National Institutes of Health
Office of Science Policy

NIH Policy for Data Management & Sharing: What is expected of researchers and how can libraries help?

Presentation to University of Chicago community for Love Data Week
February 16, 2023

Ellen Wann, PhD
Data Management & Sharing Policy Team Lead
Scientific Data Sharing Policy Division
Office of Science Policy
National Institutes of Health

Cindy Danielson, PhD
Associate Director, Systems Integration,
Office of Research Reporting and Analysis
Office of Extramural Research
National Institutes of Health

Lisa Federer, PhD
Acting Director
Office of Strategic Initiatives
National Library of Medicine
National Institutes of Health

Why does NIH Want Data to be Shared?

- **Advance rigorous and reproducible research**
  - Enable validation of research results
  - Make high-value datasets accessible
  - Accelerate future research directions
  - Increase opportunities for citation and collaboration

- **Promote public trust in research**
  - Foster transparency and accountability
  - Demonstrate stewardship over taxpayer funds
  - Maximize research participants' contributions
  - Support appropriate protections of research participants' data

Data Accessibility: Still Work to Do

"Data sharing practices and data availability upon request differ across scientific disciplines," Tedersoo et al., (2021)

- Evaluated data availability in **875 papers across nine disciplines** published 2000-2019
- Data **obtained from authors in 39.4% of requests** on average; ranged 27.9–56.1% among research fields, improved with repeated follow-up, **19.4% of requests declined**

"Reproducibility in Cancer Biology: Challenges for assessing replicability in preclinical cancer biology," Errington et al., (2021)
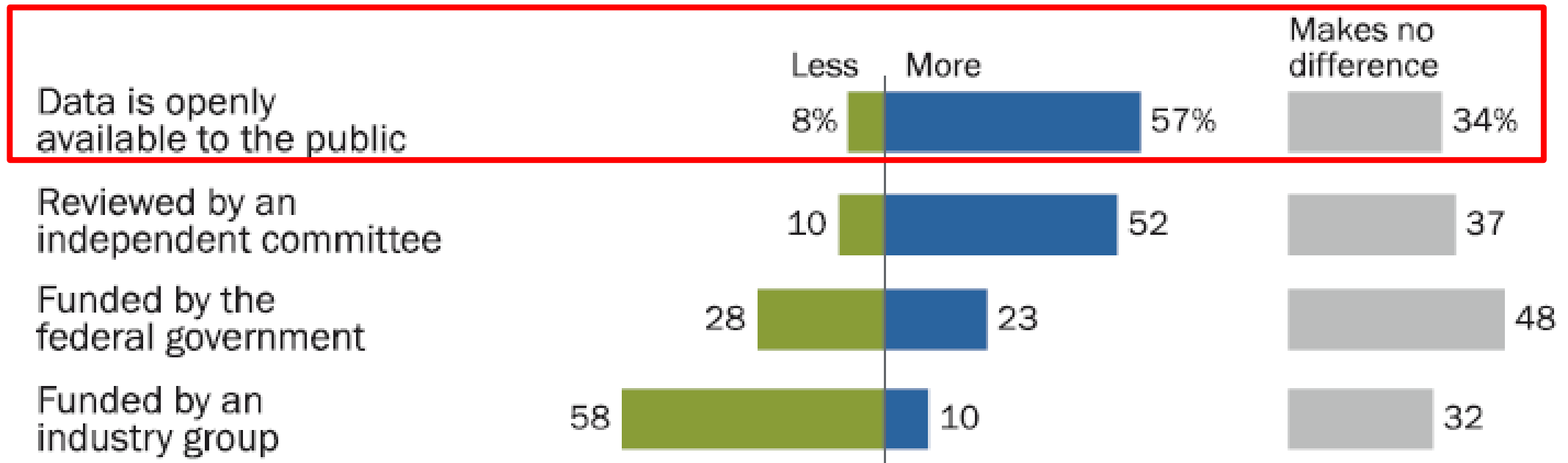
- Attempted to **repeat 193 experiments from 53 high-impact cancer biology papers;** unable to obtain data for **68% of experiments**

"Many researchers were not compliant with their published data sharing statement: mixed-methods study," Gabelica et al., (2022)

