Writing a Data Management and Sharing Plan for NIH

Ariel Deardorff, UCSF
Wasila Dahdul, UCI

2/13/2023
About Us

Wasila Dahdul (she/her)
Data Curation Librarian
UC Irvine
wdahdul@uci.edu

Ariel Deardorff (she/her)
Data Services Librarian
UC San Francisco
ariel.deardorff@ucsf.edu
About You

---

Take 1 minute to introduce yourself in the chat. Let us know:

- Your name
- Your campus
- Your role
- What kind of data you work with
Our Agenda

1. Overview of the new NIH DMS Policy (10 min)
2. Building your plan (40 min)
3. Questions (10 min)
What is this new policy anyway?
Final NIH Policy for Data Management and Sharing

Notice Number:
NOT-OD-21-013

Key Dates

Release Date:       October 29, 2020
Effective Date:    January 25, 2023

NIH's Scientific Data Sharing Website: https://sharing.nih.gov
Effective January 25, 2023

- Requires researchers seeking NIH funding to prospectively submit a plan outlining how scientific data from their research will be managed and shared.
- Researchers should “maximize the appropriate sharing of scientific data.”
- Data should be shared as soon as possible, and no later than the time of an associated publication or end of performance period (whichever comes first).
- This plan represents the minimum requirements. NIH ICOs may expect more specificity in their plans - check funding announcements for information.
What kind of research does the new policy apply to?

- Applies to all research funded in whole or in part by NIH that generates scientific data:
  - Extramural grants
  - Contracts
  - Intramural research projects
  - All other NIH funding mechanisms

- Exception: funding that does not generate data (e.g., training (T) or fellowship (F) grants).
Not all data needs to be shared

- informed consent limitations
- existing agreements prohibit sharing
- privacy or safety of research participants need protection
- explicit law or regulation prohibiting sharing
- data cannot be digitized with reasonable effort
What do you write in your DMS plan?

You will need to describe the following 6 elements:

- Data type(s) and metadata (data description)
- Related tools, software, and/or code
- Standards for the data/metadata
- Data preservation, access, and associated timelines
- Access, distribution, or reuse considerations
- Oversight of data management and sharing
TL;DR: What’s new here?

- 2-page **data management and sharing plan** that is much more detailed than previous data sharing requirements
- Baseline has shifted towards data sharing as the **default**
Building your DMS Plan
Don’t reinvent the wheel!

There are several templates and sample plans already available. You can find templates at:

- NIH’s DMS Plan Format Page (Word file)
- DMPTool - step by step plan builder tool

Sample plans

- NIH website (including genomics, clinical, survey, secondary data...)
- Your campus library or grants office
Today’s Activity: Building your DMS Plan

---

We are going to spend the rest of our time today collaboratively developing your DMS plans.

To start, take 3 minutes to locate and download a template or example plan that you can modify:

- Sample NIH Plans
- Blank template

We are going to walk through each section one by one and give you a short amount of time to jot down an outline of what you will write.
Reminder of our 6 Elements

---

1. Data type(s) and metadata (data description)
2. Related tools, software, and/or code
3. Standards for the data/metadata
4. Data preservation, access, and associated timelines
5. Access, distribution, or reuse considerations
6. Oversight of data management and sharing

Heads up! We are going to go slightly out of order!
1. **Data Type(s) - Describe your data**

---

**Element 1: Data Type**

**A. Types and amount of scientific data expected to be generated in the project:**
- Clinical Ex: Clinical data include demographic data, insurance status, medical history, medications, lab tests performed by the clinical site and central laboratory, and physical exams data, among other data pertinent to the study...

**B. Scientific data that will be preserved and shared, and the rationale for doing so:**
- Clinical Ex: Clinical data that will be preserved and shared are demographic data, insurance status, medical history, medications, lab tests performed by the clinical site and central laboratory, and physical exams data, among other data pertinent to the study...

**C. Metadata, other relevant data, and associated documentation:**
- Clinical Ex: The protocol, sample informed consent, case report forms, data dictionary, and code book will be made accessible in data repositories where data are shared...

*Your turn!* Take 5 minutes to jot down an outline of what you would write in Element 1: Data Type
2. Related Tools - **What tools are you using with your data?**

---

**Element 2: Related Tools, Software and/or Code**

A. *Genomic and Clinical Ex:* The basic statistical analyses described in the application will be done using R. We plan to use the MRI data analysis tools in the FMRIB Software Library (FSL) for multi-level modeling of group effects. BrainVoyager software will be used for anatomical segmentation to isolate regions of interest within individual subjects.

