NIH Data Management and Sharing (DMS) Policy: Generalist Repositories in the DMS Landscape

January 25, 2023
Cindy Danielson, NIH Office of Extramural Research

To: GREI Workshop

National Institutes of Health
Office of Extramural Research

- 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

National Institutes of Health
Office of Extramural Research

"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
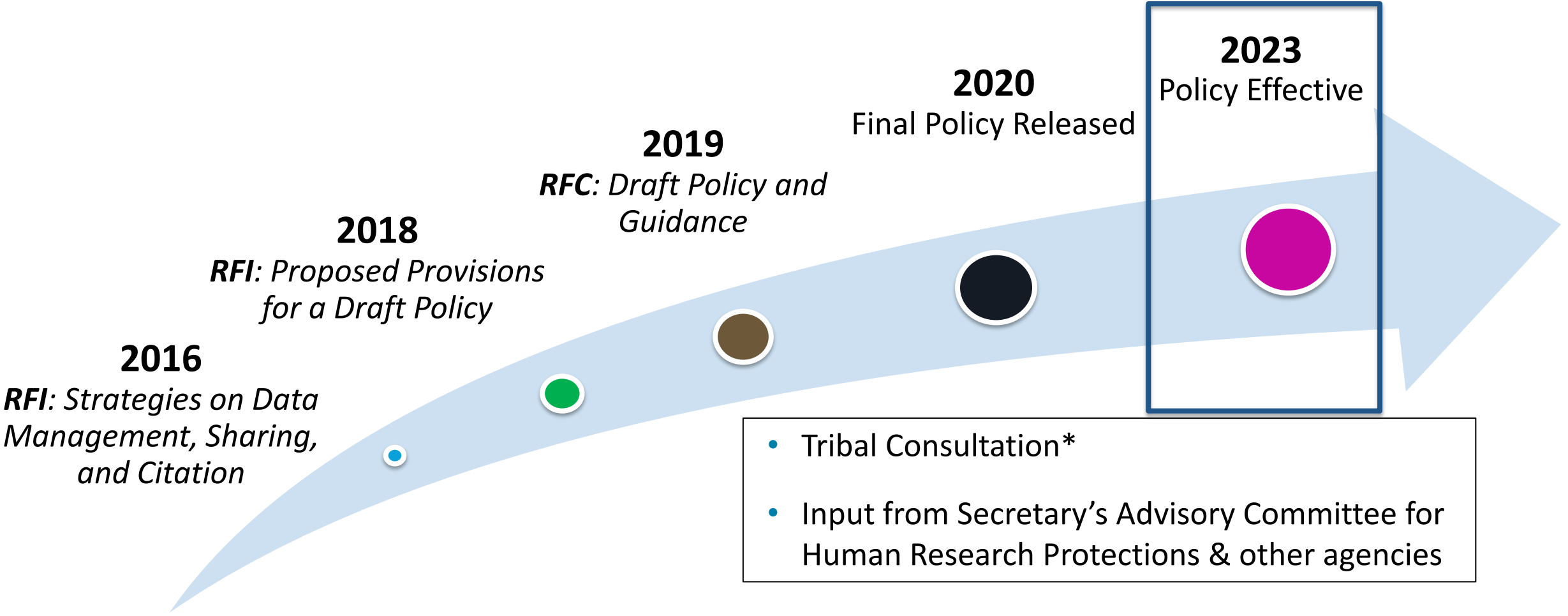
"Reproducibility in Cancer Biology: Challenges for assessing replicability in preclinical cancer biology," Errington et al., (2021)

– 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

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"

National Institutes of Health
Office of Extramural Research

- **Requirements:**
  1. **Submission** of Data Management & Sharing Plan
  2. **Compliance** with ICO-approved Plan (may affect future funding)

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

  – **What's in:** "Recorded factual material… of <u>sufficient quality to validate and replicate</u> <u>research findings</u>, regardless of whether the data are used to support scholarly publications"
  – **What's out:** Lab notebooks, preliminary analyses, peer reviews, physical objects

- **Timelines:**
  – **When to share data:** No later than <u>publication</u> or <u>end of award</u> (for unpublished data), whichever comes first

'Which brings us to my next point.'

National Institutes of Health
Office of Extramural Research

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**

National Institutes of Health
Office of Extramural Research

- Data type
  - Identifying data to be preserved and shared
- 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
- Oversight of data management
  - Plan compliance will be monitored/ managed and by whom

See Writing a Data Management & Sharing Plan for details

National Institutes of Health
Office of Extramural Research

Format of a Data Management and Sharing Plan

- ✓ Recommended to be no more than 2 pages

- ✓ Single Plan for applications subject to both DMS and GDS Policies

- ✓ Optional Format Page available and incorporated into FORMS-H application instructions

- ✓ FDP 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.

