NIH Data Management Sharing Plan Overview

Emily Boja, Ph.D. – Office of Data Sharing, National Cancer Institute
Emily.boja@nih.gov

Michelle Bulls - Office of Policy for Extramural Research Administration, National Institutes of Health
Michelle.bulls@nih.gov

Mordecai Tayebwa – Makerere University, Kampala, Uganda
Tayebwa.Mordecai@gmail.com

Saundra Harris – St. Jude Children's Research Hospital, Inc.
Saundra.harris@stjude.org

Leronda Savage, CRA – St. Jude Children's Research Hospital, Inc.
Leronda.savage@stjude.org

NIH Data Management and Sharing Policy: Driving Towards Broad and Equitable Sharing

Emily Boja, Ph.D.
Chief, Scientific Policy and Program Branch,
Office of Data Sharing, National Center Institute

NIH NATIONAL CANCER INSTITUTE

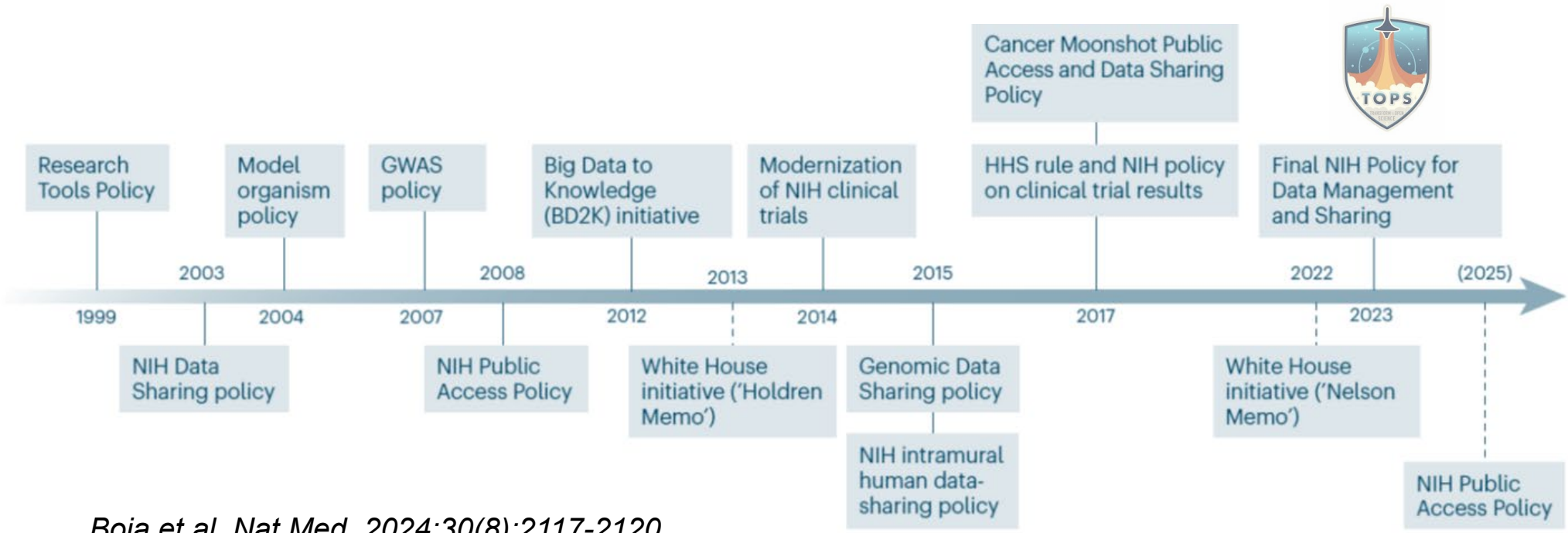
October 29, 2024

Agenda

1. Introduction
2. NIH Data Management and Sharing (DMS) Policy
3. NCI Implementation Approaches
4. Lessons Learned & Next Steps

Driving Data Sharing Towards A Broad Model

Collaborator Sharing Model	Consortium Sharing Model	Broad Sharing Model
•Sharing occurs between investigator to investigator •E.g., an individual request to access data upon publication of another researcher's work	•Sharing within large collaborative groups (e.g., sharing between investigators within a consortium) •E.g., TCGA prior to public release	•Sharing with the wider research community (e.g., NIH GDS and DMS Policies) •Most impactful to the society



Boja et al. Nat Med. 2024;30(8):2117-2120.

