# NIH Data Management and Sharing: Q&A

## Office of Research Administration
## JHU Data Services
## Welch Library

# Webinar Agenda

I. ORA Overview of the NIH Data Management & Sharing Policy
   Tom Burns, Associate Dean for Research Affairs, Office of Research Administration

II. JHU Data Services Overview
   Chen Chiu, Sr. Data Management Consultant, JHU Data Services

III. Policy Guide Overview and DMP Tool Demonstration
   Nancy Shin, Scholarly Communications Librarian, Welch Library

IV. Q&A

# NIH Data Management and Sharing Policy

# Policy

On or after January 25, 2023, all proposals for research that will generate **scientific data** must include a data management and sharing (DMS) plan for how scientific data will be preserved and shared.

Institutions and Investigators expected to:

- Plan and budget for the managing and sharing of data
- Submit a DMS plan for review when applying for funding
- Comply with the approved DMS plan

# Scientific Data

**Scientific Data** defined as: the recorded factual material commonly accepted in the scientific community as of sufficient quality to validate and replicate research findings, regardless of whether the data are used to support scholarly publications.

**Scientific data** do not include laboratory notebooks, preliminary analyses, completed case report forms, drafts of scientific papers, plans for future research, peer reviews, communications with colleagues, or physical objects, such as laboratory specimens

# Applicability

The DMS Policy applies to all research that generates scientific data, including:

- Research Projects
- Some Career Development Awards (Ks)
- Small Business SBIR/STTR
- Research Centers

Applies regardless of funding mechanism (contract, grant, cooperative agreement, etc.) DMS plan becomes a term of the award.

The DMS Policy does not apply to research and other activities that do not generate scientific data, including:

- Training (T)
- Fellowships (Fs)
- Construction (C06)
- Conference Grants (R13)
- Resource (Gs)
- Research-Related Infrastructure Programs (e.g., S06)

# Elements of DMS Plan

- Data Type
- Related Tools, Software or Code
- Standards
- Data Preservation, Access and Associated Timelines
  - Data subject to the Genomic Data Sharing Policy
- Access, Distribution, or Reuse Considerations
- Oversight of Data Management and Sharing
- Recommended to be 2 pages in length
- Format on NIH DMS site  Fillable version in Fall 2022

PREVIEW – DRAFT

## DATA MANAGEMENT AND SHARING PLAN

If any of the proposed research in the application involves the generation of scientific data, this application is subject to the NIH Policy for Data Management and Sharing and requires submission of a Data Management and Sharing Plan. If the proposed research in the application will generate large-scale genomic data, the Genomic Data Sharing Policy also applies and should be addressed in this Plan. Refer to the detailed instructions in the application guide for developing this plan as well as to additional guidance on sharing.nih.gov. The Plan is recommended not to exceed two pages. Text in italics should be deleted. There is no "form page" for the Data Management and Sharing Plan. The DMS Plan may be provided in the *format* shown below.

**Element 1: Data Type**

**A. 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.*

**B. 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.*

**C. 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.*

**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.*

**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.*

**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.*

PREVIEW – DRAFT

**Element 5: Access, Distribution, or Reuse Considerations**

**A. 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.*

**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).*

**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).*

**Element 6: 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).*

# Sample DMS Plans

| Sample | Description | NIH Institute or Center |
|---|---|---|
| Sample Plan A | Clinical and/or MRI data from human research participants | NIMH |
| Sample Plan B | Genomic data from human research participants | NIMH |
| Sample Plan C | Genomic data from a non-human source | NIMH |
| Sample Plan D | Secondary Data Analysis | NIMH |

# Assessment/Revision of DMS Plan

- Program staff will review and assess DMS Plan
- Applications will be only be funded if DMS Plan is acceptable
- Peer reviewers may review DMS Plan but not be asked to comment or factor into score.
- Revisions to DMS Plan may be requested at the JIT stage.
- Post-award revisions may be required if elements of the plan change.
- Compliance with DMS Plan will be monitored through the RPPR process.