- Requested data from **1,792 BioMed Central papers** published January 2019 with data availability statements
- 93% of authors did not respond or declined to share; **only 6.8% provided the requested data**
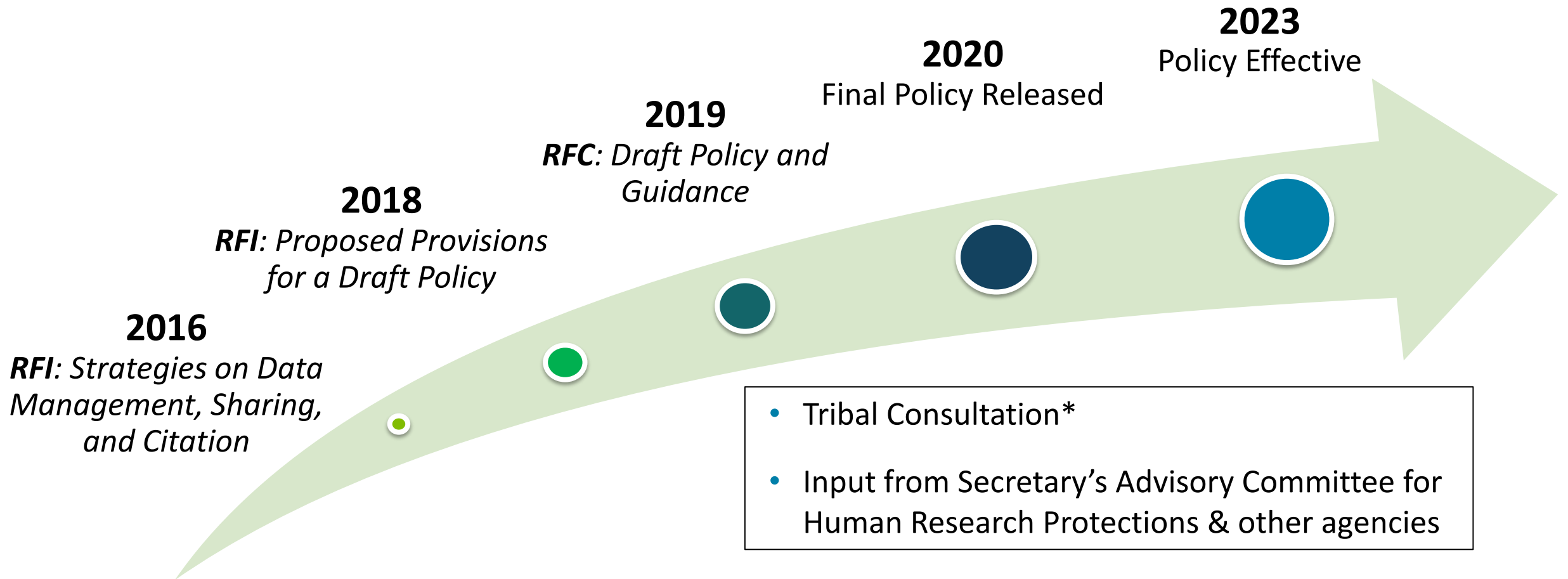
A Matter of Trust

% of U.S. adults who say when they hear each of the following, they trust scientific research findings ...

	Less	More	Makes no difference
Data is openly available to the public	8%	57%	34%
Reviewed by an independent committee	10	52	37
Funded by the federal government	28	23	48
Funded by an industry group	58	10	32

https://www.pewresearch.org/science/wp-content/uploads/sites/16/2019/08/PS_08.02.19_trust.in_.scientists_FULLREPORT.pdf

Iterative Policy Development through Consistent Community Engagement

2023
Policy Effective

2020
Final Policy Released

2019
RFC: Draft Policy and Guidance

2018
RFI: Proposed Provisions for a Draft Policy

2016
RFI: Strategies on Data Management, Sharing, and Citation

- Tribal Consultation*

- Input from Secretary's Advisory Committee for Human Research Protections & other agencies

*See "NIH Tribal Consultation Report: NIH Draft Policy for Data Management and Sharing"

NIH Policy for Data Management and Sharing

- **Submission of Data Management & Sharing Plan for all NIH-funded research** *(how/where/when)*

- **Compliance with the ICO-approved Plan** *(may affect future funding)*

- **Effective January 25, 2023** *(replaces 2003 Data Sharing Policy)*

Activities Subject to the DMS Policy

- **Applies to all research generating scientific data**, including but not limited to:
  - Research Projects
  - Some Career Development Awards (Ks)
  - Small Business SBIR/STTR
  - Research Centers
- **Does not apply to research projects <u>not</u> generating scientific data or non-research projects**, including but not limited to:
  - Training (Ts)
  - Fellowships (Fs)
  - Construction (C06)
  - Conference Grants (R13)
  - Resources (Gs)
  - Research-Related Infrastructure Programs (e.g., S06, S10)

See **Research Covered Under the Data Management & Sharing Policy**

Details [of the Policy] Matter!