Final NIH Policy for Data Management and Sharing

On and after January 25, 2023

Submit a data management and sharing (DMS) plan → share the high-value data files

Agreed-upon plans become part of the terms and conditions of award. Adhere to approved plans or request revisions to approved plans for review by NIH programs.

- Good data management & sharing is critical to good research - accelerate scientific progress
- Follow policy expectations and guidance:
  - Refer to definitions in the DMS policy: data sharing, scientific data, controlled access, etc.
  - Clearly articulate which data, when, where, and how
  - Justify any deviations from policy expectations (articulate why it's impossible or reasonable) – "justifiable limitations"
- Help us (NIH, repository, researchers, etc.) work together to navigate perceived and/or real barriers

Elements To Demonstrate Prospective Planning

Description of the data plus metadata and documentation

Related tools, software, code, etc.

Standards for the data & metadata

Data preservation, access, and associated timelines

Access, distribution, and reuse considerations

Oversight of data management and sharing

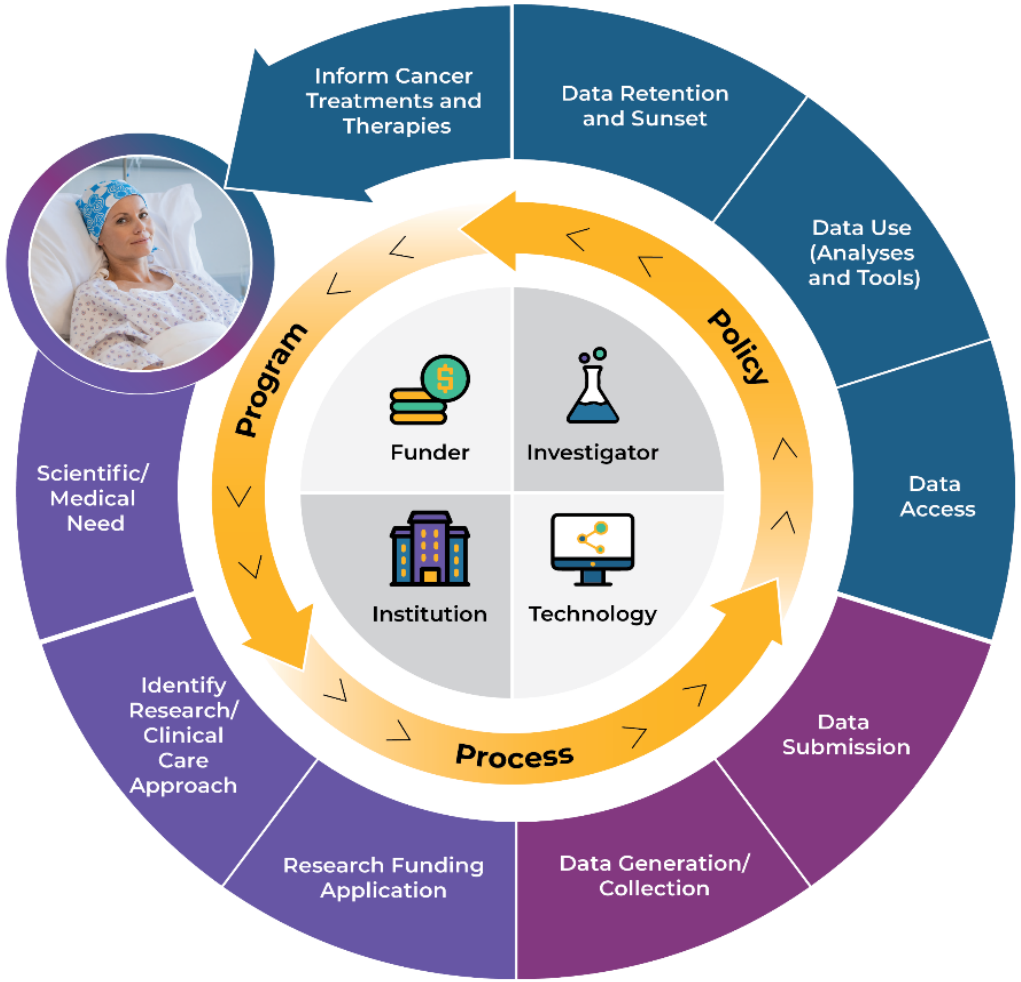
NCI's Iterative Approaches

Provide NCI staff and investigators with training to reiterate and expand upon the Offices of Extramural Research & Science Policy guidance

Develop and disseminate resources, advise programs, and enhance communication between investigators, program staff and NCI repositories

Develop evidence-based guidance for managing and sharing different data types to maximize impact on the wider research community