# Sharing Human Subjects Data

- Address data management and sharing plans during the informed consent process to ensure prospective participants understand how their data will be managed and shared;
- Outline steps they will take for protecting the privacy, rights, and confidentiality of prospective participants (i.e., through de-identification, Certificates of Confidentiality, and other protective measures);
- Assess limitations on subsequent use of data and communicate these limitations to the individuals or entities (e.g., repositories) preserving and sharing the data; and
- Consider whether access to shared scientific data derived from humans should be controlled, even if de-identified and lacking explicit limitations on subsequent use. Sharing via controlled access may be specified by certain funding opportunity announcements (FOAs) or the funding NIH Institutes or Centers.
- **Not justifiable for limiting data sharing**:
  - Data are considered to be too small
  - Data that researchers anticipate will not be widely used
  - Data are not thought to have a suitable repository

# Timeline for Sharing

- The time of an associated publication: Scientific data underlying peer-reviewed journal articles should be made accessible no later than the date on which the article is first made available in print or electronic format.

OR

- The end of the performance period: Scientific data underlying findings not disseminated through peer-reviewed journal articles should be shared by the end of the performance period unless the grant enters into a no-cost extension. If a no cost extension is permitted, then the recipient should share the data by the end of the extended performance period. These scientific data may underlie unpublished key findings, developments, and conclusions; or findings documented within preprints, conference proceedings, or book chapters. For example, scientific data underlying null and negative findings are important to share even though these key findings are not always published. Researchers should be aware that some preprint servers may require the sharing of data upon preprint posting, and repositories storing data may similarly require public release of data upon preprint posting.

# Selecting a Repository

- Encourages use of established repositories
- NIH IC may identify specific repositories in the FOA
- Primary consideration should be given to data repositories that are discipline or data-type specific.
- List of NIH-supported repositories

# Allowable Costs for DMS Plan

- Curating data
- Developing supporting documentation
- Formatting data according to accepted community standards, or for transmission to and storage at a selected repository for long-term preservation and access
- De-identifying data
- Preparing metadata to foster discoverability, interpretation, and reuse
- Local data management considerations, such as unique and specialized information infrastructure necessary to provide local management and preservation (for example, before deposit into an established repository).
- Preserving and sharing data through established repositories, such as data deposit fees.
- If the Data Management & Sharing (DMS) plan proposes deposition to multiple repositories, costs associated with each proposed repository may be included.
- **All costs must be incurred during the performance period, even where management and sharing will extend beyond the award period.**

# Unallowable Costs for DMS Plan

- Infrastructure costs that are included in institutional overhead (for instance, Facilities and Administrative costs)
- Costs associated with the routine conduct of research, including costs associated with collecting or gaining access to research data.
- Costs that are double charged or inconsistently charged as both direct and indirect costs

# Format for Budgeting

- Budget requests should be included in one of the following ways:
  - Using the R & R detailed budget form in a line item labeled “Data Management and Sharing Costs” under Other Direct Costs along with the Budget Justification attachment.
  - Using the *PHS 398 Modular Budget Form:* Use the Additional Narrative Justification attachment of the PHS 398 Modular Budget Form
- All DMS costs- including personnel costs- should be included in the Data Management and Sharing Costs line item.
- Justifications should be labeled as "Data Management and Sharing Justification" and be no more than half a page.

# Questions to Consider

- How much dedicated staff time will be required to support the data management and sharing activities proposed in the DMS plan?

- Will staff or a service center/third party vendor be performing any of the following activities:
  - Curating data
  - Developing supporting documentation
  - Formatting data according to accepted community standards, or for transmission to and storage at a selected repository for long-term preservation and access
  - De-identifying data
  - Preparing metadata to foster discoverability, interpretation, and reuse

- If service center/third party vendor is performing any of the above activities, include those costs in the DMS budget.

# Questions to Consider

- Are there data storage/repository costs applicable?
- Are there costs/fees associated with any specialized information infrastructure necessary to provide local management and preservation of the data (e.g., PMAP)?
- Are any subrecipients performing any of the DMS activities?
- **NOTE: All costs submitted in budget requests must be incurred during the period of performance of the award**- even if data sharing and storage will extend beyond termination of the award. Therefore DMS expenses may need to be prepaid prior to expiration/close out of the award.

# JHU DATA MANAGEMENT AND SHARING RESOURCES

- Biostatistics, Epidemiology and Data Management (BEAD) Core

The BEAD Core provides research support services to subscribing department investigators, and as a direct fee for service for other investigators. BEAD Core also provides assistance with grant submissions, writing data management and statistical analytic plans as well as assisting with the study design and research aims.

