National Institutes of Health
Office of Science Policy

Update on the NIH Policy for Data Management & Sharing and Implementation Activities

Presentation to Federal Demonstration Project
September 22, 2021

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

Cindy Danielson, PhD
Health Science Policy Analyst
Office of Extramural Research
National Institutes of Health

sciencepolicy@mail.nih.gov

Overview

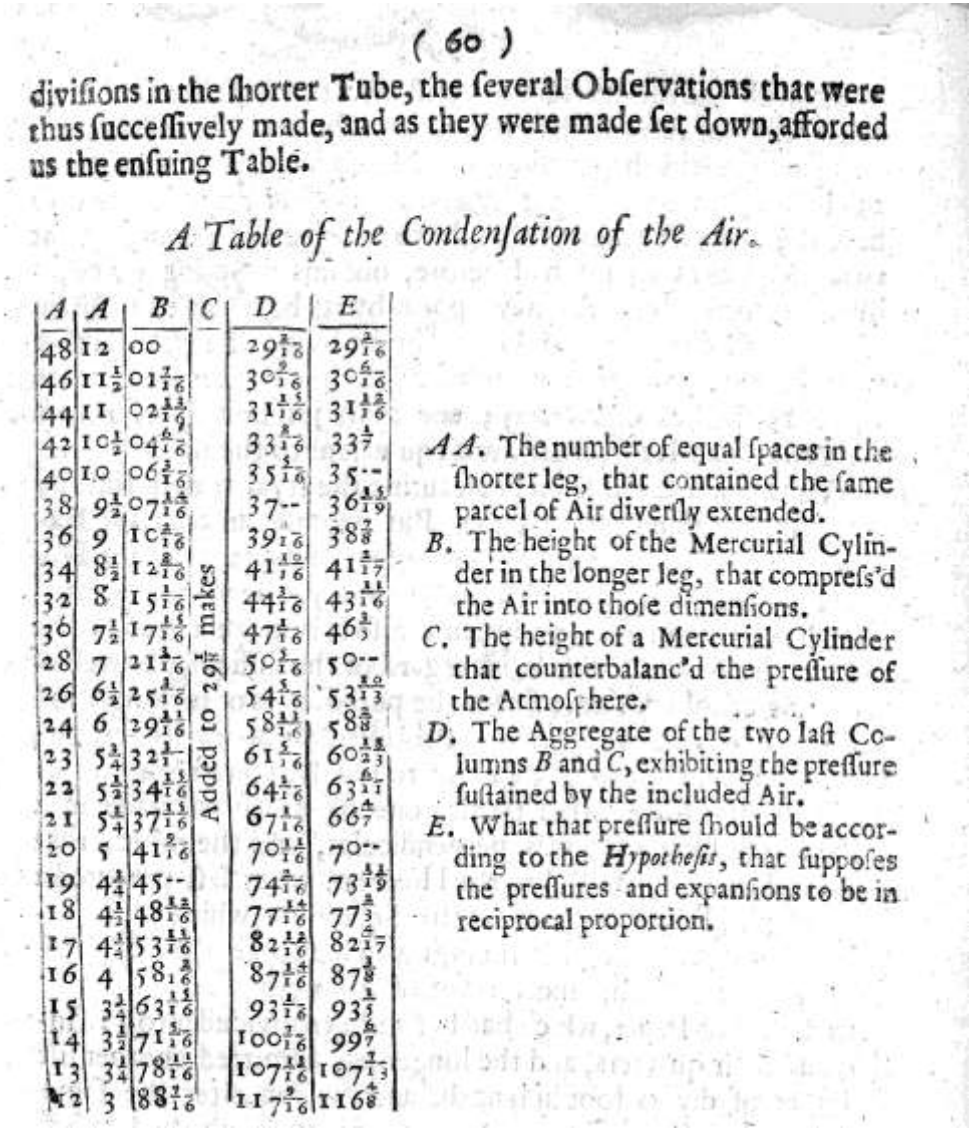
- Evolution of Scientific Data Sharing Efforts & Development of the NIH Data Management and Sharing Policy
- Overview of the NIH Data Management and Sharing Policy and Supplemental Information
- Implementation Activities and Next Steps
- Outreach and Discussion

Data Stewardship Goals

- **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

Sharing Scientific Data is Not New



(60)

divisions in the ſhorter Tube, the ſeveral Obſervations that were thus ſucceſſively made, and as they were made ſet down, afforded us the enſuing Table.

