National Institutes of Health

The NIH Data Management and Sharing Policy: Overview, Implementation, and Resources

American Medical Informatics Association Webinar
December 6, 2022

Taunton Paine, MA
Director, Scientific Data Sharing Policy Division
Office of Science Policy
National Institutes of Health

Teresa Zayas Cabán, PhD
Assistant Director for Policy Development
National Library of Medicine
National Institutes of Health

sciencepolicy@mail.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)

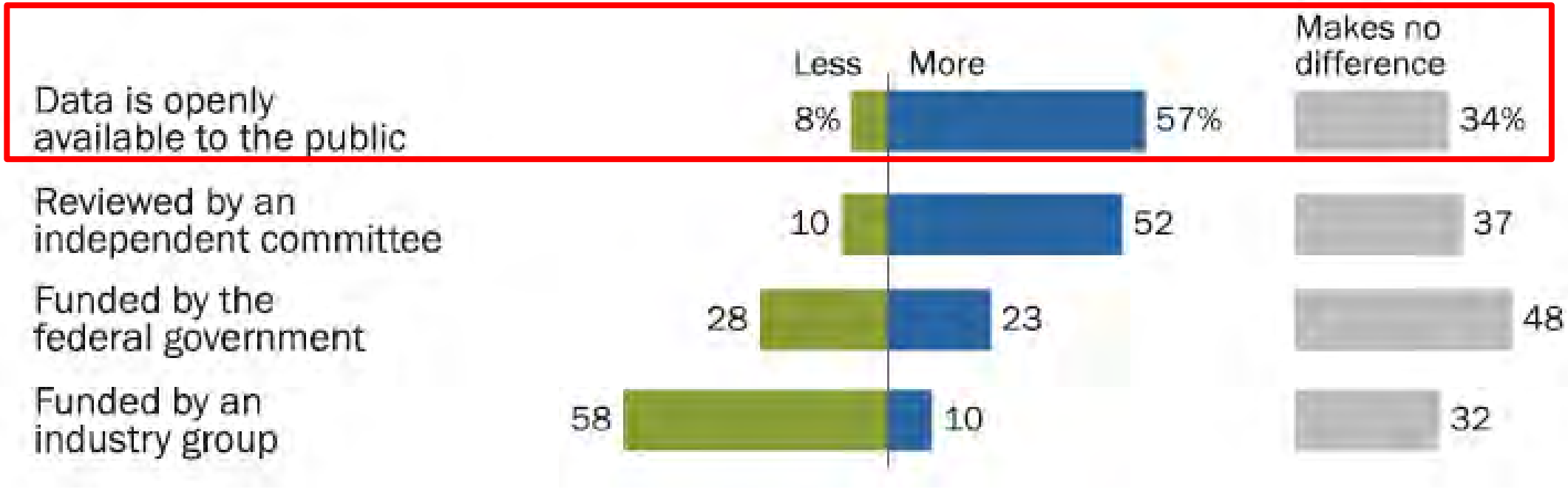
- Attempted to **repeat 193 experiments from 53 high-impact cancer biology papers;** unable to obtain data for **68% of experiments**

