

Plan Overview

A Data Management Plan created using DMPTool

DMP ID: <https://doi.org/10.48321/D1NQ2V>

Title: DMSP for Stanford Data Management Survey [Year 1]

Creator: John Borghi - **ORCID:** [0000-0001-9570-4163](https://orcid.org/0000-0001-9570-4163)

Affiliation: Stanford University (stanford.edu)

Funder: No Funding

Template: Stanford University Template

Project abstract:

The goal of this project is to gather information about current data-related practices at Stanford Medicine. This information will be used to guide the development of educational offerings and support services.

Start date: 12-01-2023

Last modified: 01-22-2024

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customize it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

DMSp for Stanford Data Management Survey [Year 1]

Data Types

The data for this project will consist of responses to a multiple choice questionnaire conducted initially on the Qualtrics platform. Participant responses will subsequently be exported into a widely used and/or open file format (e.g. .csv) for analysis.

Data will be collected as broadly as possible from the Stanford Medicine research community, We do not expect the total file size of the final dataset to exceed 50 MB.

Related Tools, Software, and/or Code

The data will be stored in open and/or common file formats (e.g. .csv) which will not require any specialized software to access or manipulate. Any custom code created to analyze data or create data visualizations will be written in a freely available programming language (e.g. R, Python) and will be made available in conjunction with the data.

Standards

To the best of our knowledge, there are not widely adopted formal standards for the organization and storage of the type of survey data we are collecting over the course of this project.

To ensure the (re)usability of the data, both for members of the research team and (potentially) for future researchers, we will maintain documentation outlining our process as well as the contents of data files (e.g. codebooks, data dictionaries) that will be accessible to the research team and will be shared alongside the data.

Data Preservation, Access, and Associated Timelines

Because of the sensitivity of survey responses, we may not share the data openly (publicly). Aggregated and/or deidentified data, documentation, and other materials (e.g. code, survey questions) may be made available through a generalist data repository.

At Stanford, data can be shared free of charge through a variety of generalist data repositories, including the Stanford Digital Repository and Dryad.

The Stanford Digital Repository (SDR) is a service built and run by Stanford Libraries. Datasets and other content in any format can be deposited into the SDR and made available immediately to the world via open access licenses and at no cost to the end user. Depositors can choose to have a Digital Object Identifier (DOI) assigned to their dataset, which is available at a persistent URL along with the required metadata describing the content. ORCID iDs may be included for every author. Datasets are replicated multiple times and stored in geo-diverse locations on different media types, providing long-term data management and data integrity. The stable technical infrastructure and dedicated staffing by digital preservation experts, funded by the University, ensures the long-term availability of the datasets.

Dryad is an open source tool for data publication and digital preservation. Datasets deposited into Dryad are permanently archived in a CoreTrustSeal-certified repository. Data files are regularly audited to ensure fixity and authenticity and are replicated with multiple copies in multiple geographic locations. Professional curators examine all Dryad deposits to ensure the validity of the data, apply robust metadata, and make certain that highly sensitive information has not been inadvertently included. Datasets deposited in Dryad are automatically assigned a Digital Object Identifier (DOI) and are indexed by Google Dataset Search and other tools to enhance discoverability.

Access, Distribution, or Reuse Considerations

The survey does not ask participants to provide personally identifiable information. If data is shared outside of Stanford, it will be shared in deidentified and/or aggregate form.

Oversight of Data Management and Sharing

Developing, executing, and monitoring this Data Management and Sharing Plan will be the responsibility of the research team (John Borghi and Mario Malicki), both of whom are experienced researchers with a background studying data management and sharing. Practices and procedures related to data management and data sharing will be outlined and reviewed with all project personnel and will also be discussed as part of regular team meetings. The team members will also be responsible for ensuring this plan remains up-to-date.