A Table of the Condenſation of the Air.

A	A	B	C	D	E
48	12	00		29 2/16	29 2/16
46	11 1/2	01 7/16		30 9/16	30 6/16
44	11	02 12/16		31 15/16	31 12/16
42	10 1/2	04 6/16		33 8/16	33 1/2
40	10	06 3/16		35 5/16	35 --
38	9 1/2	07 14/16		37 --	36 15/19
36	9	10 2/16		39 2/16	38 7/8
34	8 1/2	12 8/16		41 10/16	41 2/17
32	8	15 1/16		44 2/16	43 11/16
30	7 1/2	17 15/16		47 1/16	46 3/5
28	7	21 3/16		50 5/16	50 --
26	6 1/2	25 8/16		54 5/16	53 10/13
24	6	29 13/16		58 13/16	58 8/8
23	5 3/4	32 1/16		61 5/16	60 23/23
22	5 1/2	34 15/16		64 1/16	63 6/11
21	5 1/4	37 15/16		67 1/16	66 4/7
20	5	41 2/16		70 1/16	70 --
19	4 3/4	45 --		74 2/16	73 15/19
18	4 1/2	48 12/16		77 14/16	77 3/3
17	4 1/4	53 13/16		82 12/16	82 6/17
16	4	58 8/16		87 14/16	87 3/8
15	3 3/4	63 15/16		93 1/16	93 1/5
14	3 1/2	71 5/16		100 7/16	99 7/7
13	3 1/4	78 11/16		107 12/16	107 7/13
12	3	88 7/16		117 9/16	116 8/8

Added to 29 makes

AA. The number of equal ſpaces in the ſhorter leg, that contained the ſame parcel of Air diverſly extended.

B. The height of the Mercurial Cylinder in the longer leg, that compreſs'd the Air into thoſe dimenſions.

C. The height of a Mercurial Cylinder that counterbalanc'd the preſſure of the Atmoſphere.

D. The Aggregate of the two laſt Columns *B* and *C*, exhibiting the preſſure ſuſtained by the included Air.

E. What that preſſure ſhould be according to the *Hypotheſis*, that ſuppoſes the preſſures and expanſions to be in reciprocal proportion.

- Data table from Robert Boyle's *New experiments physico-mechanical, touching the air*, 2nd ed., 1662
- Formed basis of "Boyle's Law"

Wellcome Collection. [Oxford] : [H. Hall for T. Robinson], [1662]
https://wellcomecollection.org/works/k4u8keem

1980s: Grantee Data & Federal Records

- **1980 SCOTUS: *Forsham v. Harris*[1]**
  - NIH-funded University Group Diabetes Program found commonly prescribed diabetes drug had 2.5x increased risk of death from heart disease; FDA used the findings to create labeling requirements
  - A group of scientists requested raw data (data forms and computer tapes) through the Freedom of Information Act (FOIA); Supreme Court affirmed grantees' data were not agency records and were not subject to FOIA
- **1988 PHS Policy on Distribution of Research Resources**
  - Expected unique research resources be made readily available for research purposes to the scientific community after publication
- **1989 NHLBI "L'Enfant Memo"[2]**
  - NHLBI Director Claude L'Enfant created policy that grantees and contractors would make data available from clinical trials, epidemiological studies, and other large-scale studies within three years of major publications
  - Intended to maximize the Federal government's investment in research

1- https://www.jameslindlibrary.org/articles/the-trials-and-tribulations-of-the-university- group-diabetes-program-lessons-and-reflections/
2- https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4286227/

1990s: Sharing and Transparency

- **1996 Human Genome Project "Bermuda Principles"**
  - Expected immediate and broad sharing of data
- **1997 FDA Modernization Act and ClinicalTrials.gov**
  - Established ClinicalTrials.gov
- **"Six Cities Study" and 1999 Shelby Amendment**
  - NIH-funded **"Six Cities Study"** found in 1993 that fine particle air pollution (>2.5 microns) reduced lifespans
  - Study findings were used by EPA in 1997 for regulations
  - Findings were challenged by groups who sought access to data, but access was denied on the basis of informed consent and privacy
  - Sen. Shelby created "Shelby Amendment," later interpreted that federally-funded research data underlying regulations must be made available
  - **Beginning of broad consensus that data underlying publicly-funded studies should be made available**

