



National Institutes of Health
Office of Science Policy

The NIH Data Management and Sharing Policy: Overview and Implementation Update

Presentation to Pragmatic Trials Collaboratory Steering Committee Meeting
May 17, 2023

Taunton Paine, MA

Director, Scientific Data Sharing Policy Division

Office of Science Policy

National Institutes of Health

taunton.paine@nih.gov

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

Major NIH-wide Data Sharing Policies

Policy	Expectations	Year
NIH Data Sharing Policy	Expects investigators seeking more than \$500K in direct support in any given year to submit a data sharing plan with their application or to indicate why data sharing is not possible.	2003
Genomic Data Sharing Policy	Expects sharing of large-scale human and non-human genomic data from NIH-funded studies through a publicly available data repository. All studies with human genomic data should be registered in dbGaP , and the data should be submitted to an NIH-designated data repository . Non-human data may be submitted to any widely used data repository.	2014
Dissemination of NIH-Funded Clinical Trial Information	Expects all investigators conducting NIH-funded clinical trials to register trials at ClinicalTrials.gov, and submit results information. Complementary to Part 11 regulations.	2016

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)

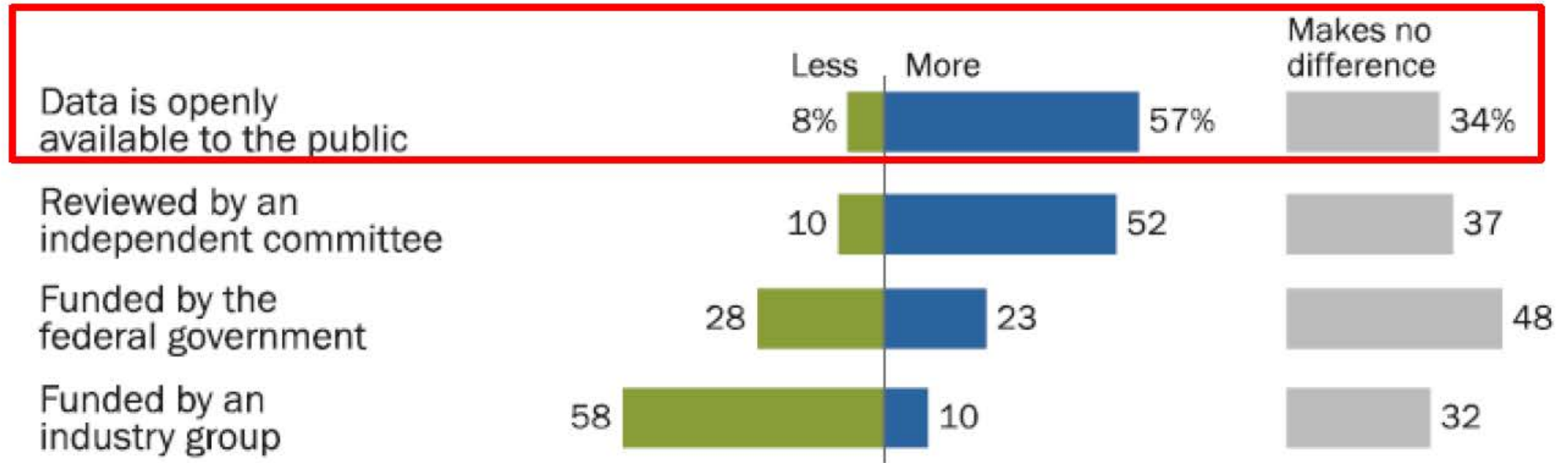
- Attempted to **repeat 193 experiments from 53 high-impact cancer biology papers**; unable to obtain data for **68% of experiments**

