

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

PAGE LIMIT: 2 pages

The DMP should address the five categories of information as specified in the PAPPG [categories below] and appropriately identify the data, metadata, samples, software, algorithms, curricula, documentation, publications, and other materials generated in the course of the proposed research. Moreover, the DMPs should describe how these materials will be disseminated, made accessible, and archived while incorporating the best practices and standards for the proposed research. DMPs should not be generic, so please edit the boilerplate text below.

You may request funds to cover costs of publication, page charges, or preparation of data as a direct cost in your budget proposal, which is evaluated as part of the merit review process. Any costs associated with implementing the DMP should be explained in the Budget Justification.

Some NSF Program Solicitations may contain specific and/or additional instructions that deviate from this guidance and/or provide exceptions to the two-page limit. Instructions in the solicitation take precedence over this guidance. - [NSF ENG DMP Guidance](#)

Data Management Plan

This Data Management Plan (DMP) addresses the NSF policy on the dissemination and sharing of research results. The project team will follow NSF's Public Access Policy, which requires scholarly scientific publications resulting from NSF research funding publicly accessible, machine-readable, and searchable at no charge to the reader. All resulting publications from this project will have the final accepted version of the peer-reviewed manuscript appropriately archived NSF PAR. The plan below encompasses data that is generated by the project team [\[info about the project team/other institutions\]](#).

1. Products of Research

[NSF Directions: 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 the project's experiments and simulations will be raw and analyzed 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, including the following if applicable:

- **Description of expected data specific to your project.**
- **Amount or size of the data**
- **Modality (text, imaging, genomic, structured...)**
- **Level of aggregation (individual, summarized...)**
- **Degree of data processing (raw, analyzed...)**

- Relationship of the data to other data, as relevant
- Use of publicly accessible archives]

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.

The PI will also outline use cases where metadata will be expected for the project. 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; sample identifiers; abbreviations or codes used in datasets; 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.

The PI will oversee and manage the project's data management and sharing policies. UC Libraries, including UCI Libraries, play an active role in supporting and advising researchers on data management, sharing, and preservation, with data sharing infrastructure and institutionally-supported full-time staff in data management.

[Add specifics about your project data, including the following if applicable:

- Roles and responsibilities of individuals in ensuring data management implementation is consistent with the DMP and potentially making updates to the DMP
- People or groups that have the right to manage the data
- People or groups that have intellectual property related to the data, data access, or data use]

2. Data Formats and Standards

[NSF Directions: 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]. Note any proprietary software or file formats needed to access data, and additional open (e.g., .csv) file formats that will be made available to enhance reusability of the data. All data will be backed up using an external hard drive, a server, or an additional computer. Additionally, project and experimental user facility data can be stored on UCI's Campus Research Storage Pool (CRSP), which is operated by the Research Cyberinfrastructure Center (RCIC). 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.

The project will have standardized documentation of electronic files such as author, file type, date file was created, and type of contents. [Metadata standards for your field can be found here by subject: <https://rdamsc.bath.ac.uk/subject-index> or longer list that can be filtered: <https://fairsharing.org/search?fairsharingRegistry=Standard&q=engineering>] Standardized filenames will be used for all electronic files and, as needed, files will be converted to more accessible formats (e.g., README, data dictionaries) for file sharing. UC librarians specializing in data services are available to assist with annotation of research data, formatting, and metadata workflows for archiving.

[Add specifics about your project data, including the following if applicable:

- Any standards to be applied to the scientific data, associated metadata, and documentation including models, formats, identifiers, definitions, unique identifiers, controlled vocabularies, taxonomies, thesauri, ontologies, code books, data dictionaries, and other data documentation.
- Whether the data standards are open or proprietary
- Indication that no appropriate data standards exist, as may be the case for some scientific fields]

3. Dissemination, Access and Sharing of Data

[NSF Directions: 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.]

The project team is committed to managing data in accordance with discipline specific standards, and those mandated by NSF and the University of California. In pursuit of these goals, project data will be made readily available to enable data discovery and distribution within a timely manner, e.g., at time of publication or within one year of publication. Data underlying peer-reviewed journal articles, conference proceedings, book chapters and other print or electronic publishing formats will be shared via established data sharing platforms. These data sharing platforms include NOMAD Repository and Archive¹ and Dryad².

For this project, the PI will use Dryad, the data sharing platform co-developed by the UC's California Digital Library. Datasets deposited and published in Dryad are made openly available under Creative Commons Zero (CC0) license. UC-members of the project team have free access to Dryad [Note: the deposit fee can be waived if a UC-affiliated researcher on the project is the data submitter (so other team

¹ <https://nomad-lab.eu/>

² <https://cdlib.org/services/uc3/dryad/>

members are then co-authors) Each dataset submitted to Dryad undergoes 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; and (4) Preservation in Core Trust Seal certified Merritt repository.

Code, software, or specialized tools needed to access, manipulate, or make use of the project data will be shared through GitHub, other open-source repositories, or available through direct contact with the research team. Software associated with a data package will also be uploaded via Dryad and published at Zenodo, which allows public software deposits with version control for the ongoing maintenance of software packages. 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. A description of any code, software or specialized tools needed as well as information or a link on how to access it will be documented on the data sharing platform, i.e., Dryad.

[Data sharing for “Research centers and major partnerships”:] For data sharing across the project team and partners during the project period, UCI has Google Workspace (unsecured) or the Campus Research Storage Pool (CRSP). CRSP is secured and 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.

[Add specifics about your project data, including the following if applicable:

- Identifying potential users of the shared data
- Any licenses or access rights for the data
- A description of whether data will be open to the public, limited, or closed and why (e.g. ethical or legal reasons or publication delay)
- A description of the shared subset of the total data generated, collected, or used (e.g., for big data, what can be shared reasonably such as metadata or partial dataset?)
- Any deadlines or embargo periods for releasing the data after they are generated
- Any other considerations that may result in limitations on the ability to broadly share scientific data (e.g., IP/NDA with industry)
- Names of the code, software, or specialized tools needed to access, manipulate, or make use of the data
- Elaborations of budget requests associated with data management]

4. Re-Use, Re-Distribution and Production of Derivatives:

[NSF Directions: 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. Datasets that do not contain sensitive data can be made available without restriction (e.g., CC0 license). 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.

[Add specifics about your project data, including the following if applicable:

- Measures to prevent the accidental or malicious modification of data
- How different tiers or levels of access will be managed
- Anything specific to your lab environment
- How cost, privacy, national security, competitiveness, or other considerations factored into DMP elements]

5. Archiving of Data

All data and resources will be retained for a minimum of three years following the end of NSF funding or three years after public release, whichever is later. Datasets deposited in Dryad for public access are permanently archived in the backend Merritt Repository for 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. After the project end date, the PI will work with UC Libraries to evaluate resources for long-term preservation (10 years+) of project data after direct project funding ends with the transfer of responsibilities for sharing and preservation.

[Optional] For data security, UC network and cloud storage access are protected by multi-factor authentication, and UC's system-wide IT infrastructure is continuously monitored for malware, known bad internet addresses, command-and-control traffic nodes, and indicators of compromise for additional security. Dryad implements and follows commercially reasonable electronic security measures to secure the systems through which information is collected or stored. Dryad stores a copy of all datasets in Zenodo for enhanced preservation services. Additionally, UC faculty and research staff are continuously updated on cybersecurity best practices through mandatory training exercises.