NCI Office of Data Sharing: Mission and Goals



Mission:

Direct a comprehensive data sharing vision and strategy for NCI which advocates for the proper balance between *broad and equitable data sharing* and the needs of the cancer research and participant communities.

Data Sharing Policy:

- Defines expectations for how NCI-supported data will be made available to the broader research community for novel discovery

Data Sharing Program:

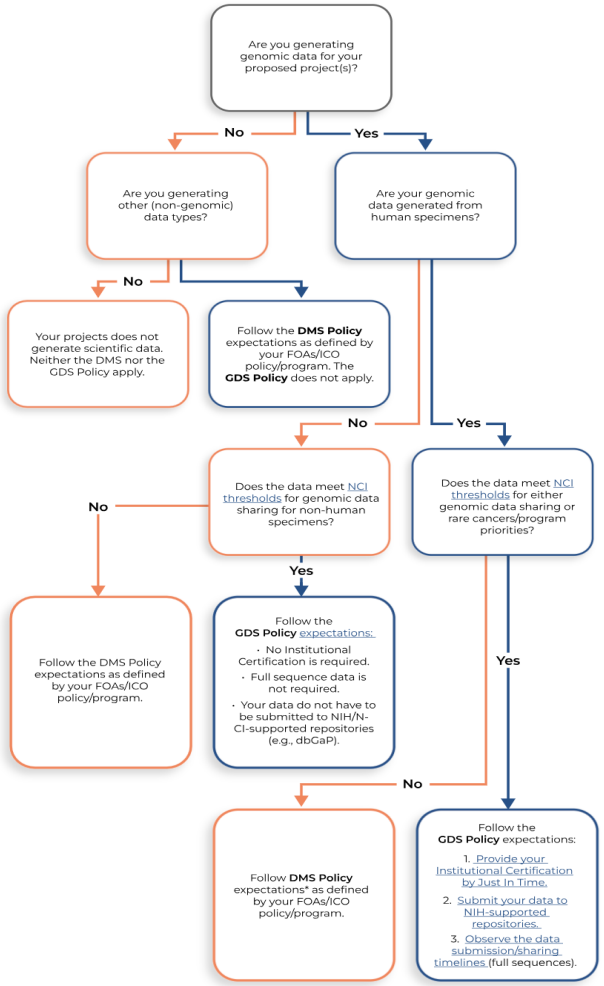
- Oversee or collaborate on scientific management of high-profile initiatives that serve as cornerstones for NCI's broader approach to data sharing (CCDI, Moonshot, Kids First)

Data Sharing Process:

- Establish & coordinate efficient registration, access and compliance procedures that align with policy expectations for NCI-supported studies

Tips: NIH Genomic Data Sharing (GDS) Policy

- GDS Policy requirements still apply and should be described in the DMS Plan.
- **Which data**: Specific data types, levels and formats expected to be shared for projects subject to the GDS Policy (human and non-human)
- **When**: Genomic data must adhere to the GDS Policy timelines:
  - Submitting data 3 months after last data generation
  - Releasing no later than 6 months after submission for human data.
- **Where**: Genomic data repositories (human and non-human)
- **How**: Provide GDS Institutional Certification with Data Use Limitations (DULs) based on consent forms and IRB determination (human only) (It is NIH policy that data should be shared with minimal restrictions as possible, consistent with any limitations on data sharing based on the consent form).



