# National Institutes of Health: Data Management and Sharing Plan (2023) – General template

## ELEMENT 1: Data Type

1. **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.*

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 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.

1. **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.*

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.

1. **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 interpretation of the data, \_\_\_\_\_\_ [e.g., metadata, documentation, protocols, data collection instruments] will be shared and associated with the relevant datasets.

## ELEMENT 2: 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.*

***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, 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.

* 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 \_\_\_\_.

***Sample language for sharing code***

Relevant resources, such as code used for data processing and analyses, will be made publicly available through GitHub (<https://github.com>), a code repository service also used by the NIH. GitHub is a web-based platform that host source codes, documentation, and project-related web content for research projects. Code documentation will include instructions on how to access data, the name of a contact person for questions, and all relevant references to publications. To ensure long-term accessibility, a copy of the GitHub code repository will be archived in Zenodo (<https://zenodo.org/>) at the time of publication. Zenodo is an open access repository that specializes in preserving software and issues DOIs for code. The code DOI will be included in each resulting publication.

## ELEMENT 3: 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.*

To facilitate their efficient use, all 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 \_\_\_\_ [CSV, TXT, DICOM] 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 \_\_\_\_ [lab notebooks, protocols, etc]. This information will be accessible to all members of the research team and will be shared alongside our data.

## ELEMENT 4: Data Preservation, Access, and Associated Timelines

1. **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).*

All dataset(s) that can be shared will be deposited in \_\_\_\_\_\_\_\_\_ [Add appropriate NIH-supported data repositories] OR \_\_\_\_\_\_\_\_ [Add appropriate subject or disease repositories]

1. **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*

The \_\_\_\_\_\_\_\_\_ [Insert repository name] provides metadata, persistent identifiers (i.e., insert whether DOI, handles, other), and long-term access. This repository is supported by \_\_\_\_\_\_\_\_[Insert funder/organization] and dataset(s) are available under a \_\_\_\_\_\_\_ [Insert license information] **OR** through a request process \_\_\_\_\_\_\_\_\_\_ [Insert information about request process].

***Sample Language for Dryad Data Repository***

Dataset(s) resulting from this research will be shared via the generalist repository Dryad, which provides metadata, persistent identifiers (i.e., DOIs), and long-term access. Dryad is the institutional data repository supported by the University of California and all data is shared under a CC0 waiver, which makes the dataset(s) publicly available. Data will be made available as soon as possible or at the time of associated publication. Dryad datasets are backed up to Merritt, the UC’s CoreTrustSeal-certified digital repository, for long-term storage and accessibility. Procedures in place to ensure dataset preservation include storage of data files in multiple geographic locations, regular audits for fixity and authenticity, and succession plans in the event of repository closure.

1. **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.*

Data will be made available as soon as possible or at the time of associated publication. The duration of preservation and sharing of the data will be a minimum of \_\_\_\_\_[*duration*] years after the end of the funding period.

## ELEMENT 5: Access, Distribution, or Reuse Considerations

1. **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.*

Data will be de-identified before it is shared. The IRB application and informed consent documents will include language describing the data management and sharing plan, explaining the motivation for sharing, and ensuring that personal identifying information will be removed prior to sharing.

**B. 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).*

Given the sensitive nature of the dataset, de-identified human subjects data will be made available in \_\_\_\_\_\_\_\_ data repository, which restricts access to the data to qualified investigators with an appropriate research question who sign a data use agreement. [Describe data repository access methods and security measures].

**C. 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).*

To protect participant privacy and confidentiality, shared data will be de-identified using the \_\_\_\_\_\_ method. [Describe de-identification method (HIPAA Safe Harbor, Expert Determination), noting any other applicable laws or policies such as HIPAA].

## ELEMENT 6: Oversight of Data Management and SharinG

The Office of Sponsored Programs at Florida Atlantic University that will be administering this award has created a data management and sharing plan compliance system as part of their process for submitting the annual NIH progress report. The Office of Sponsored Programs will check information related to the number of research participants that are deposited into GEO and dbGaP and look for the accession numbers when publications occur prior to the end of the grant award.