



**European Research Council (ERC)**

**ERC Data Management Plan**

**Template**

**ERC OPEN RESEARCH**

**DATA MANAGEMENT PLAN (DMP)**

# 

|  |  |
| --- | --- |
| **Project Acronym** | **Project Number** |
| ACRONYM |  |

***Template for the ERC Open Research Data Management Plan (DMP). The following sections should describe how you plan to make the project data Findable, Accessible, Interoperable and Reusable (FAIR). Each of the following five issues should be addressed with a level of detail appropriate to the project.***

|  |
| --- |
| **SUMMARY** *(dataset[[1]](#footnote-1) reference and name; origin and expected size of the data generated/collected; data types and formats)* |
| Four types of datasets have been identified for the ACRONYM project:   * In vitro imaging data for anatomical and functional characterization (~400GB): this includes confocal microscopy for anatomical tracing of inputs to cerebellar nuclei following viral injections, two-photon imaging of dendritic calcium signals and glutamate sensor responses. Raw data will be stored in proprietary microscope formats, as this is the equipment available in the host institute (.lif, .czi) and converted to TIFF stacks for analysis. Additional metadata will include injection coordinates, mouse line information, and immunostaining protocols. * Electrophysiological recordings of cerebellar nuclei neurons (~200GB): whole-cell patch clamp recordings in acute slices examining synaptic properties and plasticity of identified inputs. Data will be collected in WinWCP, an open-source software available on GitHub. Additional recordings will include optogenetic stimulation responses. Raw data can be loaded and analyzed in Python using the NEO library. * In vivo calcium imaging and behavioral recordings (~4TB): this combines 3D two-photon acousto-optic lens microscopy recordings through GRIN lenses for population imaging, along with calcium sensor responses during behavioral tasks. Video recordings (300Hz) will track limb movements and task performance. The data will be stored in TIFF stacks for imaging, with separate behavioral video files and DeepLabCut tracking data. * Analysis code and computational models (~1GB): custom analysis pipelines for processing all data types above, written primarily in Python and MATLAB. This includes spike inference algorithms, motion correction, behavioral analysis, and figure generation code. |

|  |
| --- |
| **1. MAKING DATA FINDABLE** *(dataset description: metadata, persistent and unique identifiers e.g., DOI)* |
| The various datasets generated within ACRONYM will be made findable using two complementary approaches: a thorough metadata description system and the use of persistent identifiers.  For the anatomical datasets, metadata will follow standardized Allen Brain Reference atlas terminology for the anatomical location of the studied structures. This is the standard frame of reference for most fine anatomical studies done in the past 10 years and is the most appropriate standard for this study. All stereotaxic coordinates used for viral injections will be provided to ensure reproducibility.  For the in vivo imaging datasets, metadata will include essential experimental parameters such as: mouse strain, sex, age, viral constructs used, imaging parameters (laser power, wavelength, frame rate), and behavioral task parameters. The standardized MINI (Minimum Information for Neuroscience Investigations) framework will be followed. For these datasets, metadata will be integrated within the Neurodata Without Borders (NWB) format, a standard format aiming at describing the data structure and context of various types of imaging experiments, which provides a way to associate heterogeneous types of recordings (for example: imaging data + video recordings + mouse running speed). Some work has already been done to convert my custom in-house format to the NWB standard.  For electrophysiological recordings, metadata will detail recording conditions (temperature, solutions), cell type identification methods, stimulation protocols, and analysis parameters. Additional metadata will specify the version of WinWCP used for acquisition and analysis code versions. These recordings will also be converted to the NWB format to facilitate their discovery and reuse.  Each experimental dataset will be assigned a Digital Object Identifier (DOI) through the chosen repository (Zenodo or any more appropriate repository). Raw data files will follow a systematic naming convention containing timestamp and experimental information: DATASET/YYYYMMDD/experiment\_X/HHmmss. A master spreadsheet will link these identifiers to experimental conditions and analysis files.  For all custom code and analysis pipelines, extensive documentation will be provided through example scripts, tutorials, and Jupyter and matlab notebooks demonstrating typical analysis workflows. These will be versioned and findable through GitHub, with major releases used for publication receiving their own DOIs through Zenodo integration. |

