Plan Overview

A Data Management Plan created using DMP Tool

Title: NIH for Marquette Template

Creator: Danelle Orange

Affiliation: Marquette University (marquette.edu)

Funder: National Institutes of Health (nih.gov)

Template: NIH-GEN DMSP (Forthcoming 2023)

Project abstract:

If you are applying for a grant with NIH, starting in 2023, this is the data management plan, including Marquette specific guidance that you should use. If you are applying for a grant with the NIH or associated entities in 2022, you are not required to use this format, although we do recommend using this form.

Last modified: 07-01-2024

Copyright information:

NIH for Marquette Template

Data Type

A general summary of the types and estimated amount of scientific data to be generated and/or used in the research. Describe data in general terms that address the type and amount/size of scientific data expected to be collected and used in the project (e.g., 256-channel EEG data and fMRI images from ~50 research participants). Descriptions may indicate the data modality (e.g., imaging, genomic, mobile, survey), level of aggregation (e.g., individual, aggregated, summarized), and/or the degree of data processing that has occurred (i.e., how raw or processed the data will be)

This project will produce _____ [Data type, e.g., imaging, sequencing, experimental measurements] data generated/obtained from _____ [e.g., instrument, method, survey, experiment, data repository]. Data will be collected from ____ [number] of research participants/specimens/experiments, generating ____ [number] datasets totaling approximately ____ [amount of data, in the following format: TB.GB/MB] in size. The following data files will be used or produced in the course of the project: _____ [list input data files, intermediate files, and final, post-processed files]. Raw data will be transformed by _____ [analysis, method] and the subsequent processed dataset used for statistical analysis. To protect research participant identities, _____ [e.g., individual, aggregated, summarized] data will be made available for sharing.

A description of which scientific data from the project will be preserved and shared.

Based on _____ [ethical, legal, technical] considerations, the following data produced in the course of the project will be preserved and shared: _____ [list] **OR** All data produced in the course of the project will be preserved and shared.

A brief listing of the metadata, other relevant data, and any associated documentation (e.g., study protocols and data collection instruments) that will be made accessible to facilitate interpretation of the scientific data.

To facilitate interpretation of the data, _____ [e.g., metadata, documentation, protocols, data collection instruments] will be shared and associated with the relevant datasets via readme.txt files for the associated data.

Related Tools, Software and/or Code

An indication of whether specialized tools are needed to access or manipulate shared scientific data to support replication or reuse, and name(s) of the needed tool(s) and software.

If no specialized tools are needed to access or manipulate the data:

_____ [Data type - Imaging data, survey data, etc] data will be made available in _____ [csv, txt, dicom, non-proprietary etc] format and will not require the use of specialized tools to be accessed or manipulated.

If specialized tools are needed to access or manipulate the data:

_____ [Data type] data will be made available in _____ format, which requires the use of specialized tools, such as _____ [include list of tools] to be accessed and manipulated.

If applicable, specify how needed tools can be accessed, (e.g., open source and freely available, generally available for a fee in the marketplace, available only from the research team) and, if known, whether such tools are likely to remain available for as long as the scientific data remain available.

- The ____ tool, which can be used to ____ is available free of charge through ____ [source name]
- The ____ tool, which can be used to ____ is available for a fee of ____ through ____ [source name].
- Custom tools to ____ will be/have been developed by the research team.
    - Requests for these tools should be directed to ____ [include details of members of the research team].
    - These tools will be shared openly via epublications.marquette.edu.

Standards

An indication of what standards will be applied to the scientific data and associated metadata (i.e., data formats, data dictionaries, data identifiers, definitions, unique identifiers, and other data documentation).

To facilitate their efficient use, all of our data and materials will be structured and described using the following standards:

If there are formal data standards for some/all of the data:

Whenever possible, we will use _____ [common data elements, standardized survey instruments, etc] to structure and organize our data.

Our ____ data will be structured and described using the ____ standard, which has been widely adopted in the ____ community. Our notebooks will be retained using Microsoft Teams and {labnotes.} [Add additional information about this standard, if applicable - e.g. implementation in data repositories, utility in combining/reusing datasets]

If there are not formal standards:

Formal standards for ____ data have not yet been widely adopted. However, our data and other materials will be structured and described according to best practices.

Data will be stored in common and open formats, such as ____ for our ____ data. Information needed to make use of this data [e.g. the meaning of variable names, codes, information about missing data, other metadata etc] will be recorded in ____ [data dictionaries/codebooks] that will be accessible to the research team and will subsequently be shared alongside final datasets.

Information about our research process, including the details of our analysis pipeline will be maintained contemporaneously, using Microsoft Teams, {and lab notebooks} , an____ [lab notebooks, protocols, etc]. This information will be accessible to all members of the research team and will be shared alongside our data.