"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

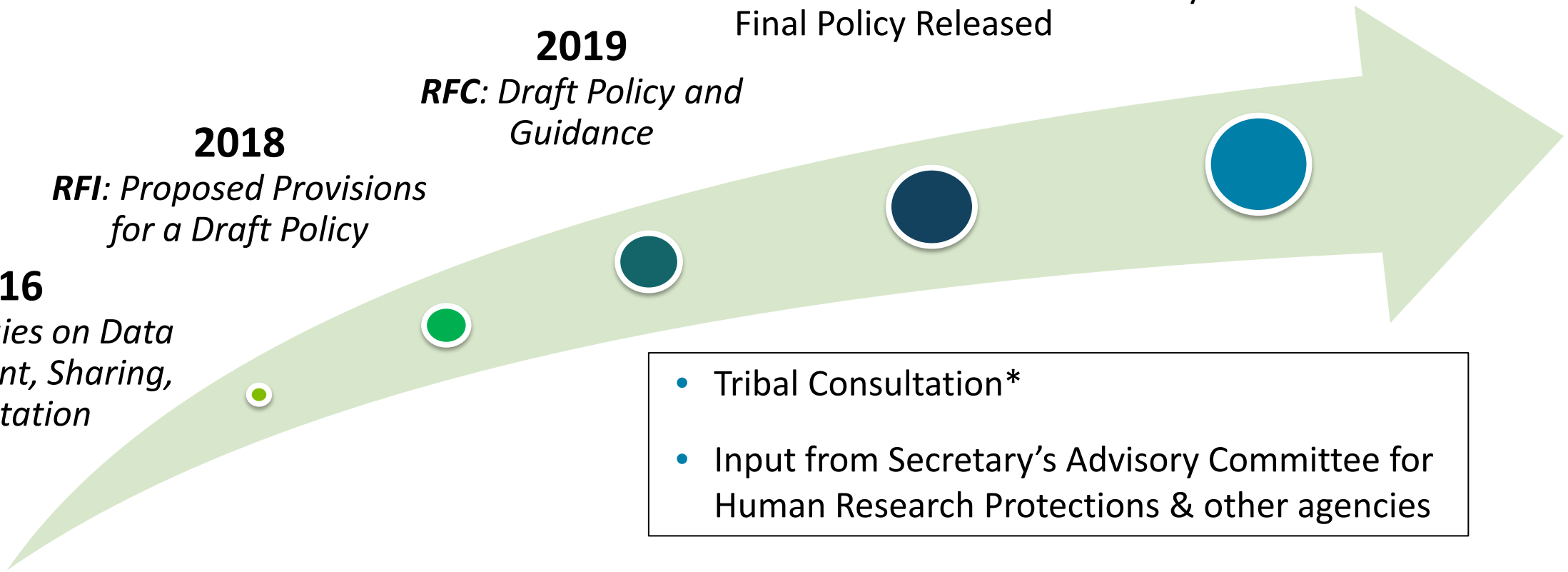
2016
***RFI**: Strategies on Data Management, Sharing, and Citation*

2018
***RFI**: Proposed Provisions for a Draft Policy*

2019
***RFC**: Draft Policy and Guidance*

2020
Final Policy Released

2023
Policy Effective



- 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
  - 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)

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
  - **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)

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

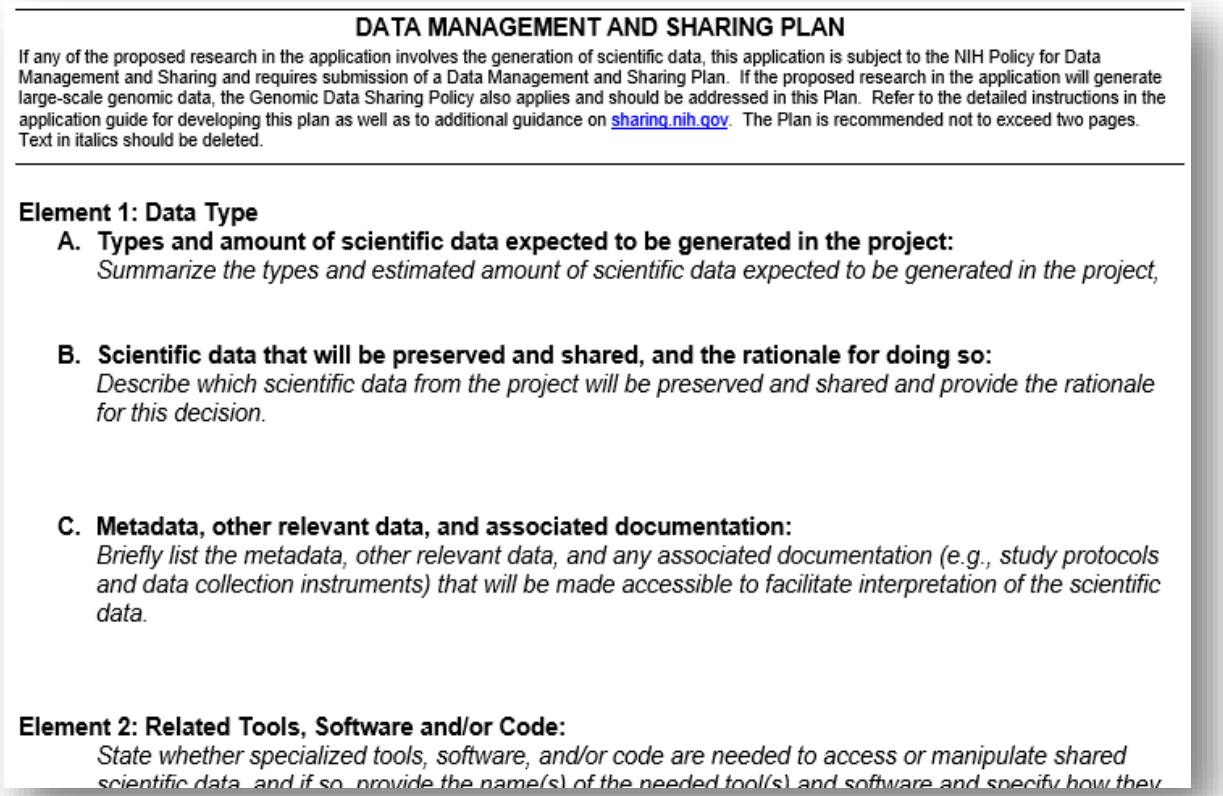
Supplemental Information: Elements of a Data Management and Sharing Plan