- **Scope:** All NIH-supported research generating *scientific data*

  - **What's in: "**Recorded factual material... of <u>sufficient quality to validate and replicate research findings</u>, regardless of whether the data are used to support scholarly publications"—relates to the proposed research questions and findings can include unpublished null results

    - May include qualitative data or data produced using fundamental basic science techniques

  - **What's out:** lab notebooks, preliminary analyses, case report forms, physical objects

- **Timelines:**

  - **When to share data?** no later than <u>publication</u> or <u>end of award</u> (for data underlying findings not published in peer-reviewed journals)

  - **How long to share data?** consider other relevant requirements and expectations (e.g., journal policies, repository policies)

Additional Expectations for Plans

- **SHARING SHOULD BE …**

  - **The default practice**

    - Data sharing should be maximized (with justifiable limitations)

    - All data should be managed; **not all must be shared**

  - **Responsibly implemented**

    - Plans should outline protection of privacy, rights, and confidentiality

    - Abide by existing laws, regulations, and policies

  - **Prospectively planned for at all stages of the research process**

Potential Limitations on Sharing

- **Data Management and Sharing Plans should <u>maximize appropriate</u> sharing:**

  - **Justifiable ethical, legal, and technical factors for limiting sharing of data include:**

    - Informed consent will not permit or limits scope of sharing or use

    - Privacy or safety of research participants would be compromised and available protections insufficient

    - Explicit federal, state, local, or Tribal law, regulation, or policy prohibits disclosure

    - Restrictions imposed by existing or anticipated agreements with other parties

    - Datasets cannot practically be digitized with reasonable efforts

  - **Reasons <u>not</u> generally justifiable to limit sharing include:**

    - Data are considered too small

    - Researchers anticipate data will not be widely used

    - Data are not thought to have a suitable repository

  - **Additional considerations:**

    - NIH respects Tribal sovereignty and supports responsible management/sharing of AI/AN participant data

    - SBIR/STTR Program Policy Directive permits withholding data for 20 years, as stipulated in agreements and consistent with program goals

Supplemental Information: Elements of a Data Management and Sharing Plan

- **Data type**
  - Data and metadata to be preserved and shared, rationale for doing so, any associated documentation
- **Related tools, software, code**
  - Tools and software needed to access and manipulate data
- **Standards**
  - Standards to be applied to scientific data and metadata
- **Data preservation, access, timelines**
  - Repository to be used, persistent unique identifier, and when/ how long data will be available
- **Access, distribution, reuse considerations**
  - Description of factors for data access, distribution, or reuse, including whether data will be controlled
- **Oversight of data management**
  - Plan compliance will be monitored/ managed and by whom

See **Writing a Data Management & Sharing Plan** for details

Format of a Data Management and Sharing Plan

- Optional DMS Plan format page available on list of Format Pages and incorporated into FORMS-H application instructions

- Plans recommended to be no more than 2 pages in length

- Federal Demonstration Partnership pilot project to test structured templates and tools for DMS Plan submission

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.

Public reporting burden for this collection of information is estimated to average 2 hours per response, including the time for reviewing instructions, searching existing data sources, gathering, and maintaining the data needed, and completing and reviewing the collection of information. An agency may not conduct or sponsor, and a person is not required to respond to, a collection of information unless it displays a currently valid OMB control number. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to: NIH, Project Clearance Branch, 6705 Rockledge Drive, MSC 7974, Bethesda, MD 20892-7974, ATTN: PRA (0925-0001 and 0925-0002). Do not return the completed form to this address.

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.

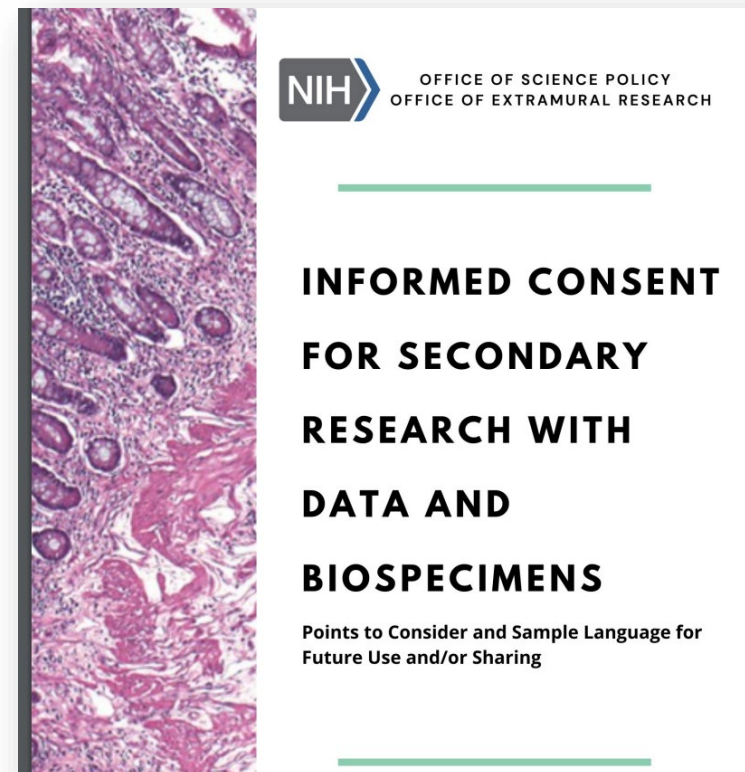# Supplemental Information: Repository Selection