- Core for Clinical Research Data Acquisition

The Core for Clinical Research Data Acquisition (CCDA) is one of the 10 **Data Trust** analytic teams responsible for assisting researchers with accessing clinical data for research

- Johns Hopkins Precision Medicine Analytics Platform (PMAP)

The PMAP platform pulls data from the Epic Medical Record and other data sources into a Data Commons, where the data are integrated together and available in a format that is operable by sophisticated machine learning and natural language processing technologies.

- JHU Data Archive

JHU Data Archive is an open access repository for the long-term management and preservation of research data. Through depositing datasets in the JHU Data Archive, researchers are able to share their research data with the public for future discovery and reuse. The JHU Data Archive is administered by professional curators, who will work with you to ensure your data is Findable, Accessible, Interoperable and Reusable (FAIR).

# BEAD Core Recommendations

- R01-level funding:
  - 40% of a data manager per year
  - 2-5% Co-I to supervise/guide data manager
- R21-level funding:
  - 20% of a data manager per year
  - 2% Co-I to supervise/guide data manager
- U's (such as a U54) (typically three RO1s combined)
  - 100% of a data manager (some efficiency gained, but depends on the three projects included)
  - 10% of a Co-I
- K-level funding:
  - 20-40 hours per year ($120/hr)

# Upcoming Sessions

ORA/Data Services Sessions:

January 10, 2023  12 pm
January 19, 2023  12 pm

# OTHER RESOURCES

- NOT-OD-21-013: Final NIH Policy for Data Management and Sharing
- NOT-OD-21-014: Supplemental Information on Elements of an NIH Data Management and Sharing Plan
- NOT-OD-21-015: Allowable Costs for Data Management and Sharing
- NOT-OD-22-213: Supplemental Information to the NIH Policy for Data Management and Sharing: Protecting Privacy When Sharing Human Research Participant Data
- NIH Data Management and Sharing Policy website
- Sample DMS Plans – to be used for educational purposes not as templates.
- Repositories for Sharing Scientific Data
- NIH Guidance on Budgeting for Data Management and Sharing
- DMPTool

# Support for NIH Data Management and Sharing Plans

**Chen Chiu**, Sr. Data Management Consultant

January 10th, 2023

**Dave Fearon**, Sr. Data Management Consultant

**Betsy Gunia**, Data Management Consultant

JOHNS HOPKINS
LIBRARIES

Data Services

JHU DATA SERVICES

# Roadmap

- About JHU Data Services
- Introduction to DMPTool
- Data sharing and repository
- Resources for NIH DMSP

JHU DATA SERVICES

# HELPING YOU NAVIGATE DATA

## WE HELP FACULTY, RESEARCHERS AND STUDENTS

FIND

USE

MANAGE

VISUALIZE

SHARE

FIND OUT MORE

**GO TO** dataservices.library.jhu.edu

**EMAIL** dataservices@jhu.edu

**SHARE AT** archive.data.jhu.edu

JOHNS HOPKINS LIBRARIES | Data Services

# Data Management Services at JHU

- Assist with data management plans (DMPs) for research grants
- Facilitate compliance with funder and journal data sharing policies
- Guidance on preparing data for online access, operating the [Johns Hopkins Research Data Repository]

