in the Research Strategy Section, we propose the generation of a spatially mapped single-cell atlas of the developing mouse brain and include specific deliverables. Our primary deliverable for each modality will be a matrix of cells × (counts in peaks for ATAC, UMIs in genes for RNA, or methylation status for DNAm) along with a dense metadata table with information for each cell. This includes the animal sex, developmental time point, punch of origin with x,y,z coordinates, assigned cluster and inferred cell type, assigned subcluster and inferred cell type, as well as a number of QC metrics (total reads, passing reads, reads in peaks, TSS enrichment, cell barcode combination, date of preparation for each stage, sequencing platform, likelihood of being a doublet, and any other relevant metrics that arise during the project).

The amount and type of data from human cells will depend on the results from the mouse studies. Data sharing plans will be updated when appropriate (likely at the start of year 4 of the grant award).

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.

The data described in section A will allow researchers to reproduce our publications and will allow them to collect additional data in a similar way to extend our results.

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.

In addition to a detailed methods section for any publications associated with this work, we will provide a detailed step-by-step protocol as a Supplementary Protocol document and maintain active protocols.io protocols for each technology and workflow.

We will additionally release protocol links as metadata to be associated with single-cell data deposited to the Neuroscience Multi-omic Archive (NeMO, https://nemoarchive.org/).

In addition to providing detailed protocols, our laboratory has hosted visiting scientists to train on the data analysis pipelines developed and deployed by the lab. We welcome the opportunity to continue these training efforts.

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 code and software that will be written to analyze the data will be deposited on GitHub (https://github.com/labname) for public access and be provided as Supplementary files for any publications. Code will be available no later than when a publication has been submitted.

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.

We will use the standards that are adopted or defined by NeMO.

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; see Selecting a Data Repository).

Mouse single-cell datasets: All single-cell epigenomics and transcriptomics data will made available through NeMO after initial data processing.

Human single cell data will be deposited to NeMO.

Upon publication we will host processed data matrixes and associated metadata as compressed downloadable archives at NeMO and, when appropriate, as supplementary information in journal publications.

We will release datasets associated with the technological advances proposed in the application once protocols are established and initial analysis performed, at which point data will be released along with a preprint prior to manuscript submission.

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.

Data will be findable for the research community through searches at NeMO. NeMO assigns unique identifiers for each sample.

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

The research community will have access to data as soon as NeMO is able to release it. NeMO will control the deletion of the data sets.

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. See Frequently Asked Questions for examples of justifiable reasons for limiting sharing of data.

There are no special considerations related to accessing or distributing the mouse data to be generated in this award.

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 the human data sets in NeMO is controlled. The NIMH Data Archive Data Access Committee serves as the data access committee for NeMO.

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).

The data are deidentified, and there is data access committee (see 5B) that is used by NeMO to evaluate the proposed use of the sensitive data from human subjects. The data archives which will hold the data are funded by NIH, so they have a Certificate of Confidentiality.

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 Office of Sponsored Programs at University X that will be administering this award has created a data management and sharing plan compliance system as part of their process for submitting the annual NIH progress report. That Office will be monitoring submission of data to NEMO.

Validation Schedule (this section is required by NIMH)

Data will be validated using the existing pipelines at NeMO. We will submit each dataset once we reach a specific data freeze milestone. Upon each data freeze we will perform an initial phase of analysis that will culminate in the production of the cell × property matrix and associated metadata, at which point the dataset will be released.

These milestones and target timelines include:

End of 1st quarter of year 2: A full spatial single-cell ATAC-seq map of an entire mouse brain at P14.

End of 4th quarter of year 2: An accompanying spatially-mapped single-cell RNA-seq dataset for a full mouse brain at P14, integrated with the ATAC dataset.

End of 2nd quarter of year 3: A full spatial single-cell ATAC-seq map of each time point.

End of 2nd quarter of year 3: A full spatial single-cell DNA methylation map for P14, integrated with RNA and ATAC datasets.

End of 4th quarter of year 3: The complete spatial single-cell RNA-seq dataset for all time points, integrated with the ATAC data.

End of 4th quarter of year 3: A full spatial single-cell map for all modalities, ATAC, RNA, and DNA methylation, integrated across modalities.