- **Encourages use of established repositories**

- **Helps investigators identify appropriate data repositories**
  - E.g., use of persistent unique identifiers, attached metadata, facilitates quality assurance

- **NIH ICs may designate specific data repository(ies)**

See Selecting a Data Repository for details

Informed Consent and DMS Policy

- Policy encourages researchers and institutions to establish robust consent processes, but:
  - Does not establish additional consent expectations
  - Does not require consent be obtained any particular way (e.g., broad consent)

- Policy recognizes limitations on data sharing based on the informed consent process

- Informed Consent Resources:
  - Points to consider
  - Sample language for future use and/or data sharing

NIH> OFFICE OF SCIENCE POLICY
OFFICE OF EXTRAMURAL RESEARCH

INFORMED CONSENT

FOR SECONDARY

RESEARCH WITH

DATA AND

BIOSPECIMENS

Points to Consider and Sample Language for Future Use and/or Sharing

Informed Consent Resource

Supplemental information for Protecting Privacy When Sharing Human Research Participant Data

- Provides a basic **framework for considering how to protect privacy** when sharing data from human participants

- Not intended as a guide for regulatory compliance

- **Broadly applicable** to different research contexts

- Establishes shared principles, provides best practices, and offers considerations for determining **whether to control access to data**

NOT-OD-22-213

Supplemental Information: Responsible Management and Sharing of American Indian/ Alaska Native Participant Data

- Information to assist in developing appropriate DMS Plans

- Emphasizes:

  - ✓ Respect for Tribal Sovereignty

  - ✓ Partnerships and mutual agreements

  - ✓ Building trust

- Developed through Tribal Consultation and stakeholder engagement beginning in 2019

NOT-OD-22-214

Supplemental Information: Allowable Costs

- **Reasonable costs allowed in budget requests (must be incurred during the performance period)**
  - Curating data/developing supporting documentation
  - Preserving/sharing data through repositories
  - Local data management considerations

- **<u>NOT</u> considered data sharing costs**
  - Infrastructure costs typically included in indirect costs
  - Costs associated with the routine conduct of research (e.g., costs of gaining access to research data)

- Over time NIH **hopes to learn more about what constitutes reasonable costs** for various data management and sharing activities

See **Budgeting for Data Management & Sharing** for details

Plan Submission and Review: A Guide

Extramural Grant Awards*

Plan Submission

With application

Brief Plan description in Budget Justification

Full Plan as separate attachment

Plan Assessment

Peer reviewers comment on (not score) budget

NIH program staff assess Plans

Plans can be revised

Plan Compliance

Incorporated into Terms and Conditions

Monitored at regular reporting intervals – mechanisms and tools to support oversight under development

Compliance may factor into future funding decisions

*Analogous requirements for contracts, Other Transaction Awards, NIH Intramural Research Program

Changes to DMS Plan Submission

- Updates made to FORMS-H version of grant application forms – updated instructions available now
- New "Other Plan(s)" field added to PHS 398 forms to collect a single DMS Plan PDF attachment
  - Separate "Data Sharing Plan" and "Genomic Data Sharing Plan" attachments no longer collected as Resource Sharing Plan(s)
  - Plans for sharing genomic data included as part of the DMS Plan in Other Plan(s)

Research Plan Section

5. Vertebrate Animals   Add Attachment   Delete
6. Select Agent Research   Add Attachment   Delete
7. Multiple PD/PI Leadership Plan   Add Attachment   Delete
8. Consortium/Contractual Arrangements   Add Attachment   Delete
9. Letters of Support   Add Attachment   Delete
10. Resource Sharing Plan(s)   Add Attachment   Delete
11. Other Plan(s)   Add Attachment   Delete
12. Authentication of Key Biological and/or Chemical Resources   Add Attachment   Delete Attachment   View Attachment

⚠ **DMS Plan will NOT be visible to peer reviewers.**

except for certain FOAs with a data sharing focus

What About the Resource Sharing Plan(s) Field?

