Sample plan created by the Working Group on NIH DMSP Guidance

Converting a resource sharing plan into a DMS Plan

OSF Link: https://doi.org/10.17605/OSF.IO/UADXR

RESOURCE SHARING PLAN

This is an example of a resource sharing plan that written by a PI prior to the 2023 NIH DMSP.

Data Sharing Plan. The proposed research will include data from banked biological material from up to 354 individuals living within a relatively restricted geographic area. The final dataset will include self-reported demographic data (age, sex) and the results of study-related tests for loiasis and other filarial infections. Even though the final dataset will be stripped of identifiable information, there remains the possibility that members of the local community might deduce the identity of participants from the de-identified database. Therefore, we will make all de-identified data from this study available only to interested parties with approval from the PI after signing a Data Sharing Agreement. The Data Sharing Agreement will include instructions to: (1) use the data only for research purposes; (2) not identify any individual participant; (3) keep the data secured at all times; and (4) destroy or return the data after analyses have been completed. After publication of the results of the study, participant level data that have been stripped of demographic information will be published as supplementary data and/or made publicly available in a data repository maintained by the Library at Institute. This study does not include development of model organisms or large-scale genomic studies.

DATA MANAGEMENT AND SHARING PLAN

This is an example of the above resource sharing plan rewritten to fit the sections outlined by the 2023 NIH DMSP template. It includes tips within colored boxes to highlight the required elements within each section.

Project Name: The Feasibility of the XYZ Method in Patients
PI Name and Affiliation: John Doe, University of Universities
Date Finalized: 2023-01-05

DOI: XX.XXXX/XXXXXX

Section 1: Data types
Describe all the data types and approximate amounts of each type to be collected during the study
Indicate which data will be preserved and shared and address security measures (if applicable)
Include descriptions of metadata and if/where study protocols will be accessible

1. Data types. The proposed research will include three data types: clinical data, proteomics data and metabolomics data. The final clinical dataset will include self-reported demographic data and the results of study-related tests for loiasis and other filarial infections. The proteomics data will include raw LC MS/MS reads from biological samples prepared from each of the 354 research subjects. The metabolomics data will include raw LC-MS and GC-MS files from subject samples. Tabular data of hits will also be saved in .csv formats. The Omics data will be made publicly available stripped of demographic information, and the de-identified clinical data will be made available by request. Protocols and details regarding instrument settings, data transformation and analysis will be made available in the accompanying plain text README document.

Section 2: Tools, software and/or code
Include the tools and computer software that will be used
Indicate if proprietary file formats will also be saved by non-proprietary means or if special software will be necessary for other users to access and reuse the data
Get help from the technicians who are running the data collection and processing

2. Tools, software and/or code. Data collection will be done in REDCap and data analysis will be performed using Microsoft Excel and R. An institutional instance of Box will be used for active data storage, which is HIPAA-compliant. For the proteomics data, resulting spectra will be converted to Mascot generic format (MGF) files using Proteome Discoverer v2.1.0.81. For the metabolomics data, raw structural information about GC-MS features will be obtained through spectral matching with the NIST 14 spectral library. Mass Profiler Professional software (Agilent) for GC-MS data and Analyst (AB Sciex) for LC-MS/MS will be used to assign peaks to raw ion chromatograms. No specialized tools or software will be needed to access or reuse the shared datasets, which will be available via a public repository (de-identified omics data), or via email after the completion of a DSA as a .csv file (de-identified clinical data).

Section 3: Data standards
Look up the data standards for each type of data output
Consult the NIH maintained list of common data elements
If there is no standard, note this information and build your own standard with a data dictionary

3. Data standards. A data dictionary will be provided for the clinical dataset that defines column headers, units of measurement and other pertinent metadata as necessary to understand and reuse the dataset. For the proteomics datasets, we will follow the Minimum information about a proteomics experiment (MIAPE) standard which was created by the Proteomics Standards Initiative of the Human Proteome Organization. For the metabolomics datasets, we will follow the guidelines dictated by the Metabolomics Standards Initiative (MSI). Both of the repositories selected for the omics data streams comply with the accepted standards for their field.

Section 4: Data preservation, access, and timelines
Provide repository name(s) and indicate how data will be findable
Make sure to discuss access and distribution for all of the types of data generated in the study
Indicate what level of data will be shared (raw, aggregate, de-identified etc.) and when and how long it will be shared (not all data needs to be shared at the same level)
Avoid hyperlinks in the DMSP, save these for the RPPR submission

4. Data preservation, access, and timelines. Concomitant with publication of the results of the study, participant level data that have been stripped of demographic information will be published as supplementary data and/or made publicly available (with restricted access as laid out in section 5 below) in the PI's institutional data repository, which will mint a DOI and continuing providing access for at least 10 years, or as long as the repository exists. The data repository also commits to maintaining at least one copy in the cloud in case of data loss. Raw proteomics data and accompanying metadata will be made publicly available through the ProteomeXchange data repository for at least 10 years. Raw metabolomics data and accompanying metadata will similarly be made publicly available (10+ years) from the Metabolomics Workbench data repository which is accessible by metabolite data aggregator, MetabolomeXchange.

Omics data will be assigned a persistent unique identifier during repository deposition and this PID will be included in relevant publications. Scripts and coding workflows will be made available via GitHub.

Section 5: Access, distribution or reuse considerations
Discuss limitations of the data sharing
Address security concerns here, including how access will be controlled

5. Access, distribution or reuse considerations. The authors have attempted to maximize FAIR sharing of the data to be collected from this project. The omics data will be made publicly available to users through their respective specialist repositories. The clinical data will be stripped of most personal identifiers but will contain some indirect identifiers (such as age, sex, and location) in order to maintain the scientific value of the results. To protect subject privacy, access to the clinical data will be controlled by the PI and released to users following the completion of a Data Sharing Agreement (DSA). The DSA will include instructions to: (1) use the data only for research purposes; (2) not identify any individual participant; (3) keep the data secured at all times; and (4) destroy or return the data after analyses have been completed.

Section 6: Oversight
Oversight will usually be the responsibility of the PI
Oversight includes revising the DMSP and adhering to submission deadlines for sharing data
All members of the team should have training on the DMSP

6. Oversight. The PI of the proposal will make the plan available to all personnel involved in the project. The PI will be responsible for ensuring faithful adherence to the DMS Plan and revising the plan annually, as the research project evolves.