JOHNS HOPKINS
LIBRARIES
Data Services



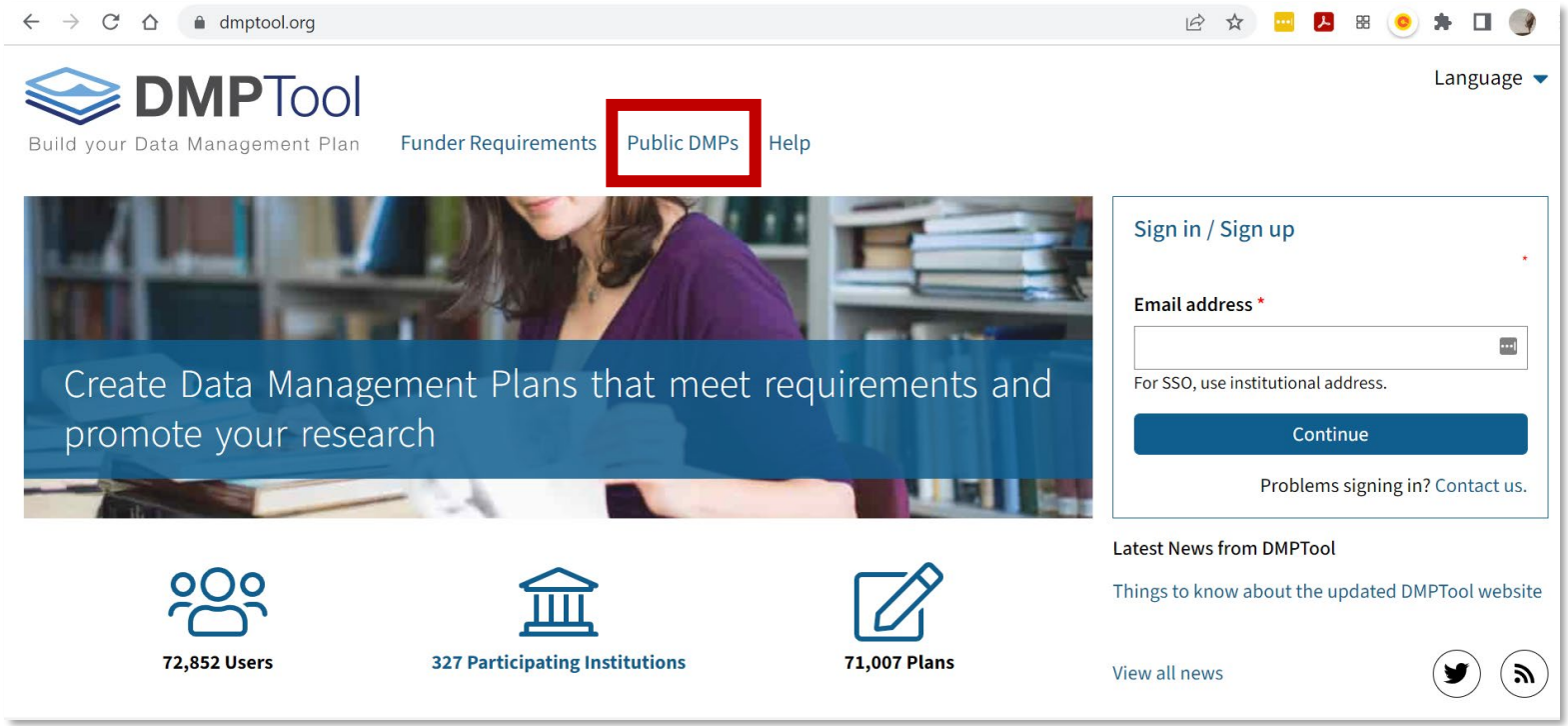

# DMPTool

# What is DMPTool?

https://dmptool.org/

- A free, online tool that provides funder-specific templates for Data Management Plans (DMPs)
- Leads researchers through the DMP writing process
  - Breaks down DMPs into relevant sections, tailored to your funder
  - Expert guidance within the tool
  - Request feedback from JHU Data Services
- 300+ participating institutions worldwide (including JHU)

# Sign in DMPTool



# Create a New Plan

## Create a new plan

Before you get started, we need some information about your research project to set you up with the best DMP template for your needs.

### * What research project are you planning?

NIH DMSP workshop demo

☐ mock project for testing, practice, or educational purposes

### * Select the primary research organization

**Research organization**

Johns Hopkins University (jhu.edu)

**- or -** ☐ No research organization associated with this plan or my research organization is not listed

### * Select the primary funding organization

**Funder**

National Institutes of Health (nih.gov)

