Caltech Library

# Data Management and Sharing Plan Example

*This example is based on the generic NIH data management and sharing plan template. Always refer to the specific guidance for your grant proposal and ICO when writing a data management plan, as requirements differ.*

## Element 1: Data Type

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

This project will collect microscopy images and video – including 2, 3, and 4 dimensional files – of fruit flies (*Drosophila melanogaster*). Most videos will capture neural activity in individual flies. Images will be acquired in the TIFF format. Individual files may be upwards of several hundred GB in size. We expect to acquire about 100 TB of content over the course of the grant period.

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

The project will share the images and videos associated with scientific articles at the time of article publication. Due to the size of the data, the project will also share limited other image and video data of high quality by the end of the grant period.

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

Images will be saved in the native file type of the microscope, Zeiss LSM, and converted to the OME-TIFF format using the BioFormats tool. OME-TIFF images use the OMEXML metadata standard to document images. We will additionally document our data in lab notebooks and digital README.txt files, which will also record relevant protocols.

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

The project will create software code in the Python programming language for the analysis of the microscopy image files. Code will be shared in the GitHub repository to maximize findability. Code will also be assigned a DOI and preserved in the CaltechDATA repository for long-term access.

Caltech holds the copyright in software generated under sponsored research projects, but software from this project will be made publicly available under a standard BSD-3-clause license. Use of standard licenses means that other users and machines have a clear understanding of the conditions of re-use for all data and software.

## Element 3: Standards:

Data will be collected and converted into OME-TIFF images, which use the OMEXML metadata standard. Shared data will leverage the DataCite metadata standard to improve findability.

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

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

# Data Management and Sharing Plan Example

All data from this project will be made available via the CaltechData repository. CaltechDATA acts as an institutional repository for data associated with Caltech, and is managed by Caltech Library. CaltechDATA is built on InvenioRDM, an open source data repository platform supported by over 20 institutional partners.

### B. How scientific data will be findable and identifiable:

Each dataset in CaltechDATA is assigned a DOI, which is a persistent identifier for that dataset and will resolve in perpetuity. CaltechDATA collects a large set of DataCite metadata to ensure that datasets are FAIR, and utilizes a wide range of persistent identifiers including ORCIDs for individuals and RORs for affiliations and funding agencies. Datasets in CaltechDATA will be findable via DataCite and other dataset search services such as the Web Of Science Data Citation Index and Mendeley Data.

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

Data will be made available at the time of associated publication or before the end of the grant period. CaltechDATA is intended as a long-term data repository but, due to the large size of data generated by this project, it is expected that data will be available for at least five years. All data is preserved using the OAIS reference model and multiple copies of all data sets are stored at the Caltech Library and remotely. The data upload charges of $ 300 / TB budgeted in this proposal covers cloud and local storage costs for five years. After five years Caltech Library will have the right to deaccession underutilized data. Caltech Library will make every effort to contact the PI and data creators before data is deaccessioned.

## Element 5: Access, Distribution, or Reuse Considerations

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

All publicly shared data will be licensed with a Creative Commons Zero (CC-0) license. This dedicates all data to the public domain and allows for unrestricted re-use. It is expected that users that employ a data set generated from this grant will provide a citation.

### B. Whether access to scientific data will be controlled:

Access to scientific data generate by this project do not need to be controlled.

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

There are no concerns about privacy, rights, or confidentiality for the data generated by this project.

## Element 6: Oversight of Data Management and Sharing:

The project PI, Dr. L. Morgan, will be responsible for the oversight and execution of this DMSP. The Caltech Library, via the CaltechDATA repository, will be responsible for the access and preservation of all shared datasets and code for at least five years upon the point at which they are made publicly available.