- Existing "Resource Sharing Plan(s)" field will remain to support Research Tools and Model Organism Sharing Policy
  - "Resource Sharing Plan(s)" attachment will continue to be visible to peer reviewers and included as an Additional Review Consideration
- eRA validations to ensure DMS Plans are attached to appropriate form field

Submitting DMS Budgets

- Direct costs to support activities proposed in DMS Plan must be indicated as "Data Management and Sharing Costs"
  - R&R Budget Form: line item in section F. Other Direct Costs

F. Other Direct Costs	Funds Requested ($)
1. Materials and Supplies	
2. Publication Costs	
3. Consultant Services	
4. ADP/Computer Services	
5. Subawards/Consortium/Contractual Costs	
6. Equipment or Facility Rental/User Fees	
7. Alterations and Renovations	
8. Data Management and Sharing Costs	
9.	
10.	

  - PHS 398 Modular Budget Form: within Additional Narrative Justification

2. Budget Justifications				
Personnel Justification		Add Attachment	Delete Attachment	View Attachment
Consortium Justification		Add Attachment	Delete Attachment	View Attachment
Additional Narrative Justification		Add Attachment	Delete Attachment	View Attachment

Justifying DMS Budgets

- Brief summary of DMS Plan and description of DMS costs (recommended ≤ ½ page) must be included within the budget justification attachment; peer reviewers may comment on requested DMS costs based on this information
  - R&R Budget Form: section L. Budget Justification

  - PHS 398 Modular Budget Form: Additional Narrative Justification

See Instructions in FORMS-H Application Guide

sharing.nih.gov

- Provides a central source of guidance related to multiple NIH data sharing policies

- Covers Data Management and Sharing, Genomic Data Sharing, Model Organisms, and Research Tools policies

- Content will be updated

Resources for: Planning and Budgeting

- Expectations for writing a Data Management & Sharing Plan

- Budgeting for data management & sharing in your application for funding

- Includes instructions and format for submitting Plans

Home > Data Management and Sharing Policy > Planning and Budgeting for Data Management & Sharing

Planning and Budgeting for Data Management & Sharing

NIH expects applicants to submit a plan for how they will manage and share their data and allows applicants to include certain costs associated with data management and sharing in their budget.
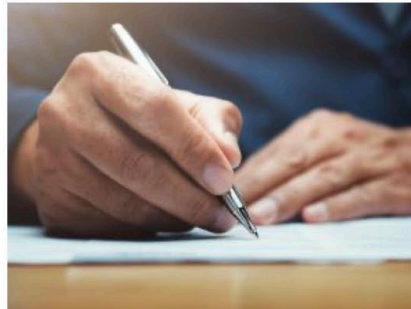
Writing a Data Management & Sharing Plan
Learn what NIH expects Data Management & Sharing plans to address.
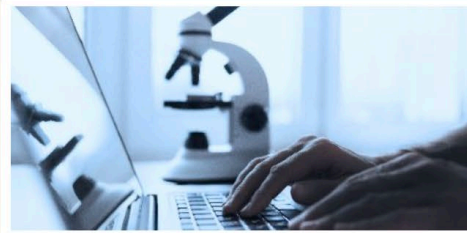
Budgeting for Data Management & Sharing
Find out what data sharing related costs may be requested in an application for funding.

Resources for: Understanding Sharing Policies

- Explore NIH sharing policies

- Tool to find which policies apply to you
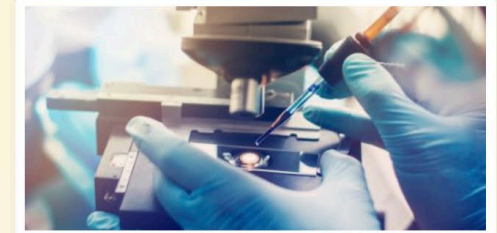
- FAQs to help understand each policy

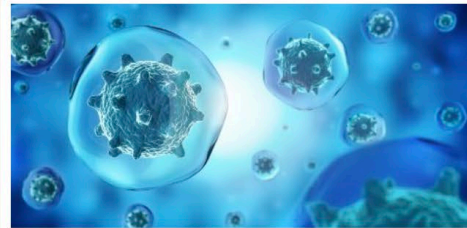**Explore the areas in which NIH has sharing policies.**

Scientific Data
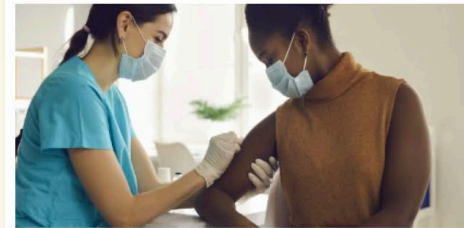
Genomic Data

Research Tools

Model Organisms

Clinical Trials ⧉

Research Publications
NIH expects that all peer-reviewed manuscripts be publicly available on PubMed Central.