Tips: DMS Plan Elements 1, 2 & 3

1. Data Type
2. Related tools, software, code
3. Standards
4. Data preservation, access, timelines
5. Access, distribution, reuse considerations
6. Oversight of data management

- What kind(s) of data and metadata?
- How much data will be generated or collected?
- What won't be shared? Why?

Glossary for Broad Data Types

1. Data Type
2. Related tools, software, code
3. Standards
4. Data preservation, access, timelines
5. Access, distribution, reuse considerations
6. Oversight of data management

- *What will another researcher or data user need to access and use my data? Note: not the same as Resource Sharing*
- *Are the tools free (e.g., GitHub)?*

1. Data Type
2. Related tools, software, code
3. Standards
4. Data preservation, access, timelines
5. Access, distribution, reuse considerations
6. Oversight of data management

- *Data Standards*
  - *NIH CDE Repository*
  - *NCI thesaurus*
  - *NCI Metathesaurus*
- *Standard file formats*
- *Data models (e.g., mCODE)*

Tips: DMS Plan Element 4

1. Data Type
2. Related tools, software, code
3. Standards
4. Data preservation, access, timelines
5. Access, distribution, reuse considerations
6. Oversight of data management

- Select established repositories for relevant data types (use a table):
  - Supplemental materials in a publication is not considered as a true repository.
  - Avoid selecting local servers and institutional repositories that are not accessible to a wide community to the extent possible.
  - Desirable repository characteristics
- How will the data be findable (e.g., indexing, persistent identifier?)
- For how long will data be available?
  - Is it perpetuity or 10-year max (may depend on the repositories, program expectations, and other policies)?
- When will the data be available?
  - At the time of publication or by the end of award, whichever is first**

***Submission and release timelines for genomic data subject to GDS: https://sharing.nih.gov/genomic-data-sharing-policy/submitting-genomic-data/data-submission-and-release-expectations*

Tips: DMS Plan Element 5 & 6

1. Data Type
2. Related tools, software, code
3. Standards
4. Data preservation, access, timelines
5. Access, distribution, reuse considerations
6. Oversight of data management

- Think like a data user
- Describe limitations if applicable
  - What type of access (e.g., open access or controlled access)
  - Any special considerations, e.g., human data, laws, data use agreements, etc.

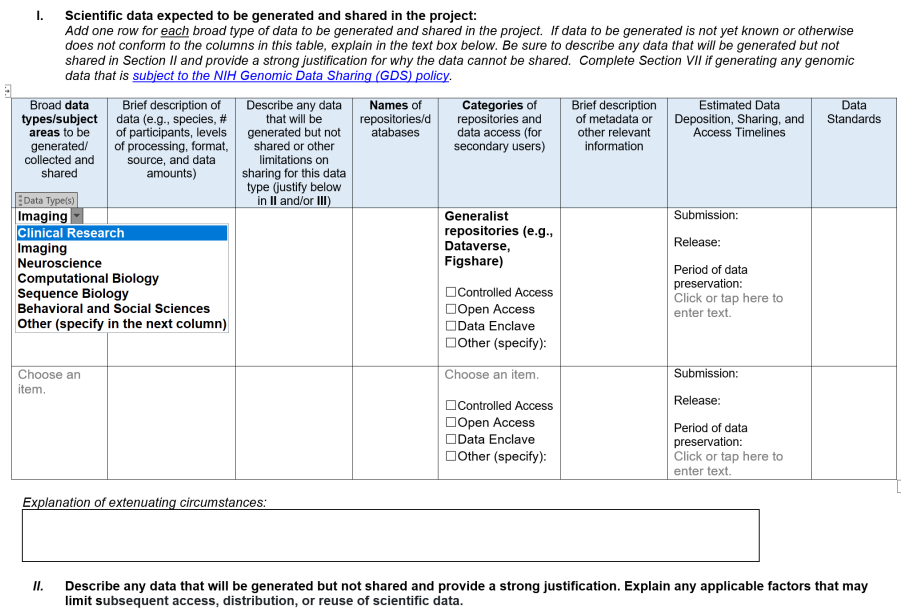
1. Data Type
2. Related tools, software, code
3. Standards
4. Data preservation, access, timelines
5. Access, distribution, reuse considerations
6. Oversight of data management

