***\*Please remember to remove all guidance (in red) and any non-relevant sections that are not used on your project.***

**DATA MANAGEMENT PLAN**

**(2 pages max)**

***The DMP should clearly articulate how the investigators plan to manage and disseminate both the physical and digital data generated by the project, taking advantage of emerging information technologies and cyberinfrastructure. The plan must include sufficient detail for evaluation of its appropriateness and feasibility during merit review. DMPs often include existing practices in the principal investigator’s laboratory and the larger research community. –*** [***NSF ENG DMP Guidance***](https://www.nsf.gov/eng/general/dmp.jsp)

1. **Products of Research**

[What types of data (experimental, computational, or text-based), metadata, samples, physical collections, models, software, curriculum materials, and other materials will be collected and/or generated in the course of the project? The DMP should describe the expected types of data to be retained, managed, and shared, and the plans for doing so. What descriptions of the metadata are needed to make the actual data products useful and reproducible for the general researcher? For collaborative proposals, the DMP should describe the roles and responsibilities of all parties with respect to the management of data (including contingency plans for the departure of key personnel from the project) both during and after the grant cycle.]

The research data resulting from this project will be in raw data, in the form of image files (e.g., tiff), videos files (e.g., mpeg), raw data sets (text or binary), and theoretical model and data analysis methods. [Add specifics about your expected data]. Metadata will be compiled to ensure that information pertaining to data format, contents, conditions of data generation, and software compatibility is documented; examples of associated metadata are descriptions of experimental running conditions and information on electronic files such as author, file type, date file was created, and type of contents. All team members are responsible for managing and labeling their data in a systematic way for ease of use by the project team as needed.

All datasets and results of analyses will be made available in [Add specifics about chosen data repository, e.g., UC’s Dryad data repository or discipline-specific data repository] upon publication. If an appropriate data repository is not available, datasets and results will be published in supplementary materials or in a summarized format in peer-reviewed journal articles, theses, conference proceedings, book chapters and other print or electronic publishing formats.

As part of the project’s broader impacts plan, we expect resulting data to include curriculum materials and collected evaluation data. Evaluation data generated could include the following: (1) Students’ answers to survey questions, discussions and exit surveys; (2) Student metrics (e.g., gender, ethnicity, major, GPA) provided by the student or UCI’s institutional offices (e.g., departments, Office of Institutional Research, Graduate Division); and (3) Overall trend results from the analysis of these data. [Add specific information about outreach/broader impacts activities]. Only aggregate data will be reported and presented; no identifiable data will be released.

1. **Data Formats and Standards**

[In what format and/or media will the data or products be stored (e.g., hardcopy notebook and/or instrument outputs, ASCII, html, jpeg or other formats)? Where data are stored in unusual or not generally accessible formats, how may the data be converted to more accessible formats or otherwise made available to interested parties? When existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies. In general, solutions and remedies to providing data in an accessible format should be offered with minimal added cost.]

The data will be stored in various formats depending on how and where the data was collected. For example, [e.g., experimental data: hardcopy notebook, instrument outputs, excel data file]. Standardized filenames will be used for all electronic files and, as needed, files will be converted to more accessible formats (e.g., ASCII, text files) for file sharing with team members or other researchers.

All data will be backed up using an external hard drive, a server or an additional computer. User facility data as well as project-wide data will be stored on UCI’s recently-introduced Campus Research Storage Pool (CRSP), which is operated by the Research Cyberinfrastructure Center (RCIC). CRSP is a network-based multi-Petabyte storage cluster. All data written to CRSP is immediately replicated for two-copy data protection. Asynchronously, data is replicated to a third offsite facility for disaster recovery. CRSP is accessible through three low-level protocols (SFTP, HTTPS and NFS). The first two enable direct access to CRSP from in-lab equipment or computers.

1. **Dissemination, Access and Sharing of Data**

[What specific dissemination approaches will be used to make data available and accessible to others, including any pertinent metadata needed to interpret the data? In this case, "available and accessible" refers to data that can be found and obtained without a personal request to the PI, for example by download from a public repository. What plans, if any, are in place for providing access to data, including websites maintained by the research group and contributions to public databases/repositories? For software or code developed as part of the project, include a description of how users can access the code (e.g., licensing, open source) and specific details of the hosting, distribution and dissemination plans. If maintenance of a website or database is the direct responsibility of the research group, what is the period of time the website or database is expected to be maintained? What are the practices or policies regarding the release of data – for example, are they available before or after formal publication? What is the approximate duration of time that the data will be kept private? “Data sharing” refers to the release of data in response to a specific request from an interested party. What are the policies for data sharing, including, where applicable, provisions for protection of privacy, confidentiality, intellectual property, national security, or other rights or requirements? Research centers and major partnerships with industry or other user communities should also address how data are to be shared and managed with partners, center members, and other major stakeholders; publication delay policies (if applicable) should be clearly stated.]