|  |
| --- |
| **2. MAKING DATA OPENLY ACCESSIBLE** *(which data will be made openly available and if some datasets remain closed, the reasons for not giving access; where the data and associated metadata, documentation and code are deposited (repository?); how the data can be accessed (are relevant software tools/methods provided?)* |
| Raw and processed data will be released using several open access repositories aligned with the data type. Data will be presented ahead of publication in internal and external seminars, international meetings, but the public data sharing will only occur at the time of the first scientific publication that uses it.  For anatomical datasets (viral tracing and confocal imaging), raw and processed data will be released using Zenodo or FigShare. Original images supporting the data will be provided in non-proprietary TIFF format. Any analysis code will be released on GitHub, with stable versions archived on Zenodo or FigShare to ensure long-term accessibility. These formats are widely used by the microscopy community and require no specific software beyond standard image analysis tools.  Electrophysiological recordings will be shared in two formats: the original WinWCP files, which can be opened using this free open-source software, and the NWB format, which ensures broader compatibility. Raw data will be converted and properly annotated. Code for data visualization and analysis will be shared via GitHub with stable versions used in publication archived on Zenodo or FigShare.  Population calcium imaging data will be of particular use to the computational modeling community, as there are currently no population recordings available in the cerebellar nuclei. This should generate interesting collaboration opportunities. The data will be provided in NWB format which enables sharing both the raw fluorescence signals and the extracted activity, along with behavioral variables. This standardized format will facilitate data reuse. However, the format of data generated during arboreal scan imaging with acousto optic lens microscope is more challenging to organize than more common widefield or patch-like imaging. Some work has been put and will be continued in converting our custom data format into files compatible with NWB, that can be reused by the larger community. Documentation will be provided. NWB files is accessible in both MATLAB and python, making the data accessible by most of the neuroscience community.  While all our code and data processing pipelines will be openly shared via GitHub, some raw datasets may be too large for standard repositories (particularly for the video recordings which can reach several hundred GB per animal). In these cases, we will share the data on request through our institutional server, while derived data products (extracted signals, behavioral variables) will be available through FigShare.  All associated publications will be deposited in the HAL open archive following CNRS requirements, with clear links to the associated datasets and code repositories. |

|  |
| --- |
| **3. MAKING DATA INTEROPERABLE** *(which standard or field-specific data and metadata vocabularies and methods will be used)* |
| Data interoperability will be achieved through the systematic use of standardized data formats, vocabularies, and cross-references across our different types of data.  All microscopy data will use the widely supported TIFF format, which is both non-proprietary and readable by all standard analysis software. All images will be annotated using Allen Brain Reference Atlas terminology for anatomical locations, ensuring consistency with the broader neuroscience community. Injection coordinates will be provided in standard stereotaxic coordinates.  For electrophysiological and calcium imaging recordings, we will convert our data into the NWB format. NWB is a standard format for neurophysiology that allows description of the data structures from multiple experimental modalities. While most calcium imaging data can be integrated into NWB, our arboreal scan recordings using acousto-optic lens microscopes require additional work due to their unique 3D random-access scanning approach. We are developing and will maintain specific tools to organize this data in a way that preserves its specific features while making it compatible with standard NWB specifications. All conversion tools will be documented and shared.  For behavioral data, we will use standardized formats for motion tracking (following DeepLabCut conventions) and task event description. Task events will be precisely timestamped and aligned with physiological recordings within the NWB format. Behavioral characterization will use standard metrics (step cycle timing, reach trajectory parameters) that are commonly used in the motor control field.  Analysis code in Python and MATLAB will be properly documented and versioned. All dependencies will be clearly specified, and we will provide environment files to ensure reproducibility. Functions will be compatible with commonly used analysis packages in the field (e.g., CaImAn for calcium imaging analysis). |