- *How will the applicant ensure the plan is being met?*
- *Who are the key personnel responsible for oversight?*
- *How often and how is oversight conducted?*

Data Management and Sharing Pilot

1. Test features of NIH DMS templates
2. Develop cost policies
3. Develop reporting policies

"Charlie Template"



I. Scientific data expected to be generated and shared in the project:
Add one row for each broad type of data to be generated and shared in the project. If data to be generated is not yet known or otherwise does not conform to the columns in this table, explain in the text box below. Be sure to describe any data that will be generated but not shared in Section II and provide a strong justification for why the data cannot be shared. Complete Section VII if generating any genomic data that is subject to the NIH Genomic Data Sharing (GDS) policy.

Broad **data types/subject areas** to be generated/ collected and shared	Brief description of data (e.g., species, # of participants, levels of processing, format, source, and data amounts)	Describe any data that will be generated but not shared or other limitations on sharing for this data type (justify below in **II** and/or **III**)	**Names** of repositories/databases	**Categories** of repositories and data access (for secondary users)	Brief description of metadata or other relevant information	Estimated Data Deposition, Sharing, and Access Timelines	Data Standards
Data Type(s) **Imaging** Clinical Research **Imaging** **Neuroscience** **Computational Biology** **Sequence Biology** **Behavioral and Social Sciences** **Other (specify in the next column)**				**Generalist repositories (e.g., Dataverse, Figshare)** ☐Controlled Access ☐Open Access ☐Data Enclave ☐Other (specify):		Submission: Release: Period of data preservation: Click or tap here to enter text.	
Choose an item.				Choose an item. ☐Controlled Access ☐Open Access ☐Data Enclave ☐Other (specify):		Submission: Release: Period of data preservation: Click or tap here to enter text.	

Explanation of extenuating circumstances:

II. **Describe any data that will be generated but not shared and provide a strong justification. Explain any applicable factors that may limit subsequent access, distribution, or reuse of scientific data.**

Reporting Data Management and Sharing (DMS) Plan Activities in Research Performance Progress Reports (RPPRs)