- **Data type**
  - 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

Format of a Data Management and Sharing Plan

- ✓ Plans recommended to be no more than 2 pages in length
- ✓ Optional format page will be available
- ✓ 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.

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

DMS Plan format page will be added to list of Format Pages and incorporated into FORMS-H application instructions by Fall 2022

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
  - ***Generalist Data Repository Ecosystem Initiative Webinar: December 8***

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

Best Practices for Responsible Management and Sharing of AI/AN Participant Data

Understand	Understand Tribal sovereignty and laws, regulations, policies, and preferences
Engage	Engage early with Tribes when developing a data management and sharing plan, before research begins, and continue throughout research
Establish	Establish mutually beneficial partnerships
Agree	Agree who will manage data (e.g., Tribe, researcher, trusted 3rd party)
Consider	Consider additional protections, as necessary

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**

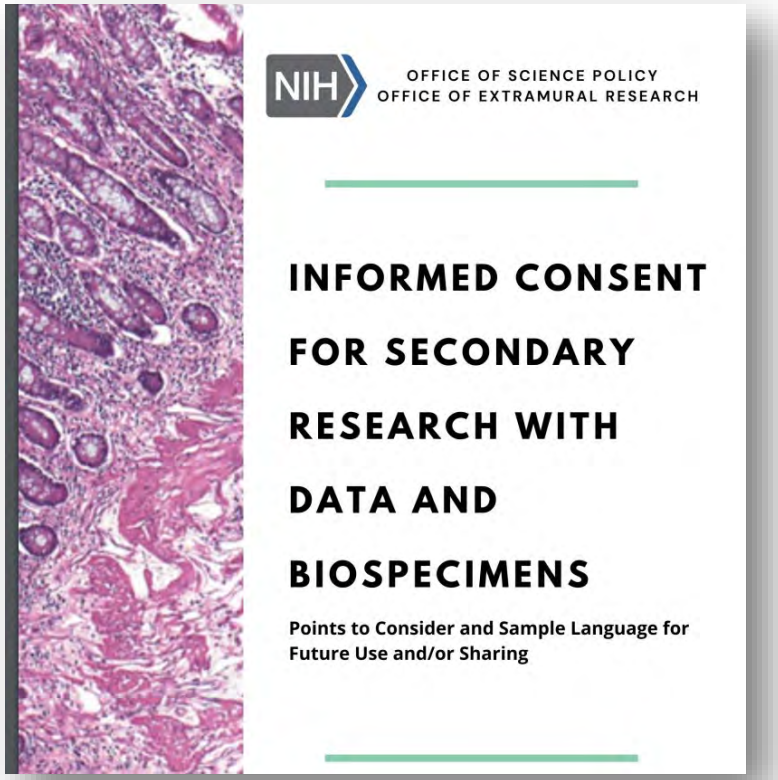
NOT-OD-22-213

Best Practices for Protecting Privacy When Sharing Human Research Participant Data