Precedent for Making Data from NIH Supported Research Publicly Available

Example NIH-wide Data Sharing Policies

NIH Data Sharing Policy (2003)

Establishes expectation that research data from large awards (>$500K) will be shared

NIH Genome-Wide Association Studies Policy (2007) & NIH Genomic Data Sharing Policy (2015)

Establishes expectations for sharing large-scale genomic data

NIH Intramural Human Data Sharing Policy (2015)

Establishes expectation that all human data from the intramural research program should be shared consistent with applicable laws, regulations, and policies

NIH Policy on the Dissemination of NIH-Funded Clinical Trial Information (2016)

Establishes expectation for the timely registration and submission of results information for all NIH-funded clinical trials

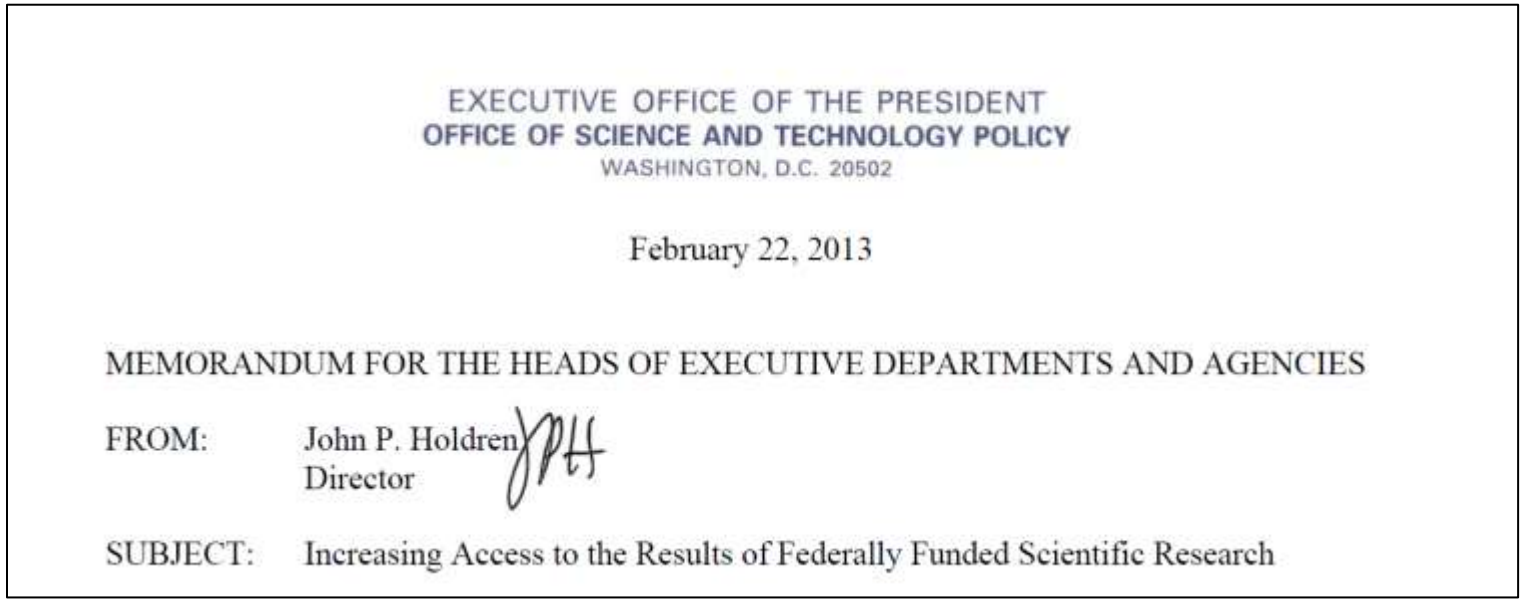
Examples of NIH ICO and Domain Specific Data Sharing Policies*