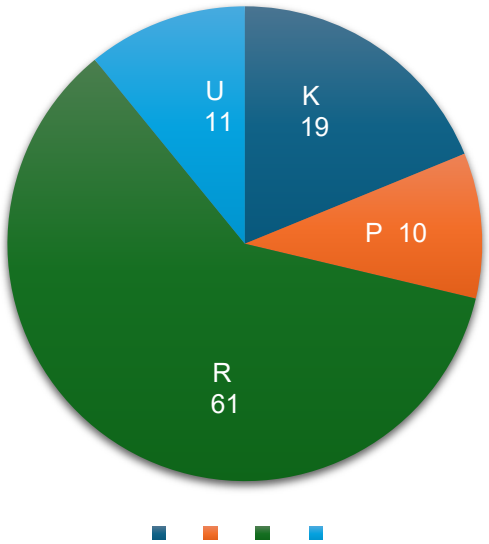
NOT-OD-24-175.html

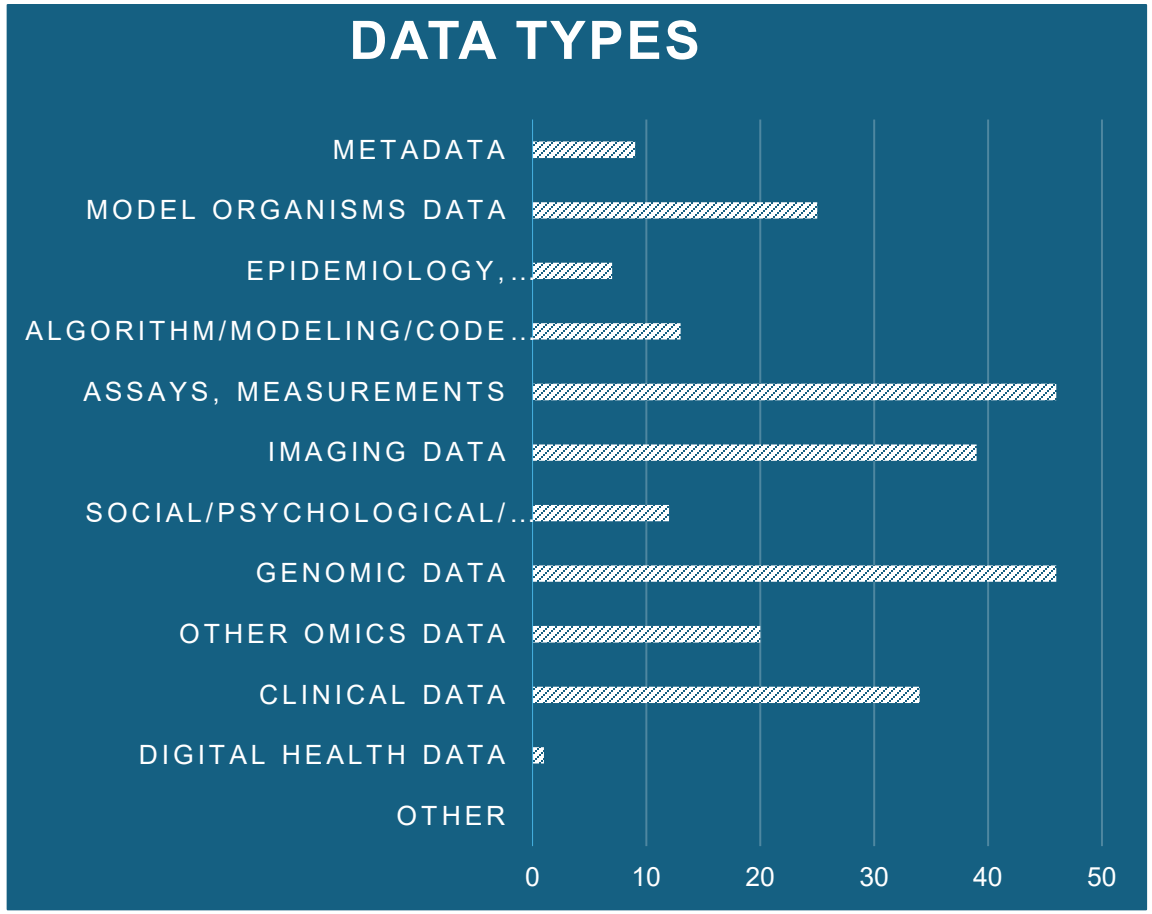
On October 1, 2024, NIH will issue new DMS questions that align with the NIH Final Policy on Data Management and Sharing, asking recipients for updates on the status of data sharing, repositories, and unique identifiers for data that have been shared.

https://thefdp.org/demonstrations-resources/nih-data-management-sharing-pilot/

Evaluate DMS Plans: Learning from Real Data

- 101 randomly sampled DMS plans from *NCI grants within fundable range*: 60 plans were acceptable as written, 31 have minor corrections, 10 have major revisions.
- Overall, 90% will meet policy expectations with little or no efforts.
- 66 randomly sampled DMS plans *from awarded grants* also subsequently evaluated.