1. De-identify to the greatest extent while maintaining scientific utility; Use Common Rule and HIPAA Privacy Rule standards
   - Consider risks from information even when de-identified
   - Share identifiable data only with explicit consent
2. Use agreements for transferring data
   - Communicate limitations on use, include prohibitions on re-identification or recontact
3. Understand applicable legal protections and limitations on disclosure

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



Informed Consent Resource

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

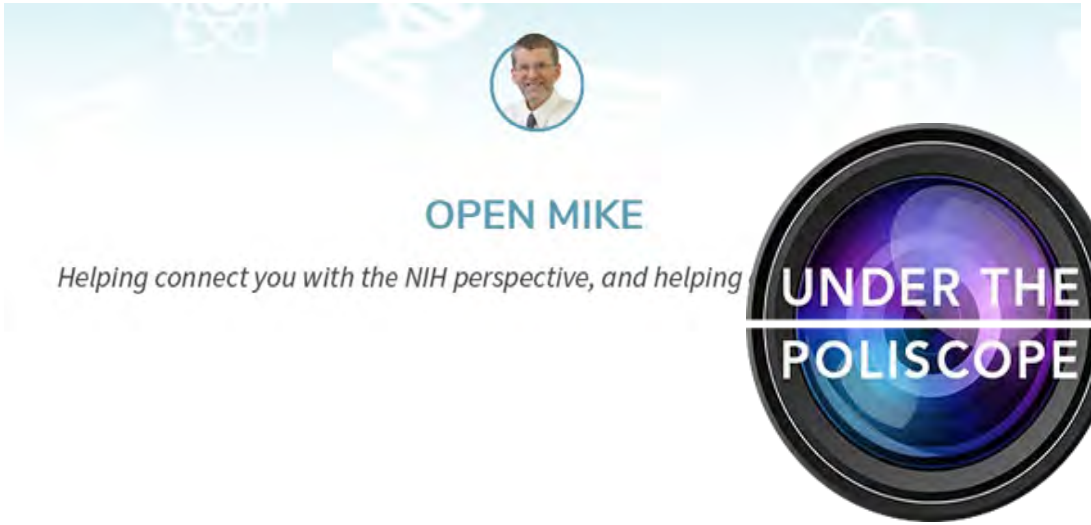
Compliance may factor into future funding decisions

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

Monitoring Compliance with DMS Plans

- Approved DMS Plan becomes a Term and Condition of Award
- Recipient reports progress on implementing approved DMS Plan in Research Performance Progress Report (RPPR)
  - RPPR questions will be updated
- NIH reviews compliance annually
  - Failure to comply may result in an enforcement action, including additional special terms and conditions or termination of award, and may affect future funding decisions
- *Plans may be made publicly available in the future*