**- or -** ☐ No funder associated with this plan or my funder is not listed

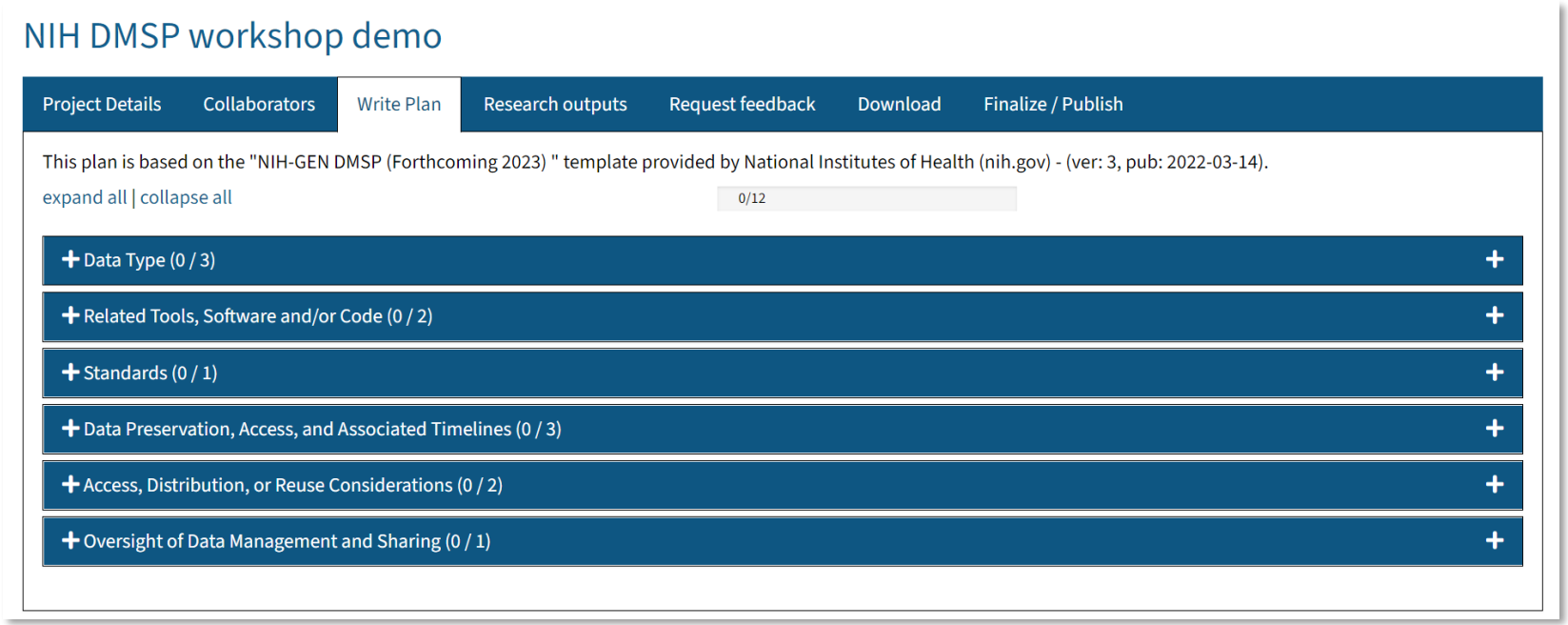
### Which DMP template would you like to use?

NIH-GEN DMSP (Forthcoming 2023)

We found multiple DMP templates corresponding to your funder.

Create plan | Cancel

# Write Your Plan



## + Data Type (0 / 3)

Briefly describe the scientific data to be managed, preserved, and shared.

**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)**

B *I*

Box to type your text

Save

NIH Fill-In-The-Blank

**NIH example answer**

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.