Recurring Issues Observed

Overall	• Elements of the NIH format page are not used the same way - **need to review Plan as a whole** • Conflicting information (contradictions across elements) • Vague reasons for not sharing • Inconsistency between data types in DMS plans and Research Strategy • Planning to share only through “publication” and “conferences”
Which data	• Unclear which data will be ***generated*** versus ***shared*** • Lacking important details: species/source, time course or not, formats shared, amount, metadata
When	• Missing adherence to GDS policy timelines for human genomic data • Stating by the time of publication (but not end of project period whichever occurs first) • Duration follows *local* data retention cycles (but should follow *repository retention timelines*)
Where	• Not using an established **repository** or not committing to one • Planning to share only through “publication” and “conferences” • Naming a data repository that is inappropriate (not broadly accessible, data type) • When listing multiple repositories, not indicating which data goes to which • Over-reliance on “generalist repositories” (should prioritize discipline-specific repositories)
How	• Sharing “**by request**” or with **Investigator-control** (even if using a repository) • Plan for sharing is too restrictive and/or limitations are not adequately justified
Budget	• DMS budget justifications are missing or not clear • No explicit support for DMS Plan activities (e.g., personnel for preparing submissions)

Tips: DMS Plan DOs and DON'Ts

Dos

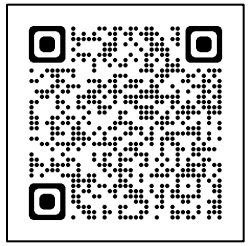
- ✓ Be a stand-alone document (one plan)
- ✓ Use a template
- ✓ Keep it concise
- ✓ Demonstrate thoughtful planning on management and sharing of data
- ✓ Provide your best estimate for 6 Plan element specifics
- ✓ Avoid jargons and spell out acronyms
- ✓ Consider all applicable policies
- ✓ Harmonize information on scientific data generated and shared in DMS plans with that in the Research Strategy

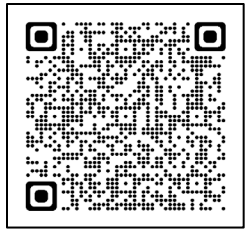
Don'ts

- X Be part of "Resource Sharing Plan" (Confuse DMS Plan with software/ code sharing)
- X Simply state "we will follow the policy"
- X Write more than 1 Plan
- X Reference page numbers elsewhere in applications for DMS information
- X Include confidential or proprietary information
- X Skip Plan elements
- X Copy and paste NIH Sample Plans

Some NIH and NCI Resources

Resources	Use
Policy Decision Tool	Use this web application to see which NIH policies might apply to your research
Toolkit for DMS Policy	Learn more about the DMS policy by visiting the National Library of Medicine
OER/OSP Guidance on Informed Consent for Secondary Research with Data and Biospecimens	Write your informed consent documents and maximize data sharing
Applicable Activity Codes for DMSP	Review this list of activity codes to see if you need to write a DMS Plan for your research
DMPTool	Walk you through creating DMS plans based on NIH templates.
Tips for Writing a DMS Plan	Help you write your plan for NCI-funded or sponsored research.
Decision Tree for GDS and 2023 DMS Policy Harmonization	Answer questions to see if your research must apply the GDS Policy, 2023 DMS Policy, or both
Glossary for Broad Data Types for DMS Plans	Review this glossary to get ideas for accurately describing your research data.
NIH-Supported Scientific Data Repositories	Explore this list of repositories which house various types of data.
NIH Federal Demonstration Partnership	Write your DMS plan using one, or preferably both, templates offered by the Federal Demonstration Partnership.





NIH Data Sharing Index (S-index) Challenge

Mandatory Registration [March 2025]

Phase 1 Deadline [June 2025]

Phase 2 Deadline [February 2026]

Join us in promoting data sharing by developing a robust metric to reward exemplary data sharers

NIH National Institutes of Health
Turning Discovery Into Health

Scan to learn more about this Challenge

NIH Federal Demonstration Partnership Data Management and Sharing Pilot Update

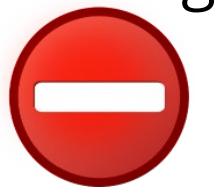
MICHELLE G. BULLS, DIRECTOR
OFFICE OF POLICY FOR EXTRAMURAL RESEARCH ADMINISTRATION

October 29, 2024

National Institutes of Health
Office of Extramural Research

Federal Demonstration Partnership (FDP) DMS Pilot - Evolution of the DMS Plan Template

- Initially, OER developed an optional DMS Plan Format Page that aligns with the recommended elements of a DMS Plan.
- Many ICs wanted to issue IC-specific templates.