|  |
| --- |
| **4. INCREASE DATA RE-USE** *(what data will remain re-usable and for how long, is embargo foreseen; how the data is licensed; data quality assurance procedures)* |
| Population imaging data will be of particular value to the computational and modeling communities, as multiple groups are currently developing models of cerebellar functions. Having access to actual population activity in the cerebellar nuclei will help improve our understanding of cerebellar processing and could improve bio-inspired computational and robotic models. As this kind of recording has never been done in the cerebellar nuclei, it is likely to generate interesting collaborations.  Data quality will be assured at multiple levels. During experiments, quality control metrics will be automatically logged: image stability for anatomical data and population imaging, electrophysiological recording parameters (access resistance, cell healthiness), and behavioral performance metrics. For arboreal scanning data, custom tools have been developed to ensure proper signal extraction and quantification of recording stability in 3D. All these quality metrics will be provided alongside the data.  A significant challenge lies in organizing the arboreal scan data format compared to more standard imaging methods. We are developing and will maintain tools to convert our 3D random-access scanning data into formats that can be reused by the larger community. We will provide documentation of both the raw data format and conversion tools to facilitate reuse by other teams interested in applying this technique.  Data that will be used for publication will be released at the time of publication, which may exceed the end of the project. This will ensure proper documentation and validation of the datasets before public release. The initial release will include raw data alongside any derived data products used in publications.  All data will be made available under CC-BY 4.0 license, allowing reuse with appropriate citation. Analysis code and software tools will be released under the Apache 2.0 license to ensure compatibility with existing codebase and to encourage adaptation and reuse while maintaining appropriate attribution requirements. An example of reuse would be the validation of computational models of the cerebellum against our population recording data, or methodological development building upon our arboreal scanning approach. |

|  |
| --- |
| **5. ALLOCATION OF RESOURCES and DATA SECURITY** *(estimated costs for making the project data open access and potential value of long-term data preservation; procedures for data backup and recovery; transfer of sensitive data and secure storage in repositories for long term preservation and curation)* |
| All datasets are stored using a multisite backup strategy. At the acquisition level, data is stored locally on the workstation PC for fast data analysis, then copied to the institute's local server. Long-term storage and computational resources have been secured through the University of Strasbourg Computer Service Centre (Centre de Calcul de l'Université de Strasbourg, CCUS, https://hpc.pages.unistra.fr/). This infrastructure ensures both data preservation and the computational power needed for analysis of large imaging datasets.  The required hardware resources are covered by the project: workstations for data acquisition and analysis, personal computers, and local storage in the institute. There are no additional costs required for hosting code on GitHub and data on FigShare/Zenodo, as these platforms provide free use to academic accounts. The only constraint is the time required to format the data and ensure proper organization. However, as many journals now require data sharing, this process has been integrated into our general workflow. Open access publication costs have been budgeted in the project to ensure the visibility of the results and associated data.  The longevity of the data will be primarily limited by the lifespan of the platforms and of the code running it. GitHub is the major code-hosting platform and is owned by Microsoft, which should provide long-lasting financial support. Nonetheless, code hosted on GitHub uses the standard git protocol and could be moved to another platform (e.g. GitLab) or locally if required. Zenodo and FigShare are both supported by many academic players and can be easily replaced, although any DOI published pointing to these repositories would be hard to change.  Code is written in Python and MATLAB which have excellent reputations for backward compatibility. Code dependencies and versions will be clearly specified. For the more specific arboreal scanning analysis tools, proper documentation will ensure their usability beyond the lifetime of the current software versions.  Data collected in this project is not restricted by intellectual property constraints. Mouse lines will be used following institutional guidelines and proper licenses. |

**DISCLAIMER. Please note that the ERC Data Management Plan is not a part of the Ethics Review. It is the responsibility of the Principal Investigator to inform the ERCEA Ethics Team of any ethics issues/concerns regarding the collection, processing, sharing and storage of data in relation to the project.**

1. *Several datasets may be included into a single DMP.* [↑](#footnote-ref-1)