“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**

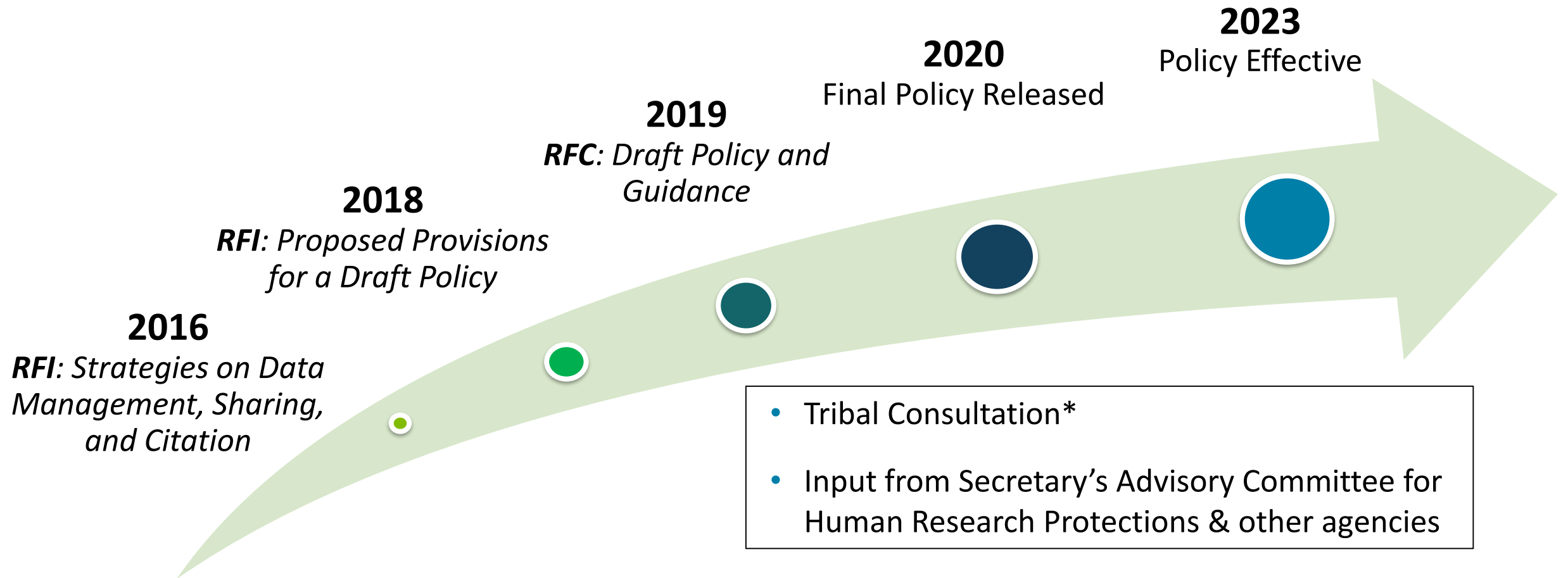
A Matter of Trust

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



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



*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 not 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 sufficient quality to validate and replicate research findings, 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 publication or end of award (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)

Potential Limitations on Sharing

- **Data Management and Sharing Plans should maximize appropriate 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 not 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

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](https://www.nih.gov/sharing). 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

Elements of the Optional DMS Plan Format Page

Element 1: Data Type

- A. Types/amount of scientific data to be generated
- B. Scientific data to be preserved and shared, and the rationale for doing so
- C. Metadata, other relevant data, and documentation

Element 2: Related Tools, Software and/or Code

Element 3: Data Standards

Element 4: Data Preservation, Access, and Associated Timelines

- A. Repository where scientific data/ metadata archived
- B. How scientific data will be findable and identifiable
- C. When and how long scientific data will be available

Element 5: Access, Distribution, or Reuse Considerations

- A. Factors affecting subsequent access, distribution, reuse
- B. Whether access to scientific data will be controlled
- C. Protections for privacy, rights, and confidentiality of human research participants

Element 6: Oversight of Data Management and Sharing

See [Writing a Data Management & Sharing Plan](#) for details and [Format Page](#)

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

Supplemental Information: **Repository Selection**

Specialized Data Repositories

- 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
 - GenBank
 - NIMH Data Archive
 - BioData Catalyst
 - ImmPort
 - BioLINCC

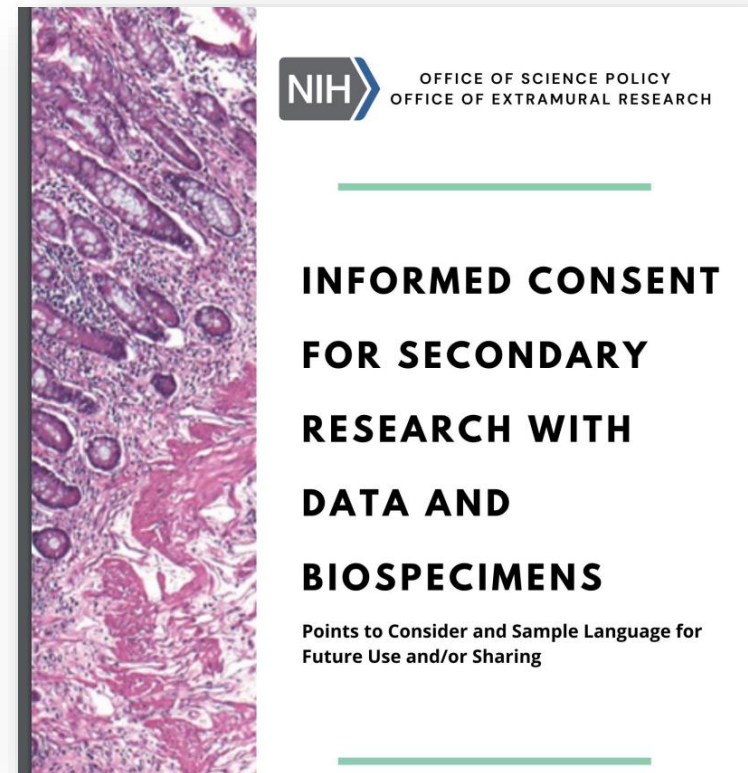
Supplemental Information: **Repository Selection**

Other Established Data Repositories

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

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



See [Informed Consent Resource](#) for details

Supplemental Information: 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**

See [Principles and Best Practices](#) for details



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
- **NOT 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*

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

The screenshot shows the homepage of sharing.nih.gov. At the top, there is a yellow navigation bar with the text "U.S. Department of Health & Human Services | National Institutes of Health". Below this is the NIH logo and the text "SCIENTIFIC DATA SHARING". To the right of the logo is a search bar with the word "Search" and a magnifying glass icon. Further right are links for "NIH Staff", "FAQ", and "Contacts & Help". Below the navigation bar is a horizontal menu with five items: "DATA MANAGEMENT AND SHARING POLICY", "GENOMIC DATA SHARING POLICY", "OTHER SHARING POLICIES", "ACCESSING DATA", and "ABOUT". The main content area features a large blue background with a network diagram of nodes and lines. The headline reads "Expediting the Translation of Research Results to Improve Human Health." Below the headline is a section titled "FEATURED NEWS & EVENTS" with a sub-headline "Gearing Up for 2023: Implementing the NIH Data Management and Sharing Policy" and a "View More" button.

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](https://twitter.com/NIH_OSP)
- osp.od.nih.gov/blog/

