

Plan Overview

A Data Management Plan created using DMPTool

DMP ID: <https://doi.org/10.48321/D16919>

Title: The study and improvement of data collection and management for infectious diseases and immunodeficiencies

Creator: Emily Ricotta - **ORCID:** [0000-0001-9928-8275](https://orcid.org/0000-0001-9928-8275)

Affiliation: National Institutes of Health (nih.gov)

Principal Investigator: Emily Ricotta

Project Administrator: Emily Ricotta

Funder: National Institutes of Health (nih.gov)

Funding opportunity number: AI001367

Template: NIH-Default DMSP

Project abstract:

Nonrandomized, observational population studies play a critical role in evaluating health outcomes after exposures of interest such as an infectious disease, an immune disorder, or a particular medical intervention. The strength of these studies is that, when well-designed and executed, they can be more representative of the exposure mechanism and population under study than a randomized control trial (RCT), which is considered the gold standard of medical research. However, while efforts have been made to standardize RCT protocols during infectious disease outbreaks, and for disease treatment in general, protocol standardization of nonrandomized studies is currently lacking.

This research aims to 1) identify gaps in study design and data collection methods during

emerging infectious disease outbreaks and 2) develop standardized tools to fill these gaps to enhance the efficiency and reproducibility of infectious disease research, particularly during emerging infectious disease outbreaks.

Start date: 10-01-2022

End date: 09-30-2025

Last modified: 02-07-2023

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

The study and improvement of data collection and management for infectious diseases and immunodeficiencies

Data Type

Types and amount of scientific data expected to be generated in the project: *Summarize the types and estimated amount of scientific data expected to be generated in the project.*

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 primarily secondary data and data collection tools. A scoping review will generate a dataset comprised of articles published early during outbreak/epidemic scenarios as well as summary metrics of the data collected within those articles. Informant interviews with the corresponding authors of the manuscripts included in the review may be conducted. Tools will include protocol development guides, data collection and management aides, and training materials.

Literature review data will be kept in excel spreadsheets or will be stored in software such as Covidence. Interviews will be conducted as surveys and will be stored in REDCap. Tools will be developed in Word and will be stored in Box.

Scientific data that will be preserved and shared, and the rationale for doing so: *Describe which scientific data from the project will be preserved and shared and provide the rationale for this decision.*

Data produced during the literature review will be shared with the manuscript at the time of publication. Tools will be shared openly with the NIH community and on Github. Raw data will be retained in secure, NIH-approved storage platforms including REDCap (author surveys) and NIH Box.

Metadata, other relevant data, and associated documentation: Briefly list 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 the interpretation and reuse of the data, a README file and data dictionary will be

generated and deposited into a repository along with all shared datasets. The README file will include method description and other tools (e.g., software, databases). The data dictionary will define and describe all variables in the dataset.

Related Tools, Software and/or Code

State whether specialized tools, software, and/or code are needed to access or manipulate shared scientific data, and if so, provide the name(s) of the needed tool(s) and software and specify how they can be accessed.

Sharable data will be made available in csv format and will not require the use of specialized tools to be accessed or manipulated.

Standards

State what common data standards will be applied to the scientific data and associated metadata to enable interoperability of datasets and resources, and provide the name(s) of the data standards that will be applied and describe how these data standards will be applied to the scientific data generated by the research proposed in this project. If applicable, indicate that no consensus standards exist

Literature reviews will follow PRISMA guidelines for scoping reviews, including synthesis of evidence: <https://www.prisma-statement.org/Extensions/ScopingReviews?AspxAutoDetectCookieSupport=1>

Data Preservation, Access, and Associated Timelines

Repository where scientific data and metadata will be archived: Provide the name of the repository(ies) where scientific data and metadata arising from the project will be archived; see [Selecting a Data Repository](#))

Data produced during the literature review and associated metadata will be shared with the manuscript at the time of publication. If direct data sharing via the journal where the review is published is not possible, data will be deposited in Zenodo.

Tools will be shared openly with the NIH community and on Github.

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

Data sharing via the journal where the review is published will result in a digital object identifier associated with the manuscript itself. If data is published in Zenodo, the repository provides metadata and assigns a digital object identifier to all data shared in the repository and enables long-term data access. Zenodo is funded by the European Commission through CERN and OpenAIRE. Users must specify a license for all publicly available files. Licenses for closed access files may be specified in the description field. All meta data is openly available under CC0 license. By uploading content, no change of ownership is implied and no property rights are transferred to CERN. All uploaded content remains the property of the parties prior to submission.

<https://about.zenodo.org/policies/>

When and how long the scientific data will be made available: Describe when the scientific data will be made available to other users (i.e., no later than time of an associated publication or end of the performance period, whichever comes first) and for how long data will be available.

Scientific data will be made available no later than the time of publication.

Access, Distribution, or Reuse Considerations

Factors affecting subsequent access, distribution, or reuse of scientific data: NIH expects that in drafting Plans, researchers maximize the appropriate sharing of scientific data. Describe and justify any applicable factors or data use limitations affecting subsequent access, distribution, or reuse of scientific data related to informed consent, privacy and confidentiality protections, and any other considerations that may limit the extent of data sharing. See [Frequently Asked Questions](#) for examples of justifiable reasons for limiting sharing of data.

All finalized data will be shared in an uncontrolled fashion once deposited in the appropriate repository.

Whether access to scientific data will be controlled: State whether access to the scientific data will be controlled (i.e., made available by a data repository only after approval).

Access to data will not be controlled and will be available to interested users at the time of publication in an unrestricted fashion.

Protections for privacy, rights, and confidentiality of human research participants:
If generating scientific data derived from humans, describe how the privacy, rights, and confidentiality of human research participants will be protected (e.g., through de-identification, Certificates of Confidentiality, and other protective measures).

In order to ensure consent for data sharing (corresponding author surveys), 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.

All records will be kept confidential to the extent provided by federal, state, and local law. Authorized representatives of NIAID may inspect all documents and records required to be maintained by the investigator, including but not limited to, medical records. Records will be kept locked and data will be coded. Any personally identifiable information maintained for this study will be kept on restricted-access computers and networks. Personally identifiable information will only be shared with individuals authorized to receive it under this protocol. Individuals not authorized to receive personally identifiable information will be provided with coded information only, as needed. Clinical information will not be released without written permission of the participant, except as necessary for monitoring by the IRB, NIAID, and the Office for Human Research Protections (OHRP).

Oversight of Data Management and Sharing

Describe how compliance with this Plan will be monitored and managed, frequency of oversight, and by whom at your institution (e.g., titles, roles).

The following individuals 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:

- Emily Ricotta, Principal Investigator, Division of Intramural Research, NIAID, emily.ricotta@nih.gov
-