- NIH and the FDP are conducting a pilot with the goal of developing a single NIH template for DMS Plans.
- Pilot included testing of two DMS Plan templates, ‘Alpha’ and ‘Bravo.’
- Feedback was collected from the applicant community, as well as NIH Program Officials.
- NIH and FDP used this feedback to develop a new DMS plan template, ‘Charlie.’

Using Feedback to Develop a New Template

- Utilize tables where possible
  - Helps to ensure the required information is provided for every data type
  - Facilitates monitoring and reporting

- Utilize drop-down menus where appropriate
  - Standardizes the information provided while also providing guidance on the information expected

- Use established lists/terms where possible
  - Example: using the Trans-NIH BioMedical Informatics Coordinating Committee (BMIC) Subject Areas to capture broad data types

- Section for describing limitations arising from consent/privacy considerations only needs to be completed if data derived from humans will be generated

- Design question wording to also provide guidance on expectations
  - Example: requesting different information related to the data sharing timeline to emphasize that there could be differences in deposit versus release timelines

FDP DMS Pilot – What's next?

- NIH is finalizing its plans for User Acceptance Testing of the Charlie template – stay tuned!
- NIH will update the RPPR questions in the future, to align with the final format page.
- Pilot Phase 2: Cost Policies
    - Establish common cost principles, identify types of costs required, and determine how to identify additional/unforeseen costs that may be required to meet the spirit of the data sharing policy.
    - **Next steps:** Small working group brainstorming steps to catalogue the activities that should be included in a cost/effort calculator.

DSMP Challenges in LMIC

- **Institutional Challenges:**

1. Limited institutional capacity for data management
2. Inadequate data governance structures.
3. Lack of standard operating procedures (SOPs) for data management.
4. Insufficient resources (financial, human, infrastructure).
5. Limited access to reliable internet and technology.
6. Inadequate data storage

- **Technical Challenges:**

1. Limited use of data management software and tools.
2. Inconsistent data formats and standards.

DSMP Challenges in LMIC

- **Regulatory Challenges:**

1. Weak or absent data protection laws.
2. Lack of clarity on data ownership and intellectual property.
3. Inconsistent regulations across countries or regions.
4. Limited enforcement of data management policies.

- **Social and Cultural Challenges:**

1. Limited awareness and understanding of data management importance.
2. Cultural and social norms around data sharing and privacy
3. Limited trust in data management systems.
4. Language barriers and limited accessibility.

DSMP Challenges in LMIC

- **Financial Challenges:**

1. Limited funding for data management infrastructure.
2. High costs of data collection and management.
3. Limited resources for data analysis and interpretation.
4. Dependence on external funding sources.

- **Human Resource Challenges:**

1. Limited skilled personnel for data management.
2. Limited training and capacity-building opportunities.
3. Inadequate data management expertise

Mitigation measures

1. Establish international partnerships and collaborations.
2. Invest in data infrastructure and tools
3. Provide training and capacity-building programs
4. Engage in open and transparent communication
5. Monitor and adapt to changing regulations and standards.

NIH Data Management and Sharing (DMSP)

October 29, 2024

Presented by:
Saundra Harris and Leronda Savage

Data Management Sharing Plan

Central Office Perspective

❖ **Challenges**

1. Coordination of compliance offices
2. Identifying resources within the institution
3. Expectations at JIT stage
4. Monitoring
5. Budget considerations

Data Management Sharing Plan

Central Office Perspective

❖ **Wins**

1. Participant of the FDP pilot program
2. Data Management Librarian
3. Process for review at proposal stage
4. DSMP questions incorporated in the grant management system

Data Management Sharing Plan

Departmental Administration

1. Preparation with budget
2. Preliminary review
3. Single point of contact
4. Department Process – Coordination of overview of DMSP for new faculty

Data Management Sharing Plan

- Questions/Panel Discussion