

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

Data Management Plan

This Data Management Plan (DMP) addresses the DOE policy that data sharing and preservation will enable validation of results. This plan encompasses data that is generated by the project team [info about the project team/other institutions].

Data Collected, Generated, or Used

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].

Standards

The project will have standardized documentation of electronic files such as author, file type, date file was created, and type of contents. Documentation will be maintained in the form of README files or data dictionaries. 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, and abbreviations or codes used in datasets.

[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>] 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]

Related Tools, Software and/or Code

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 contract with the research team. When possible, code will be stored as text files for use in the GitHub environment. 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. 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. Integration with Zenodo

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

- A description of any code or specialized tools needed to make use of the data, which may include:
- Names of the code, software, or specialized tools needed to access, manipulate, or make use of the data
- How these can be accessed (e.g., open source and freely available, generally available for a fee in the marketplace, available only from the research team or some other source)]

Data Sharing

The project team is committed to managing data in accordance with discipline specific standards, and those mandated by the DOE 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 the end of the project. Data underlying peer-reviewed journal articles, conference proceedings, book chapters and other print or electronic publishing formats, and presentations will be shared as either supplemental information, available to other researchers through the direct email contact with the individual team member, or available for download from established data sharing platforms. The latter includes NOMAD Repository and Archive¹, UC Dryad², and general research data repositories at the national laboratory or academic institution's library.

The most common method for disseminating the data types in this project is via publications in the peer-reviewed press and presentations at relevant meetings. To the extent practical and necessary for validation, the data produced in this project will be directly uploaded as supplementary information to published journal articles in a format that is open, machine-readable, and digitally accessible to the public. The published article will indicate how these data can be accessed. The project team will follow DOE's Public Access Policy, which requires scholarly scientific publications resulting from DOE 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 in DOE PAGES.

For this project, the PIs will use Dryad as the data sharing platform, where 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, the data publication platform co-developed by the California Digital Library and Dryad. [Note: the deposit fee can be waived if a UC-affiliated researcher on the project is the data submitter (so other team 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; (4) Preservation in Core Trust Seal certified Merritt repository; and (5) Standardized views, downloads, and citation metrics in accordance with Make Data Count.

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

- 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 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)
- Any deadlines or embargo periods for releasing the data after they are generated

¹ <https://nomad-lab.eu/>

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

- Any other considerations that may result in limitations on the ability to broadly share scientific data
- Identifying potential users of the shared data]

Data Preservation

All data and resources will be retained for a minimum of three years following the end of DOE funding or three years after public release, whichever is later. 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. 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.

Data Protection: Security and Integrity

During the project period, data will be stored on personal password-protected computers and backed-up on secured campus resources, such as UCI Google Workspace, Microsoft OneDrive, or UCI's Campus Research Storage Pool (CRSP). UCI Google Drive may only be used for Protection Level 1 and 2 data, and [Protection Level 3 and 4 data](#) should be stored on [Microsoft OneDrive](#). CRSP must not be used to store personally-identifiable information that would fall under guidelines such as [FERPA](#) (e.g. Student data), and [HIPAA](#) (health-care data)

CRSP is operated by the Research Cyberinfrastructure Center (RCIC) and is a network-based multi-Petabyte storage cluster with 7x24 support. 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.

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.

[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]

Oversight of Data Management

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

[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]

Rationale

The project team is committed to the efficient and effective sharing of its digital data with the scientific community, stakeholders and the public to facilitate validation of the research and to advance science for the broader research community.

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

- How the DMP influences the potential impact of the data within the immediate field and in other fields, and any broader societal impact
- How cost, privacy, national security, competitiveness, or other considerations factored into DMP elements
- Elaborations of budget requests associated with data management]