Generic Template for De-Identified Data Shared on Figshare

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.

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. [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, { lab notebooks, Electronic Lab Notebook, protocols.io, 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 [UR Figshare link]. [U R F i g s h a r e l i n k] is the institutional repository for University of Rochester and it's primary mode of sharing research. Automatic backups of all figshare.com data files and metadata are performed nightly, directly to Figshare-managed Amazon Web Service (AWS) S3 buckets.

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

The [UR Figshare link] provides metadata, persistent identifiers (DOI), and long-term [10 year] access. This repository is supported by the University Rochester and dataset(s) are available under a Creative Commons attribution license

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.**

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.

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 documentation 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. Only the minimal necessary PHI will be collected for the purposes of the study, and all team members are HIPAA trained. To protect participant and family member privacy and confidentiality, shared data will be de-identified according to all federal and state guidelines and in accordance with HIPAA policy rule following the safe-harbor method.

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.**