Not sure where to start?

Find which policies apply to you

Resources for: Which Policies Apply to You?

- Explore NIH sharing policies

- Tool to find which policies apply to you

- FAQs to help understand each policy

Which Policies Apply to My Research?

Question

First, let's start by identifying what type of outputs your research will generate.

Some projects generate **data** such as DNA sequences, medical imaging, or statistical analyses of other data. Others may generate novel **model organisms** or **research tools** such as algorithms, mice, or laboratory equipment.

Note that activities such as training, infrastructure development, and other non-research activities may not directly generate research outputs at all.

Your research may produce more than one of these types of outputs. Please go through the decision tool for one output at a time and then select the "start over" button to explore additional outputs.

- DATA
- MODEL ORGANISMS
- RESEARCH TOOLS
- NONE OF THE ABOVE

Need help answering this question? See these resources:
Definitions of Model Organisms and Related Resources
Definition of Research Tools

BACK   NEXT

Sample NIH DMS Plans Available

- 10+ sample NIH DMS Plans available for educational purposes, including:

  – Human clinical and/or MRI data (NIMH)

  – Human genomic data (NIMH, NHGRI, NIDDK)

  – Human & non-human genomic data (NIMH)

  – Secondary data analysis (NIMH, NIDDK)

  – Human clinical and genomics data (NICHD)

  – Human survey data (NICHD)

  – Model organism (Zebrafish) data (NICHD)

  – Technology development (NHGRI)

  — Clinical data (NIDDK)

  — Non-human basic research (NIDDK)

DATA MANAGEMENT AND SHARING PLAN
An example from an application proposing to collect single cell genomic data from mice and humans.
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 (but this has not been done in the sample below). 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.

As detailed in the Research Strategy Section, we propose the generation of a spatially mapped single-cell atlas of the developing mouse brain and include specific deliverables. Our primary deliverable for each modality will be a matrix of cells × (counts in peaks for ATAC, UMIs in genes for RNA, or methylation status for DNAm) along with a dense metadata table with information for each cell. This includes the animal sex, developmental time point, punch of origin with x,y,z coordinates, assigned cluster and inferred cell type, assigned subcluster and inferred cell type, as well as a number of QC metrics (total reads, passing reads, reads in peaks, TSS enrichment, cell barcode combination, date of preparation for each stage, sequencing platform, likelihood of being a doublet, and any other relevant metrics that arise during the project).

The amount and type of data from human cells will depend on the results from the mouse studies. Data sharing plans will be updated when appropriate (likely at the start of year 4 of the grant award).

See **Writing a Data Management & Sharing Plan** for details

How to Get Started on a DMS Plan

Steps for preparing for the DMS Policy:

- **Identify existing resources within your institution** that may be able to assist you, such as data librarians

- Review examples of relevant sample plans on sharing.nih.gov

- Try **drafting a Data Management and Sharing Plan** for your work based on the recommended elements (NOT-OD-21-014)

- **Review your past data sharing practices** to meet other funder or publisher expectations and consider what you may need to update for the new DMS Policy

- **Read FOA** for any ICO or program-specific DMS requirements

Ask a librarian!

- **Data librarians** have expertise in data management and sharing, metadata, and more
- **Liaison librarians** may have familiarity with repositories and standards commonly used within a discipline
- **Scholarly communications librarians** may be able to provide guidance on selecting a license for sharing data

Network of the National Library of Medicine National Center for Data Services

NCDS provides training designed to teach librarians to:

- provide expertise in policies and practices supporting **open science and data sharing**
- use NIH & NLM resources and common tools for **data discovery and (re)use**
- **analyze, evaluate, communicate, and visualize** data
- **recommend or apply ethical practices** in data science

NCDS Resources

DATA DRIVEN DISCOVERY:
Building Capacity for Data-Driven Research

NNLM supports the goals of the _NLM Strategic Plan 2017-2027: A Platform for Biomedical Discovery and Data-Powered Health_. This effort is led by the National Center for Data Services. Two of the key goals of this plan are to _accelerate discovery and advance health through data-driven research_ and to _build a workforce for data-driven research and health_. Through funding, education, and outreach, the Network serves as a resource for librarians, students, and interested individuals to learn about library roles in data science and data management and emerging trends in supporting biomedical research.