- NIMH Data Sharing Policy
  - Expects all raw and analyzed data from NIMH-funded human subjects research to be deposited into the NIMH Data Archive
- NIH Data Sharing Policy for Autism Data
  - Expects all raw and analyzed data from human subjects research related to autism to be deposited into the NIMH Data Archive
- NHLBI Clinical Trials and Epidemiological Studies Data Sharing Policy
  - Expects data submission to BioLINCC or another suitable repository no later than 3 years after clinical trial or epidemiological study completion or 2 years after the main paper is published
- NCI Cancer Moonshot Public Access and Data Sharing Policy
  - Expects a Public Access and Data Sharing Plan for making publications resulting from Cancer Moonshot funding and their underlying primary data publicly available immediately to the extent possible

*Non-comprehensive list of NIH data sharing policies

2000s: Towards Sharing by Default

- **2013 "Holdren Memo"**
  - Required by 2010 America COMPETES Reauthorization Act
  - White House Office of Science and Technology Policy directed all Federal Depts. and Agencies with R&D budgets $100 million+ to work toward requiring data management and sharing plans be submitted from all applicants



EXECUTIVE OFFICE OF THE PRESIDENT
OFFICE OF SCIENCE AND TECHNOLOGY POLICY
WASHINGTON, D.C. 20502

February 22, 2013

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

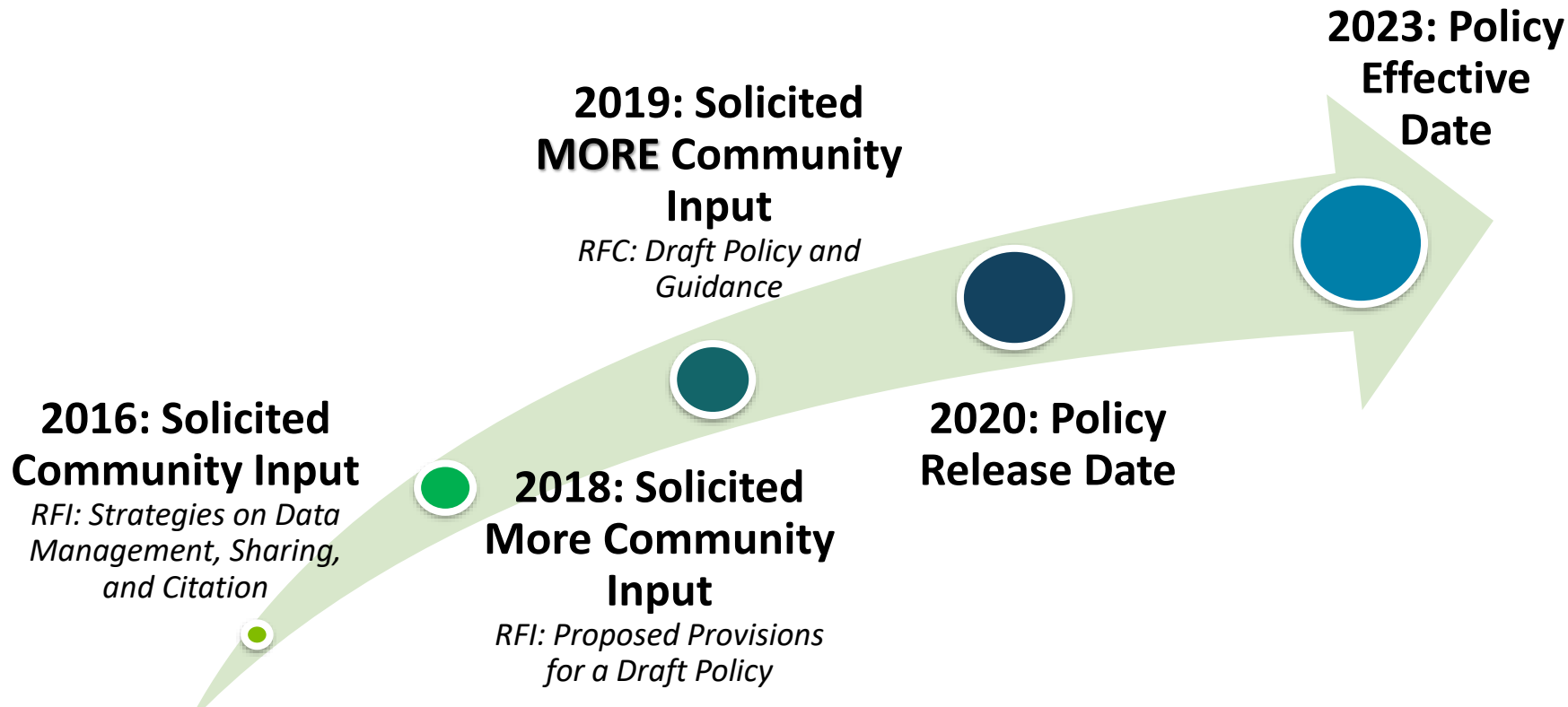
FROM: John P. Holdren
Director

SUBJECT: Increasing Access to the Results of Federally Funded Scientific Research

An Iterative Policy Development Process

- Sought public comment repeatedly
- Tribal Consultation*

  **Details provided in "NIH Tribal Consultation Report: NIH Draft Policy for Data Management and Sharing"*
- Intersection with other government agencies & Secretary's Advisory Committee for Human Research Protections



Overview

- Evolution of Scientific Data Sharing Efforts & Development of the NIH Data Management and Sharing Policy
- Overview of the NIH Data Management and Sharing Policy and Supplemental Information
- Implementation Activities and Next Steps
- Outreach and Discussion

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)*
- **Supplemental info available to assist**
- **Aims to foster data stewardship**

Building on the 2003 Data Sharing Policy

- **Key differences between the 2003 and 2020 Policies:**

Policy Provision	2003 Policy	2020 Policy
Scope	• Awards >$500K a year	• All awards generating scientific data
Data to share	• Final research data	• Scientific data regardless of whether data underlie a publication
Mode of sharing data	• To be described in Plan	• Use of established repository named in Plan
Plan elements	• Flexible level of detail	• Detailed guidance provided in Supplemental Information

- **Graduated compliance date**
  - Replaces the 2003 Data Sharing Policy for new and competing applications submitted for the January 25, 2023 receipt date