Data will be shared, as appropriate, via publication in journal articles, conference proceedings and presentations, and by making the final data sets and metadata formally available to other researchers through the direct email contact with the individual faculty member, [X] website download, and data sharing platforms. Due to differences between team members with respect to research areas and privacy/IP requirements, data storage and sharing practices must be determined at the individual faculty member or institutional level. The PI will oversee and work with campus resources as necessary that all data has protection of privacy, confidentiality, intellectual property, national security, or other rights or requirements. The PI will also oversee and have team members work with UCI Libraries to follow the NSF Public Access policy in which the final accepted version of the peer-reviewed manuscript will be appropriately archived (i.e., NSF-PAR).

CRSP can be used for data sharing across the project via access with an individual UCI netID – outside-UCI team members can be assigned a temporary UCI netID. Data sharing via “read only” format is also available for outside researchers through a website interface.

Additionally, data can be deposited to Dryad, the open and curated data publishing platform for UC researchers to share and publish their data. California Digital Library and Dryad co-develop this service that is designed to be a simple data publication tool for researchers to meet funder and publisher mandates. Each dataset goes through a curation process to check for findability, accessibility, interoperability, and reusability. Other features of Dryad include: (1) Publish data through publisher and researcher workflow integrations; (2) Link data with related articles, code, and other datasets; (3) Permanent data citations with a DataCite DOI for attribution; (4) Preservation in Core Trust Seal certified Merritt repository; (5) Standardized views, downloads, and citation metrics in accordance with Make Data Count.

Software and open-source code developed in the course of a project can be made publicly available through a software repository with version control features [Add specifics about software repositories GitHub, CRAN, Zenodo, etc…].

1. **Re-Use, Re-Distribution and Production of Derivatives:**

[What are your policies regarding the use of data provided via general access or sharing? For data to be deemed “re-usable,” it must be accompanied by any metadata needed to reproduce the data, e.g., the means by which it was generated, detailed analytical and procedural information required to reproduce experimental results, and other pertinent metadata. Practices for appropriate protection of privacy, confidentiality, security, intellectual property, and other rights should be communicated. The rights and obligations of those who access, use, and share your data with others should also be clearly articulated. For example, if you plan to provide data and images on your website, will the website contain disclaimers or condition regarding the use of the data in other publications or products?]

The terms of re-use and re-distribution of data for general access or sharing will be determined at the individual faculty member level. The data acquired, preserved, and shared in the context of this project is governed by the University of California's policies pertaining to intellectual property, record retention, and data management as referenced by the California Digital Library. Any research data requiring the protection of the rights and welfare of human research subjects, such as evaluation of outreach activities, will be governed by campus IRB requirements. Prompt file of records of inventions and provisional patents will be done in partnership with UCI Beall Applied Innovation to protect intellectual property and reduce any delays/impediments to the dissemination of research data while retaining intellectual property rights of the University, investigators and students. Permission is required before reproducing any content (e.g. images, text, figures) that have been published in journals; individual journal requirements must be followed.

Datasets deposited and published in Dryad are made available under Creative Commons Zero (CC0) license, and data files must not contain sensitive or personally identifiable information. Software developed in the course of a project can be made publicly available for reuse under a software license, e.g., MIT open source, Apache version 2. Educational and outreach materials can also be made publicly at a project website or other repository under a license for reuse, such as CC-BY.

1. **Archiving of Data**

[When and how will data be archived and how will access be preserved over time? For example, will hardcopy logs, instrument outputs, and physical samples be stored in a location where there are safeguards against fire or water damage? Is there a plan to transfer digitized information to new storage media or devices as technological standards or practices change? Will there be an easily accessible index that documents where all archived data are stored and how they can be accessed? If the data will be archived by a third party, please refer to their preservation plans (if available). Where no data or sample repository exists for collected data or samples, metadata should be prepared and made publicly available over the Internet and the PI should employ alternative strategies for complying with the general philosophy of sharing research products and data as described above.]

RCIC has multiple Petabyte systems which are intended for long-term “scratch” storage. Four different parallel files systems provide about 3.2PB of usable capacity. While data is protected against multiple disk failures, all data in these file systems are single copy. Users can choose directories for “selective backup” to replicate data to another storage system. Replication runs daily.

Additionally, the project team has free access to the Dryad data repository, the data publication platform codeveloped by the California Digital Library and Dryad. Datasets deposited in Dryad are permanently archived in the backend Merritt Repository for both public access and long-term storage. Merritt is a Core Trust Seal certified repository, and datasets are retained indefinitely, files replicated in multiple geographic locations, files regularly audited for fixity and authenticity, and succession plans are in place in the case of repository closure.

The published data will be available in print from publishers or electronically in pdf format. All data and resources will be retained for a minimum of 3 years following the end of NSF funding or three years after public release, whichever is later.