Comments & Guidance

Guidance | Comments

NIH | JHU | DMPTool

**NIH Guidance**

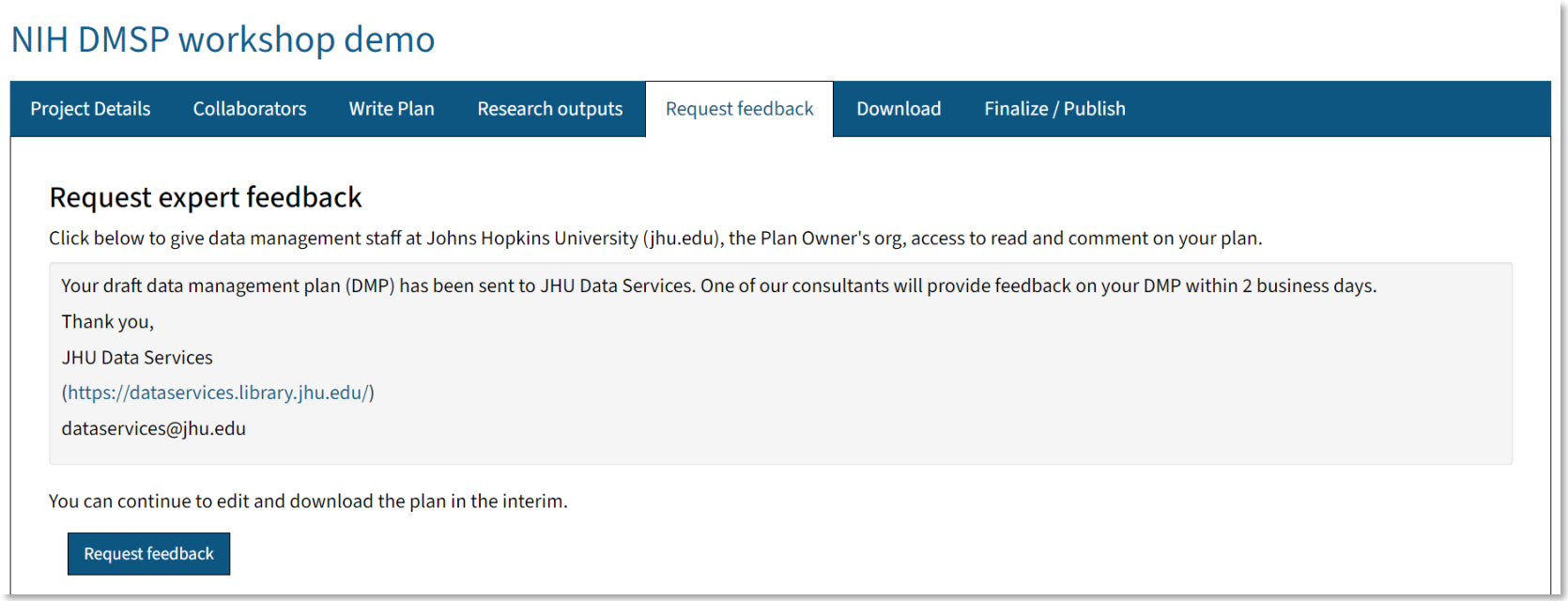
The final DMS Policy defines Scientific Data as: "The recorded factual material commonly accepted in the scientific community as of sufficient quality to validate and replicate research findings, regardless of whether the data are used to support scholarly publications. Scientific data do not include laboratory notebooks, preliminary analyses, completed case report forms, drafts of scientific papers, plans for future research, peer reviews, communications with colleagues, or physical objects, such as laboratory specimens."

Even those scientific data not used to support a publication are considered scientific data and within the final DMS Policy's scope. We understand that a lack of publication does not necessarily mean that the findings are null or negative; however, indicating that scientific data are defined independent of publication is sufficient to cover data underlying null or negative findings.