The Devil is in the Details…

- **Scope:** All NIH-supported research generating *scientific data*
  - **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
  - **Does not include** lab notebooks, preliminary analyses, peer reviews, physical objects

- **Timelines:**
  - For **when to share data,** no later than publication or end of award (for unpublished data)
  - For **how long to share data,** consider relevant requirements and expectations (e.g., repository policies, retention requirements, journal policies) for minimum time frames

'Which brings us to my next point.'

Additional Expectations for Plans

- **SHARING SHOULD BE …**
  - **The default practice**
    - Maximize appropriate data sharing; plans may justify exceptions (i.e., ethical, legal, technical factors)
    - All scientific data should be managed; not all scientific data must be shared

  - **Responsibly implemented**
    - Plans should outline protection of privacy, rights, and confidentiality
    - Existing laws, regulations, and policies continue to apply
  - **Prospectively planned for**
    - During informed consent, including communicating how data will be used and shared
    - Data submission, including whether access to data, even if de-identified, should be controlled

Supplemental Information: Elements of a Data Management and Sharing Plan

- **Recommended elements of a 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

Supplemental Information: Repository Selection

- **Encourages use of established repositories to improve FAIRness:**
  - Includes broader data repository ecosystem supported by other organizations (public and private)
  - Provides considerations for storing human data
    - e.g., fidelity to consent, restricted use compliant, privacy, plan for breach, download control, violations, and request review
- **Helps investigators identify appropriate data repositories**
  - Provides desirable characteristics
    - e.g., use of persistent unique identifiers, attached metadata, facilities quality assurance
  - Refers to list of NIH-supported Data Repositories
- **NIH ICs may designate specific data repository(ies)**

Supplemental Info to the Policy: Allowable Costs

- **Reasonable costs allowed in budget requests**
  - 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)

Plan Submission and Review

Extramural Grant Awards*

Plan Submission

With application for funding in Budget Justification section

Plan Assessment

Peer reviewers only comment on (not score) budget

NIH program staff assess Plans

Plans can be updated

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, OTAs, IRP*

Overview

- Evolution of Scientific Data Sharing Efforts & Development of the NIH Data Management and Sharing Policy
- Overview of the NIH Data Management and Sharing Policy and Supplemental Information
- Implementation Activities and Next Steps
- Outreach and Discussion

Implementation Areas of Consideration

What's Next?