NIH DMSP Toolkit
Learn about the NIH Data Management and Sharing Policy, including information about data management plans, repositories, and data sharing.

Data Services
Discover resources to help you get started in or level up providing data services at your institution.

NNLM Data Glossary
Connects and defines concepts, services, and tools relevant to librarians working in data-driven discovery.

NLM Data Discovery
Access National Library of Medicine datasets.

https://www.nnlm.gov/ncds/resources

Data Curation Network

Aim: to build a trusted community-led network of curators advancing open research by making data more ethical, reusable, and understandable
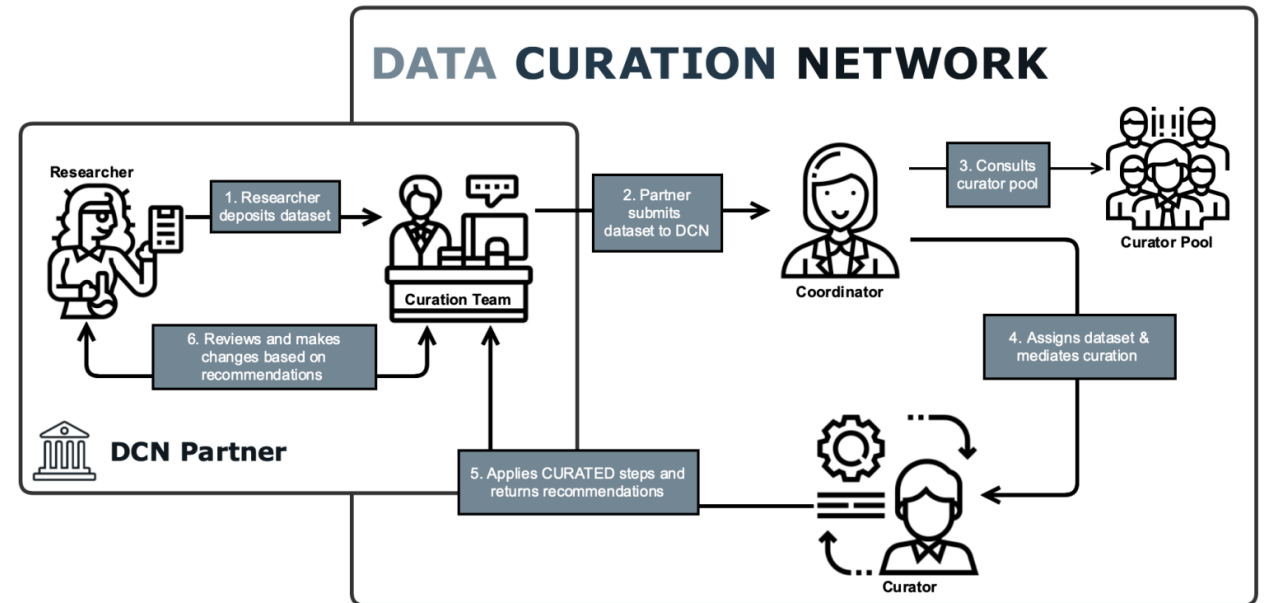
DCN CURATE(D) model
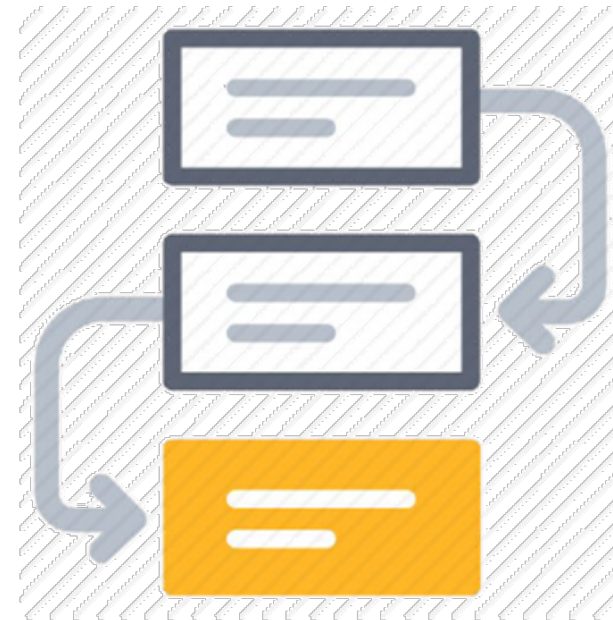
Check files
Understand documentation
Request missing information
Augment the submission
Transform the format
Evaluate for FAIRness
Document throughout

https://datacurationnetwork.org/resources/workflows/

Roadmap for 2023 and Beyond

– **OSP *Under the Poliscope* and *Open Mike* blogs** provide a general roadmap for what to expect