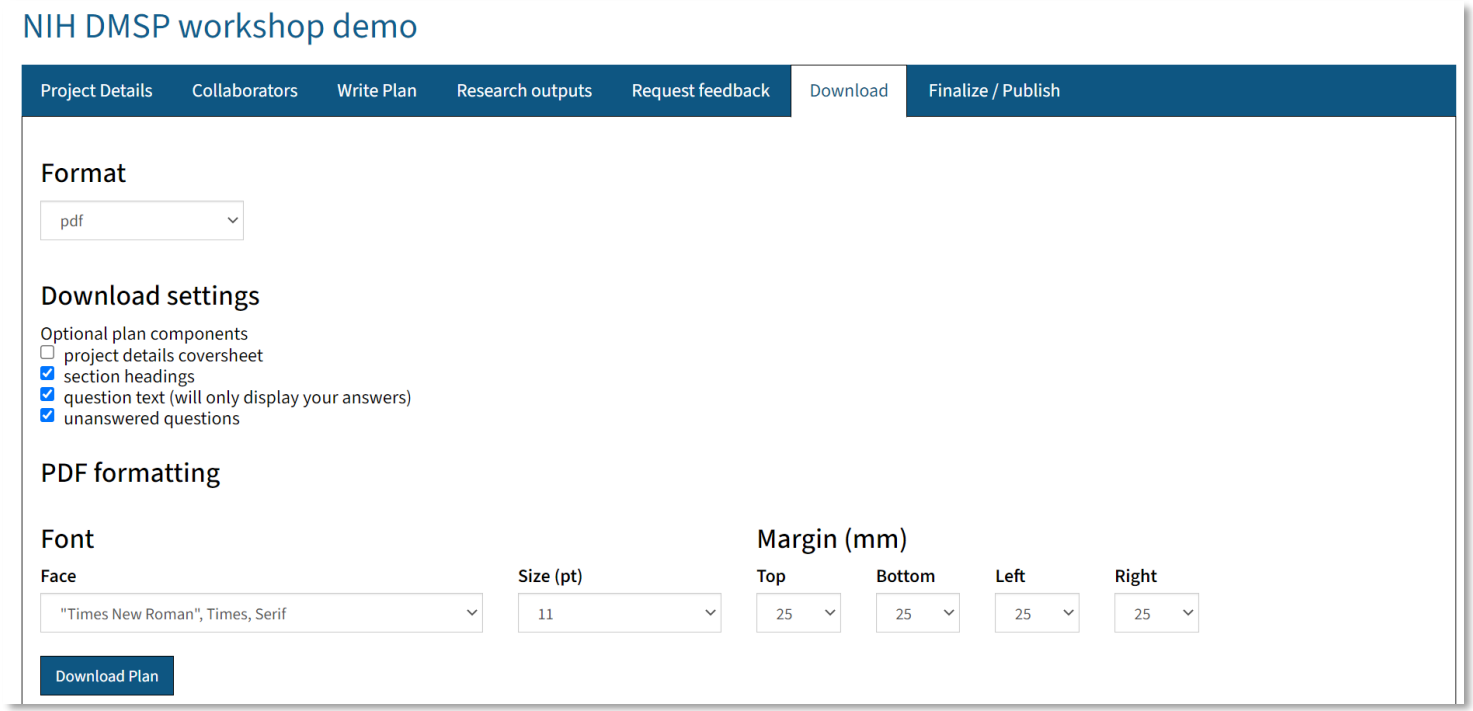
**Additional Guidance**

Research projects vary widely in the types of data produced. In this section, you will describe the categories, amounts, and degree of processing of your data.

# Send Your Plan for Feedback



# Download Your Plan



# Help with DMPs and DMPTool

- NIH Data Management and Sharing Plans support (flyer and website)
- View our online module: Using the DMPTool to Write your Plan
- Email Data Services: dataservices@jhu.edu
- Live webinars and recordings
  - Writing an NIH Data Management and Sharing Plan (recording)
    - Wednesday 1/25, 12-1pm (register)
  - Writing a Data Management Plan with DMPTool (recording)
    - Wednesday 1/25, 1:30-2:30pm (register)

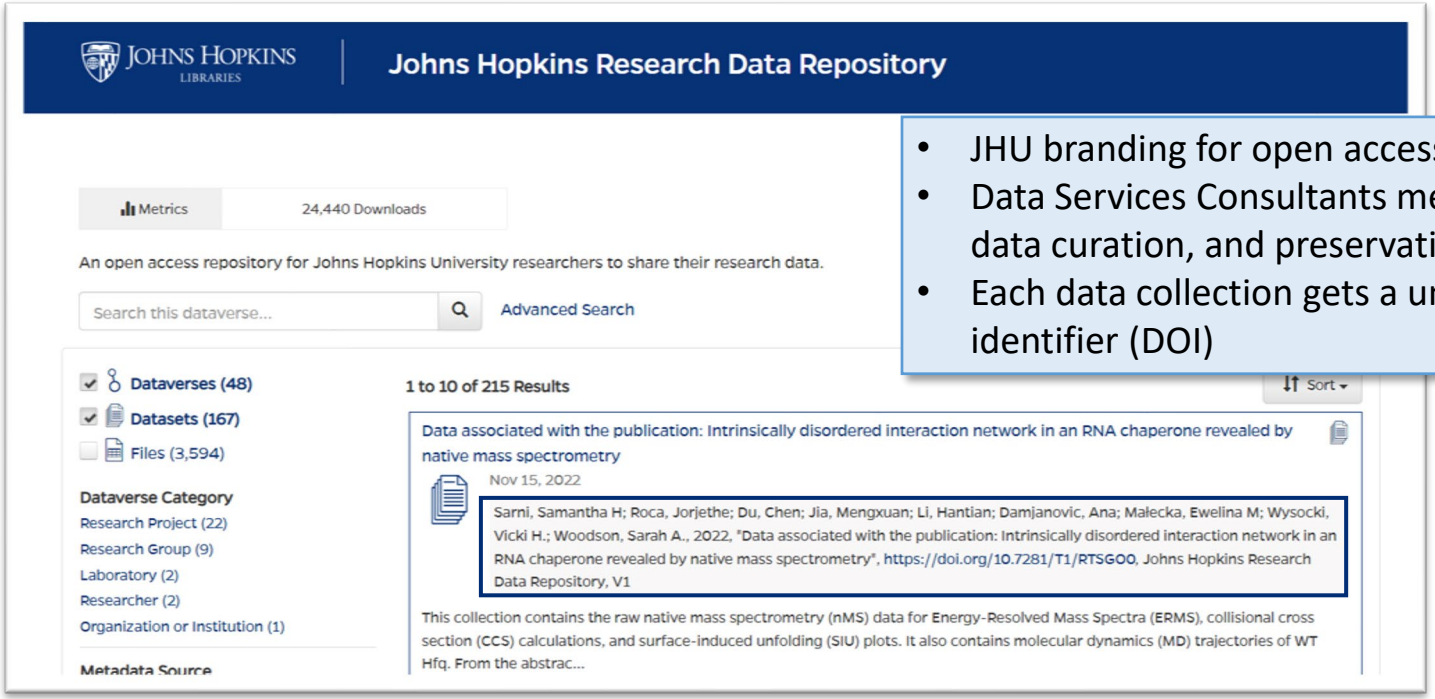
# Where to Find Sample Plans?