- **Collaborative trans-NIH effort** – Working with trans-NIH stakeholders and committees to implement the policy
- **Approaches and workflows** – Determine the appropriate roles, responsibilities, and processes by which ICs will assess Plans and monitor compliance
- **System changes** – Enhance award management systems and develop tools to support the submission, assessment, and compliance monitoring of Plans
- **Public posting of Plans** including how they will link to repositories, employment of persistent identifiers such as DOI, and FAIR principles
- **Planning communications and guidance** to ensure investigators, institutions, and NIH staff are prepared for the Policy

Example Questions Received on Implementation

- How will information in a Data Management & Sharing Plan be captured?
  - Will there be a template to submit Plans?
  - Will the Plan be submitted as a free text document, or will some elements be collected in a structured form to enhance consistency and facilitate assessment?
  - What information will peer reviewers need to determine whether the budget is reasonable?
- How to promote consistency, minimize redundant Plans, and ensure accountability?
  - What types of applications will need to submit a Plan, and how will exceptions be handled?
  - Will awards subject to multiple data sharing Policies need to submit multiple Plans?
  - How will Plan submission and compliance for multi-site network studies be handled?
- What guidance will be offered to help applicants prepare Plans?
  - What criteria will be used to assess the appropriateness of a Plan?
- How will NIH monitor policy compliance and handle non-compliance?
  - How will institutional non-compliance be handled during the award and after?

Are there other questions or points we should consider?

Resources for Implementation

What's Next?

- **Engage in outreach** to develop additional supplemental information (including tribal-specific considerations)
- **Develop resources** to inform data management and sharing costs (informed by the 2020 NASEM report on forecasting costs & April 2021 NASEM workshop)
- **Develop approaches** for incentivizing good data sharing practices
- **Clarify interactions** with other NIH-wide (e.g., GDS Policy) and ICO-specific data sharing policies
- **Develop FAQs and other resources** to aid policy implementation

Implementing Recommendations from the NIH Tribal Consultation

- **The Policy clarifies NIH's respect for Tribal sovereignty:**
  - Data should be managed and shared in accordance with Tribal laws and sovereignty respected in the absence of written Tribal laws

- **NIH plans to develop guidance to:**
  - Facilitate respectful partnerships between AI/AN communities and researchers to align data management and sharing goals
  - Emphasize respect for sovereignty
  - Ensure Tribal input into NIH data management and sharing practices to increase trust and participation in biomedical research

Policy Assessment

What's Next?

- **Establishing evaluation metrics:**
  - To assess short-term goals…
    - Compliance with Plan submission
    - Compliance with the data management and sharing outlined in Plans
    - What data management and sharing costs are budgeted
  - To assess long-term goals…
    - Increasing high-quality data sharing that results in secondary data usage
    - Increasing public transparency through data management and sharing
    - Promoting rigorous and reproducible research

Incentivizing Data Sharing

What's Next?

- **Consider promoting usage of PIDs** to facilitate FAIR data sharing and policy compliance monitoring
  - Align PID language in relevant NIH instructions (e.g., in Final RPPR)
  - Determine ways to utilize PIDs in policy compliance monitoring
  - Establish a standard way to report PIDs in NIH applications and reports
  - Explore potential policy mechanisms to increase PIDs (e.g., trainee ORCID iDs)
- **Mechanisms to encourage data citation** – Explore ways researchers can be rewarded for being good data sharers through data citation (e.g., in reports and applications)