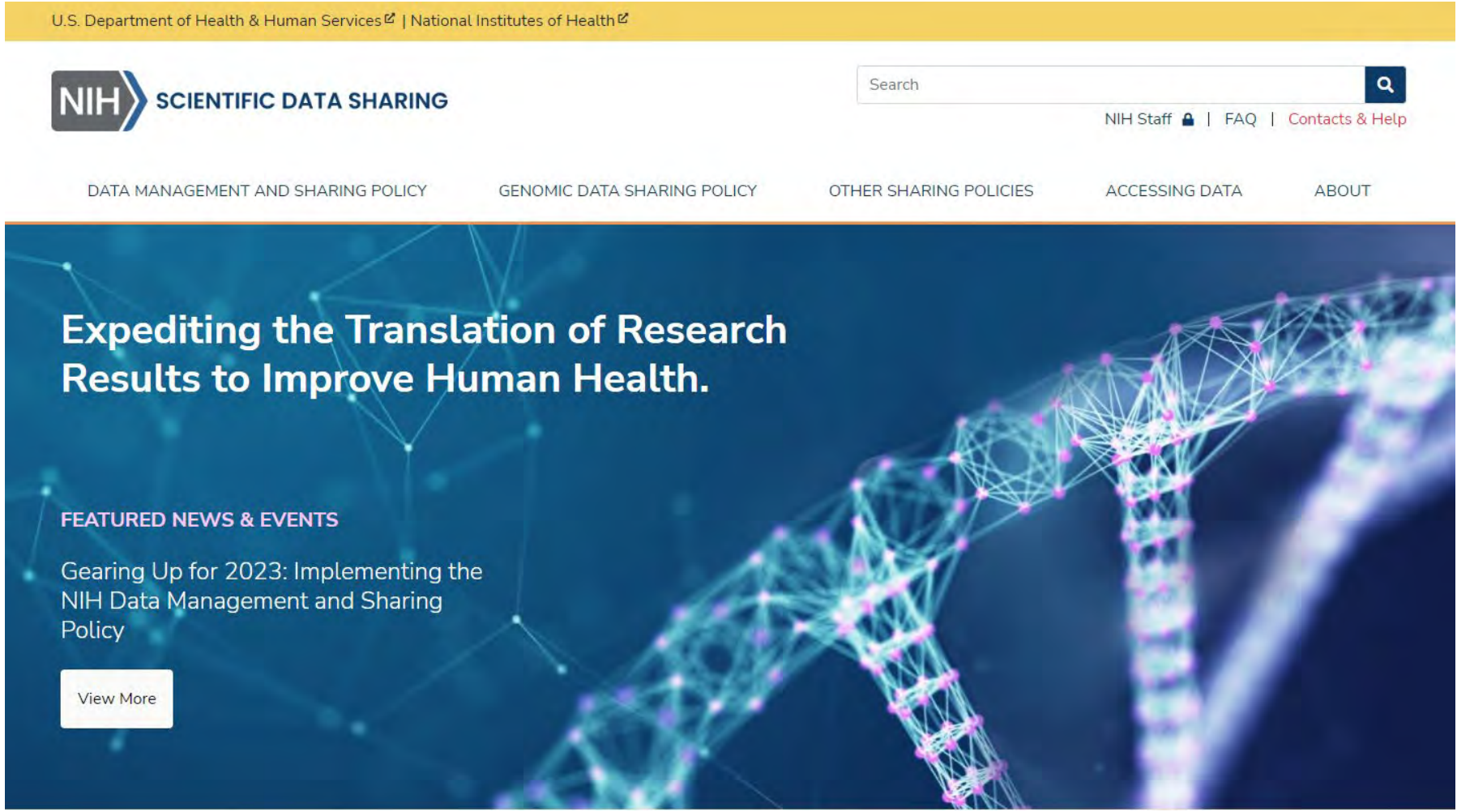
Roadmap to 2023 and Beyond



- Recent **OSP *Under the Poliscope* and *Open Mike* blogs** provide a general roadmap for what to expect leading to 2023 and afterward
- **Out now!**
  - NIH 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
- **Before 2023:**
  - Final Plan format page, additional FAQs and guidance
- **Beyond 2023:**
  - Ongoing assessment of the Policy for short- and long-term goals
  - Incentives for data sharing

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



DMS Policy FAQs

Policy Scope

- How the DMS Policy fits with other NIH data sharing policies
- Interaction of DMS Policy with expectations of other funders, collaborators
- Applicability to projects establishing repositories or creating data infrastructure with no research

Managing and Sharing Scientific Data

- Whether all data are expected to be shared
- When data should be shared
- Whether timeline for sharing data changes with a no cost extension
- Potentially justifiable reasons for limiting sharing of data
- Expectations for SBIR/STTR projects, secondary research, behavioral research, qualitative data

Considerations for Scientific Data Derived from Human Participants

- Protections for human research participants
- Whether broad consent is a requirement of the Policy

Compliance and Enforcement

- How noncompliance is handled

Contract-specific considerations

FAQs

NIH Genomic Data Sharing Policy

- **Purpose**
  - Sets expectations and responsibilities to ensure broad, responsible, and timely sharing of genomic data
  - Expects consent for research use
- **Scope**
  - Applies to all NIH-funded research generating ***large-scale human*** or ***non-human*** genomic data and secondary research using these data
- **Became effective January 25, 2015**

Input Requested on the Future of the GDS Policy

- **Published RFI (NOT-OD-22-029) November 30, 2021**
  - 60 comments received
- **Issued NOT-OD-22-198, August 31, 2022**
  - Harmonizes expectations for GDS Plans with DMS Plans, including format, submission timing, review, and compliance
- **Still considering input received on:**
  - Standards for NIH-supported data repositories
  - Whether to accept other standards for data de-identification, including expert determination, and under what conditions
  - Whether to permit records linkage, and under what conditions
  - Whether to expand the scope of the GDS Policy's sharing and/or protections to include other research scenarios (e.g., projects of smaller size) or data types (e.g., proteomics, metabolomics)