B. *Social Science Ex:* Scientific data will be processed and analyzed with STATA and shared in many widely accessible formats, including SAS, STATA, SPSS, dBase, Excel, and ASCII.

C. *Basic Science Ex:* All code and software that will be written to analyze the data will be deposited for public access and be provided as Supplementary files for any publications...

**Your turn!** Take 5 minutes to jot down an outline of what you would write in Element 2: Related Tools
4. Preservation, Access, Timeline - where are you sharing?

Element 4: Data Preservation, Access, and Associated Timelines

A. **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](#).

B. **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.

C. **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.
Selecting a data repository

Primary consideration should be given to data repositories that are discipline or data-type specific

- Search for NIH Institute and Center required or recommended data repositories
- [Scientific Data Guide to Data Repositories](#)

Use a generalist repository if no appropriate discipline or data-type-specific repository is available

- UC’s [Dryad Digital Repository](#) makes data openly available to the larger research community
- List of [NIH generalist repositories](#) (GREI), including [Dryad](#)
- Small datasets (up to 2 GB in size) may be included as supplementary material for articles submitted to PubMed Central
4. Preservation, Access, Timeline

Element 4: Data Preservation, Access, and Associated Timelines

A. Repository where scientific data and metadata will be archived:
   ○ Human Genomic Ex: DASH (Clinical and laboratory data); dbGaP/Sequence Read Archive (SRA, Genomic data)

B. How scientific data will be findable and identifiable:
   ○ Human Genomic Ex: Clinical and laboratory data will be findable and identifiable using a DOI created by DASH. Genomic data will be findable and identifiable using a dbGaP study accession number and SRA sequence record accession numbers generated by NCBI...

C. When and how long the scientific data will be made available:
   ○ Human Genomic Ex: The study team will submit an initial batch of processed and cleaned clinical and laboratory data to DASH after 500 participants have been enrolled...

Your turn! Take 5 minutes to jot down an outline of what you would write in Element 4: Preservation, Access, Timeline
Element 5: Access, Distribution, or Reuse Considerations

A. Factors affecting subsequent access, distribution, or reuse of scientific data:
   ○ Genomic and Clinical Ex: Research participants will be consented for data sharing of their individual genomic and clinical data via controlled access. Our institution will provide an Institutional Certification upon registering the study in dbGAP...

B. Whether access to scientific data will be controlled:
   ○ Genomic and Clinical Ex: Individual-level genomic and clinical data will be shared via controlled-access...

C. Protections for privacy, rights, and confidentiality of human research participants:
   ○ Genomic and Clinical Ex: Data will be de-identified according to HIPAA and the Common Rule. Participants will have the opportunity to opt-out of such sharing or to withdraw their data from the database by contacting the study team or the university’s research administration office...

Your turn! Take 5 minutes to jot down an outline of what you would write in Element 5: Access, Distribution, Reuse
3. Standards - How will this work be described and organized?

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.
Finding standards

- Some fields have community-developed standards while others do not. Indicate if no standards have been established.
- Places to find standards:
  - Your selected data repository
  - NIH Common Data Elements (clinical data)
  - Fairsharing.org
  - DCC Disciplinary Metadata
  - Pubmed/Google search
3. Standards: Activity

Element 3: Standards:

- **Social Science Ex:** To facilitate data use, the study will use standard processing and documentation protocols adopted by the Inter-university Consortium for Political and Social Research (ICPSR) for data formats and dictionaries as well as for variable names, descriptions, and labels...

- **Basic Science Ex:** We will use the standards that are adopted or defined by the Neuroscience Multi-omic Archive (NeMO)

- **Clinical Ex:** Data will be standardized to CDISC format whenever possible. Medical laboratory data will be standardized using LOINC (Logical Observation Identifiers Names and Codes)...

*Your turn!* Take 5 minutes to jot down an outline of what you would write in Element 3: Standards
Element 6: Oversight of Data Management and Sharing

- Ex: Monitoring of and compliance with this plan will be the responsibility of the project's Principal Investigator. The plan will be implemented and managed by professional staff working under the direction of the PI...

- Ex: Data will be submitted by a project data manager from the PI’s project team. The data manager will oversee data collection, analysis, storage, and sharing. Compliance with the plan will be monitored by the PI routinely...
Today’s Takeaways

● New NIH requirement for data management and sharing now in effect
● There are templates and sample plans available - use them!
● Help is available from your campus library and office of research
  ○ Start here for your campus library data contacts
  ○ Local campus policy pages
    ■ UCSF NIH Policy Overview page
    ■ UCI NIH Data Sharing Policy Webpage
What questions do you have?