Implementation: Comparison of Key Data Sharing Policies

Policy Provision	NIH Data Management and Sharing Policy	NIH Genomic Data Sharing Policy
Scope	• All NIH-supported research generating **scientific data**.	• NIH-supported research that **generates 'large-scale' human or non-human genomic data** as well as the use of these data for subsequent research.
Timeline for Sharing Data	• Shared scientific data should be made accessible **as soon as possible**, and **no later than the time of an associated publication**, or the **end of performance period**, whichever comes first.	• Non-human data: Should be made available no later than the **date of initial publication** • Human data: Submit data in a **timely manner** or no later than the **date of initial publication** • Smaller submission intervals expected for various data processing levels (**e.g., 3 months**)
Plan Elements	• Data type • Related tools, software and/or code • Standards • Data preservation, access, and associated timelines • Access, distribution, or reuse considerations • Oversight of data management and sharing	• Data type • Data repository • Timeline for sharing data • IRB or analogous review (human data) • Appropriate use of data (human data)
Plan Review	• Plan is peer-reviewed for reasonable costs and reviewed by NIH Programmatic Staff for adequacy.	• Plan is peer-reviewed for appropriateness.
Repository Specifications	• Encourages use of **established repositories**	• Genomic research data expected to be submitted to an **NIH-designated data repository**

Example Questions and Feedback Received

- **Common questions or considerations**
  - What unpublished data is appropriate to share?
  - Does NIH plan to provide sufficient repositories for increased data sharing?
  - Data sharing timelines and compliance for SBIR grantees
  - Data sharing expectations for qualitative data

Example Questions Received on Costs

- What is/is not an allowable cost?
  - What is considered a direct cost and what is an F&A cost?
  - Would hiring data scientists be considered an allowable cost?
  - Are costs for cloud storage considered allowable data management?
- Are grantees allowed to anticipate costs after the period of the performance and request funds for the costs?
- Where should costs associated with collecting data and gaining access to data be requested?

Overview

- Evolution of Scientific Data Sharing Efforts & Development of the NIH Data Management and Sharing Policy
- Overview of the NIH Data Management and Sharing Policy and Supplemental Information
- Implementation Activities and Next Steps
- Outreach and Discussion

Policy Presentations and Outreach

- Washington University, St. Louis
- Johns Hopkins University Institute for Clinical and Translational Research Lecture series
- Loyola University, Chicago
- Module for Research Data Management Librarian Academy
- Council on Government Relations
- INCF Assembly publishers roundtable
- STM Spring Conference Society Day 2021
- NSTC Subcommittee on Open Science
- NIH advisory councils:
  - NIH Council of Councils
  - NIA
  - NIAMS
  - NIDDK
  - NIDDK
  - NHGRI

NASEM Workshop on Changing the Culture of Data Management and Sharing

- **Workshop aimed to…**
  - Identify community training and resource needs
  - Learn about challenges anticipated in implementing policy expectations
  - Understand what the community envisions successful data sharing to look like and how to measure it
- **~2,000 participants**
- https://www.nationalacademies.org/our-work/changing-the-culture-of-data-management-and-sharing-a-workshop

Implementation Considerations from the NASEM Workshop

- **Implementation requires a system-wide culture shift**
  - Need for aligned incentives and resources from NIH and all biomedical ecosystem stakeholders (e.g., other funders, institutions, publishers, data repositories, and associations)
- **Impactful data sharing is key to successful policy implementation**
  - Data management practices mindful of secondary data users are necessary for useful data sharing
  - Collecting and analyzing data sharing metrics are needed to understand the value of data sharing
  - Data citation adoption promotes data reuse and aids compliance monitoring
  - Trust needs to be earned to enable data sharing

Community-led Efforts Promoting Data Sharing

AAU and APLU report recommending how institutions and Federal agencies can improve access to research data

FASEB's data reuse program

Request for Information: Developing Consent Language for Future Use of Data and Biospecimens

NOT-OD-21-131

- NIH requests information on “Points to Consider” and sample language meant to serve as a resource for investigators as they tailor appropriate informed consent language to their studies
- Comments should be submitted by **September 29, 2021**
- **https://osp.od.nih.gov/rfi-comment-informed-consent-sharing/**

Discussion

- Which areas are of highest priority for policy implementation?
- Are there areas the FDP community could contribute to which would help NIH and the scientific community implement the Policy?
- What do you consider to be administratively burdensome? Are there strategies to reduce the burden of complying with this Policy?

Thank You!

- OSP Data Management and Sharing Website
- 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

sciencepolicy@mail.nih.gov