– **Out now!**
- NIH 2-part webinar series & FAQs
- Supplemental information for researchers working with AI/AN Participants
- Supplemental information for protecting privacy when sharing research data
- Notice for Genomic Data Sharing Plan harmonization

– **Ongoing in 2023 and beyond:**
- Additional FAQs and guidance
- Ongoing assessment of the Policy for short- and long-term goals
- Incentives for data sharing

OPEN MIKE

Helping connect you with the NIH perspective, and helping

UNDER THE POLISCOPE
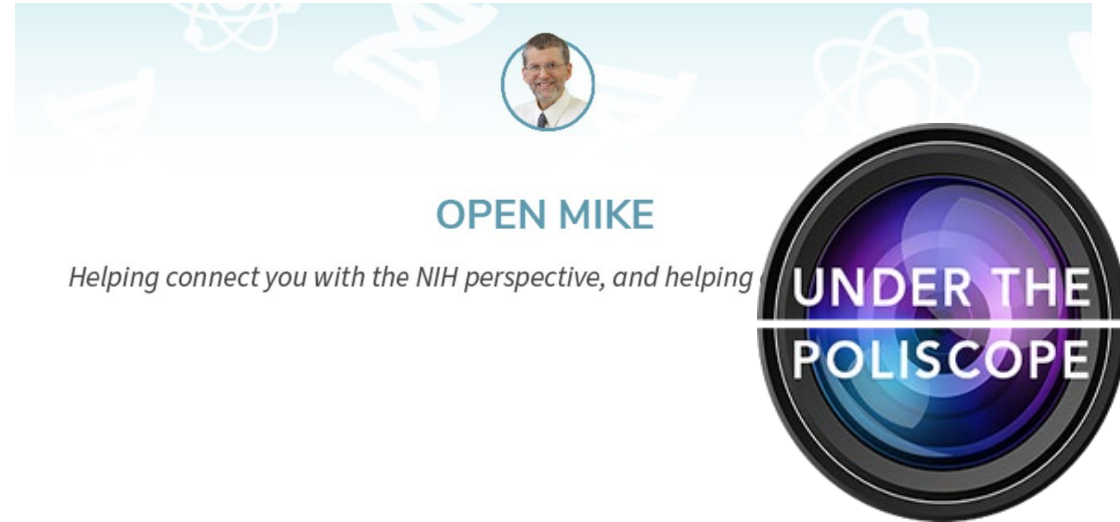
White House
Office of Science and Technology Policy
2022 Public Access Memo

- Directs Federal agencies supporting research to develop plans to ensure:

  - Publications resulting from federally funded research are made freely available and publicly accessible in repositories without embargo

  - Scientific data underlying publications are made accessible at time of publication, and develop approaches for sharing scientific data not underlying publications

  - Collection and sharing of appropriate metadata for publications and data, and digital persistent identifiers for publications, data, researchers, and awards/projects

- NIH will seek public input on implementation plans

[OSTP Memo](#)

Thank You!

Policy and Supplemental Information:

- **NOT-OD-21-013** – Final NIH Policy for Data Management and Sharing

- **NOT-OD-21-014** – Supplemental Information to the NIH Policy for Data Management and Sharing: Elements of an NIH Data Management and Sharing Plan

- **NOT-OD-21-015** – Supplemental Information to the NIH Policy for Data Management and Sharing: Allowable Costs for Data Management and Sharing

- **NOT-OD-21-016** – Supplemental Information to the NIH Policy for Data Management and Sharing: Selecting a Repository for Data Resulting from NIH-Supported Research

Resources:

- **NIH Data Sharing Website** – sharing.nih.gov

- **NIH Office of Science Policy DMS Policy Website** – history and background on the NIH DMS Policy

- **Frequently Asked Questions** – sharing.nih.gov/faqs

- **NIH Data Management and Sharing Policy Webinar Series** – Implementation of the NIH DMS Policy

- **News & Events** – Latest news and upcoming events

Contact:

- Questions – sharing@.nih.gov

- Follow us on Twitter – @NIH_OSP

- osp.od.nih.gov/blog/