**DesignSafe Data Management Plan Guide**

This document is intended as a Data Management Plan (DMP) guide for researchers who will use the NHERI DesignSafe cyberinfrastructure (CI). Each research project will have specific details that will prevent a one-size-fits-all data management template from working for all users. Below is guidance on the five main areas required by the National Science Foundation (NSF) for DMPs, along with information about the DesignSafe functionalities that can support your data management needs. Also, you may use the DMP tool developed by the California Digital Libraries to prepare your two page DMP (<https://www.lib.utexas.edu/datamanagement/creating_plans)>. This tool functions as a template and includes all the elements that are required by NSF.

The five main areas that must be covered in an NSF data management plan are:

1. Types of data
2. Data and metadata standards
3. Policies for access and sharing
4. Policies for re-use and redistribution
5. Plans for archiving and preservation

Please contact the NHERI Experimental Facilities to populate the sections below with appropriate technical information as related to your proposed experiments.

1. **TYPES OF DATA**

In your data management plan, include a brief summary of the types, formats and prospective sizes of data that you will generate, as well as a summary of the existing data and models from other sources/repositories that you will reuse. Include information about any other materials needed to make these data useful to others, such as code, experimental plans and designs, white papers and help me files. In the context of a DMP, code is also considered data and can be licensed and publicly shared.

1. **DATA AND METADATA STANDARDS**

Your data need to be adequately described so it can be retrieved and reused by your team and by others. Some features of the DesignSafe CI can be described in your data management plan in order to demonstrate how you will meet this requirement. DesignSafe will have built-in support for descriptive tagging and automatic extraction of technical metadata about your files. In collaboration with the user community DesignSafe is developing a data model to map the structure of your research workflows, as well as metadata vocabularies, for the different hazards domains that will deposit data in the DesignSafe Data Repository, called the Data Depot. In this way your dataset will be described and published with all the information required for others to reuse it. It will also use the DataCite metadata schema for data citation purposes.

1. **POLICIES FOR ACCESS AND SHARING**

The DesignSafe CI is an end-to-end data management, analysis and publication platform. Your DMP should state that all data collected and generated will be deposited in the DesignSafe Data Depot from the inception of the research project. For those working with NHERI Experimental Facilities, experiment planning documentation and data incoming from the experiments will be shared with the researcher through the CI. While you are actively working on the research project you will be able to share your data with your team members and any other user to which you give authorization. The research team will be responsible for curating its data progressively using tools and facilities provided by the CI for the purpose of keeping track, describing, and organizing the data while conducting analysis using the DesignSafe Discovery Workspace. Data in the DesignSafe Data Depot will be version controlled, and it will be possible to gather and input provenance information throughout data analysis. If you see fit, you may describe your data workflow to best reflect the way in which you will manage your data.

The DesignSafe CI includes a public data repository that supports searching, browsing, sharing, downloading and reusing free and unrestricted (beyond citation requirements) data and supporting documents. After data is curated and ready to be published, it will receive a DOI for persistent identification, citation, and sharing.

1. **POLICIES FOR REUSE AND REDISTRIBUTION**

Within the DesignSafe CI, you will choose a license for your material. Because the DesignSafe Data Depot is an open repository, the following licenses will be offered:

* For datasets: [ODC-PDDL](http://opendatacommons.org/licenses/pddl/1-0/) and [ODC-BY](http://opendatacommons.org/licenses/by/)
* For copyrightable materials (for example, documents, workflows, designs, etc.): [CC0](https://creativecommons.org/publicdomain/zero/1.0/) and [CC-BY](https://creativecommons.org/licenses/by/4.0/)
* For code: any open, non commercial license (for example, [GPL](http://www.gnu.org/licenses/gpl.html))

You should select appropriate licenses for your data after researching which license best fit your needs and institutional standards. You should note in your DMP the licenses you plan to use. Note that datasets are not copyrightable materials.

1. **PLANS FOR ARCHIVING AND PRESERVATION**

Depositing your data and associated materials in the DesignSafe Data Depot will meet NSF requirements for preservation. Your data management plan should note that the CI will persistently maintain all uploaded data on storage resources at the Texas Advanced Computing Center, and these resources are redundant and geographically replicated. Features will be in place to ensure the authenticity, integrity, security and persistence of the datasets for open access. The DesignSafe CI is committed to the continuity of data preservation and will ensure preservation beyond the conclusion of the DesignSafe project.