"Refreshing NIH's Genomic Data Sharing Policy," December 1, 2021, blog by Dr. Lyric Jorgenson, NIH

Advancing Open Science and Data Science at NIH and NLM

Data Infrastructure	Modernized Data Ecosystem	Data Management, Analytics, and Tools	Workforce Development	Stewardship and Sustainability
•Optimize data storage and security •Connect NIH data systems	•Modernize data repository ecosystem •Support storage and sharing of individual datasets •Better integrate clinical and observational data into biomedical data science	•Support useful, generalizable, and accessible tools and workflows •Broaden utility of and access to specialized tools •Improve discovery and cataloging resources	•Enhance the NIH data-science workforce •Expand the national research workforce •Engage a broader community	•Develop policies for a FAIR data ecosystem •Enhance stewardship

Accelerating Discovery and Data-Powered Health

Accelerate discovery and advance health through data-driven research

Reach more people in more ways through enhanced dissemination and engagement

Build a workforce for data-driven research and health

Standards and Interoperability in Research Data Sharing

Fast Healthcare Interoperability Resources (FHIR®) Standard

Notice Number: NOT-OD-19-122

Key Dates

Release Date: July 30, 2019

Related Announcements

NOT-HL-20-815
NOT-OD-19-014
NOT-OD-18-134
NOT-OD-19-150
NOT-OD-20-146
NOT-HL-21-010

Issued by

OFFICE OF THE DIRE

Purpose

The purpose of this no
clinical data for resear

Accelerating Clinical Care and Research through the Use of the United States Core Data for Interoperability (USCDI)

Notice Number:
NOT-OD-20-146

Key Dates

Release Date: July 30, 2020

Related Announcements

NOT-OD-19-122 - Fast Healthcare Interoperability Resources (FHIR®) Standard

NOT-OD-19-150 - Request for Information (RFI): Use of the Health Level Seven International (HL7®) Fast Healthcare Interoperability Resources (FHIR®) for Capturing and Sharing Clinical Da
for Research Purposes

NOT-OD-19-014 - Request for Information (RFI) on Proposed Provisions for a Draft Data Management and Sharing Policy for NIH Funded or Supported Research

NOT-OD-18-134 - Request for Information (RFI): Soliciting Input for the National Institutes of Health (NIH) Strategic Plan for Data Science

Issued by

Office of The Director, National Institutes of Health (OD)

Purpose

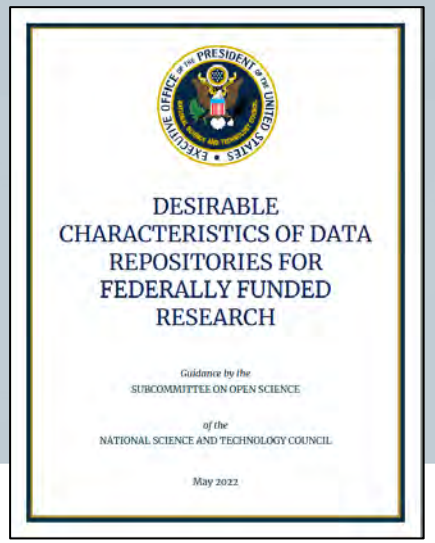
Purpose

The purpose of this notice is to encourage NIH-supported clinical research programs and researchers to adopt and use the standardized set of data classes, data elements, and associated vocabulary standards specified in the United States Core Data for Interoperability (USCDI) standard.[1] The use of USCDI will facilitate the use of clinical data in research studies and enable

PubMed Central

db GaP
GENOTYPES and PHENOTYPES

ClinicalTrials.gov
A service of the U.S. National Institutes of Health



Critical Role of Informatics

Expediting Access to Results of Federally Funded Research

- Policy guidance for Federal agencies supporting research to develop or update plans to ensure:
  - Publications are made freely available and publicly accessible in repositories without embargo
  - Scientific data underlying publications are made accessible at the time of publication
  - Digital persistent identifiers are included in published research outputs

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 – sciencepolicy@mail.nih.gov
- Follow us on Twitter – @NIH_OSP
- osp.od.nih.gov/blog/