Data Preservation, Access, and Associated Timelines

The name of the repository(ies) where scientific data and metadata arising from the project will be archived.

All dataset(s) that can be shared will be deposited in epublications.marquette.edu. Epublications.marquette.edu is the institutional repository for Marquette University and it's primary mode of sharing research.   Epublications.Marquette is backed up to Amazon Glacier, an S3 cloud storage system and follows OAI-PMH standards for replication and transfer. The files are tested annually for fixity and authenticity and succession plans exist in the event of repository closure.

[OR]

All dataset(s) that can be shared will be deposited in epublications.marquette.edu (institutional bepress repository) as a link, [Add appropriate NIH-supported data repositories], AND/OR _____ [Add appropriate subject or disease repositories]

How the scientific data will be findable and identifiable, i.e., via a persistent unique identifier or other standard indexing tools.

The epublications.marquette.edu provides metadata, persistent identifiers (handles), and long-term access. This repository is supported by Marquette University and dataset(s) are available under a Creative Commons attribution license **OR** through a request process using Microsoft teams forms, and Selected Works.

When the scientific data will be made available to other users (i.e., the larger research community, institutions, and/or the broader public) and for how long.

Data will be made available as soon as possible or at the time of associated publication.

Access, Distribution, or Reuse Considerations

Describe any applicable factors affecting subsequent access, distribution, or reuse of scientific data related to:

- **Informed consent (e.g., disease-specific limitations, particular communities' concerns).**
- **Privacy and confidentiality protections (i.e., de-identification, Certificates of Confidentiality, and other protective measures) consistent with applicable federal, Tribal, state, and local laws, regulations, and policies.**

[Marquette University classifies data into 4 different categories. Researchers should chose the correct one for their activities and detail their data type and actions related to access, distribution, and reuse.

- **Category 1**: Publicly Available Information: This includes publications, unrestricted databases, and other materials available on-line that are not identified as confidential by the provider of the information.

  Publicly available data is made available via the Epublications.marquette.edu repository as well as any subject repositories. Please detail those repositories and timeline to getting the items into the selected repositories.

- **Category 2**: Compiled or Generated Restricted: This data is generated during research or compiled as part of the research. This may include data sets created through measurements of materials, insects, or animals, compiled from publicly available information, lab notes, images, survey results.

  This data can be made available via the digital repository. Please detail what security procedures you will be taking to ensure both access and security if necessary. The Epublications repository can allow you to control access to specific individuals or based on location, please contact digitalprograms@marquette.edu for specific details)

- **Category 3**: Compiled or Generated Controlled: This is data that is not only confidential, but for which access is also controlled by federal or state law, such as personally identifiable information, personal health information, student records, or controlled unclassified information.

  This information must be deidentified before **any** publication. Please include your plan to anonymize this data prior to publication and security procedures taken to prevent unauthorized use. Please contact digitalprograms@marquette.edu for specific language regarding restricting access on Epublications should you choose to add the data set to the repository.

- **Category 4**: Licensed Data: This is data that has been licensed from a third party for use in conducting a research project.

  Third Party data is not something the University can publicly post. Please detail the ways that others can get access to the same data via that 3rd party.]

Whether access to scientific data derived from humans will be controlled (i.e., made available by a data repository only after approval).

- **Any restrictions imposed by federal, Tribal, or state laws, regulations, or policies, or existing or anticipated agreements (e.g., with third party funders, with partners, with Health Insurance Portability and Accountability Act (HIPAA) covered entities that provide Protected Health Information under a data use agreement, through licensing limitations attached to materials needed to conduct the research).**
- **Any other considerations that may limit the extent of data sharing.**

In order to ensure participant consent for data sharing, IRB paperwork and informed consent documents will include language describing plans for data management and sharing data, describing

the motivation for sharing, and explaining that personal identifying information will be removed.

To protect participant privacy and confidentiality, shared data will be de-identified using the _____ method. [Describe de-identification method, noting any other applicable laws or policies such as HIPAA].

Oversight of Data Management and Sharing

Indicate how compliance with the Plan will be monitored and managed, frequency of oversight, and by whom (e.g., titles, roles).

The following individuals [or just the position titles if unknown] will be responsible for data collection, management, storage, retention, and dissemination of project data, including updating and revising the Data Management and Sharing Plan when necessary.

* Name, Position Title, Host Institution, ORCID, email

Sample Language for budgeting requirements

This project includes the following costs associated with data management and sharing.

For data curation and the development of related documentation, the project is requesting $_____. These funds will allow us to prepare data for sharing including de-identification of data, the incorporation of metadata to ensure discoverability and the data transfer process to _____repository for preservation and access. **An additional cost of $ _____ is required to cover data deposit fees for _____ repository, which will cover _____ years of hosting.**