- DMPTool's ongoing collection of publicly shared data management plans that can be filtered by funder, institution and subject
- DMP examples from University of Arizona
- Sample NIH Data Management and Sharing plans for Clinical, Secondary, and Genomic research from NIMH
- Example DMS Plans on GitHub

JOHNS HOPKINS
LIBRARIES
Data Services

# Data Sharing

# Open Access: Johns Hopkins Research Data Repository




- JHU branding for open access research data
- Data Services Consultants mediate deposit, data curation, and preservation
- Each data collection gets a unique persistent identifier (DOI)

Johns Hopkins Research Data Repository information and FAQ http://dataservices.library.jhu.edu/archiving

# Controlled-Access: ICPSR and Vivli


- A data repository for research in the social, behavioral sciences and public health data
- With the options of Open or Controlled-access data repositories
- JHU is one of their member institutions

https://www.icpsr.umich.edu/icpsrweb/ICPSR/


- A controlled-access data repository for clinical research data
- JHU is one of their member institutions

https://vivli.org/

# De-Identification Tools

NIST National Institute of Standards and Technology U.S. Department of Commerce

NIST maintains a list of deidentification tools

Protocols on how they de-identify DICOM images and list of software used

CANCER IMAGING ARCHIVE

NIH National Library of Medicine Lister Hill National Center for Biomedical Communications

NLM Scrubber: Freely available clinical text deidentification tool

Applications to assist in de-identification of human subjects research data

JOHNS HOPKINS LIBRARIES | Data Services

# Consent Form Language for Sharing Data

For publishing medical data to an online repository, it is usually essential to get participant's consent, even for restricted repositories and for most de-identified datasets

- State where de-identified data will be shared
- State to whom these de-identified data will be shared
- Inform them of small risk that data may be re-identified

**Work with your IRB to develop a consent form for data sharing**

# Consent Form Language for Sharing Data

- JHU SOM IRB
  - Boilerplate for data sharing
  - Organization Policy on Informed Consent Process and Documentation
- JHU BSPH IRB
  - Consent form templates
- NIH's Informed Consent template (link)
  - Always get approval from your IRB about the consent language

# Contact JHU Data Services

**GO TO**
dataservices.library.jhu.edu

**EMAIL**
dataservices@jhu.edu

**SHARE DATA AT**
archive.data.jhu.edu

## Helping you


DATA

FIND OUT **More**

https://youtu.be/r4VqNMw6Q8s

JOHNS HOPKINS LIBRARIES | Data Services

# Overview of NIH Data Management and Sharing Policy Guide

Nancy Shin, MLIS, AHIP

*Scholarly Communications Librarian*
*Welch Medical Library*