DMS Plan format page added to list of Format Pages and incorporated into FORMS-H application instructions

National Institutes of Health
Office of Extramural Research

- Several sample NIH DMS Plans available for educational purposes, including:
  - Human clinical and/or MRI data (NIMH)
  - Human genomic data (NIMH & NHGRI)
  - Human & non-human genomic data (NIMH)
  - Secondary data analysis (NIMH)
  - Human clinical and genomics data (NICHD)
  - Human survey data (NICHD)
  - Model organism (Zebrafish) data (NICHD)
  - Technology development (NHGRI)

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

National Institutes of Health
*Office of Extramural Research*

- Expectations for writing a Data Management & Sharing Plan

- Includes instructions and optional format for submitting Plans

- Addresses Genomic Data Sharing Policy expectations

- Recent FAQs that address 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.
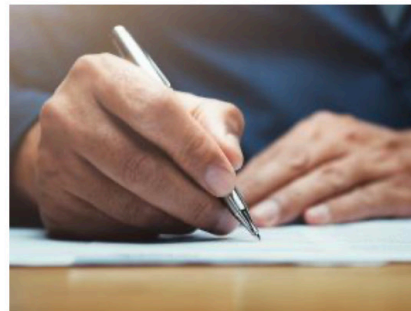
Writing a Data Management & Sharing Plan

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.

National Institutes of Health
Office of Extramural Research

- Encourages use of established repositories
  - Generally improves FAIRness of the data
- Helps investigators identify appropriate data repositories
  - Desirable characteristics, 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

National Institutes of Health
Office of Extramural Research

- Prioritizes data-type and discipline-specific data repositories

- Refers to NIH-supported data repository list outlining:

  – Repository description (e.g., data-types accepted, research community served, tools available),

  – Supportive NIH IC(s),

  – Whether and when new data are accepted, and

  – How to submit data

- Examples include:

  dbGaP   BioData Catalyst

  GenBank   ImmPort

  NIMH Data Archive   BioLINCC

- ## If no appropriate discipline or data-type specific repository is available, consider other potentially suitable options:

  - Institutional repositories

  - PubMed Central (small datasets only)

  - Generalist data repositories, including:

Dataverse	Open Science Framework
Dryad	Synapse
Figshare	Vivli
IEEE Dataport	Zenodo
Mendeley Data	

National Institutes of Health
Office of Extramural Research

NIH-Supported Repositories

- Filterable list of 80+ NIH Repositories

NIH-supported Scientific Data Repositories*

Institute or Center ⬍	Repository Name ⬍	Repository Description	⬍
NCI ⌄		Keyword Filter	
NCI	Cancer Nanotechnology Laboratory (caNanoLab)	caNanoLab is a data sharing portal designed to facilitate information sharing in the biomedical nanotechnology research community to expedite and validate the use of nanotechnology in biomedicine. caNanoLab provides support for the annotation of nanomaterials with characterizations resulting from physico-chemical, in vitro, and in vivo assays and the sharing of these characterizations and associated nanotechnology protocols in a secure fashion.	
NCI	Imaging Data Commons	The National Cancer Institute (NCI) Cancer Research Data Commons (CRDC) aims to establish a national cloud-based data science infrastructure. Imaging Data Commons (IDC) is a new data repository of CRDC supported by the Cancer Moonshot. The goal of IDC is to enable a broad spectrum of cancer researchers, with and without imaging expertise, to easily access and explore the value of de-identified imaging	

Other Repository Resources

- Generalist repositories
- Nature's Data Repository Guidance
- Registry of Research Data Repositories

See **Repositories for Sharing Scientific Data** for details

Decision Tool

Which Policies Apply to My Research?

NIH has a variety of sharing policies in place for research that it funds. This tool will assist in helping you determine which of the following NIH policies apply to a particular project:

- Genomic Data Sharing Policy
- 2003 Data Sharing Policy
- Data Management & Sharing Policy (in effect January 25, 2023)
- Model Organism Sharing Policy
- Research Tools Policy

START HERE

NIH Institute and Center Data Sharing Policies

Institute or Center	Data Sharing Policy Name	Description of Data Sharing Policy
NCI ⌄		
NCI	NCI Clinical Trial Access Policy	NCI believes that the full value of NCI-suppo Clinical Trials can be realized only if the resul published as rapidly as possible. The Clinical at ensuring public availability of results from trials from all NCI-funded research grants, co and/or contracts that support covered interve

See Other Sharing Policies for details

National Institutes of Health
Office of Extramural Research

15

Steps you can take to begin preparing for the DMS Policy:

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

National Institutes of Health
Office of Extramural Research

For More Information…

Website: NIH Scientific Data Sharing

FAQs: DMS Policy FAQs

Email Box: Sharing@nih.gov

Webinar Series: NIH DMS Policy Implementation