National Institutes of Health
Office of Science Policy

Implementing the NIH Data Management and Sharing Policy within the NIH IRP

Presentation to OHSRP Lecture Series

February 16, 2023

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

Chuck Dearolf, PhD
Director of Program Development and Support
Office of Intramural Research
National Institutes of Health

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

Examples of NIH ICO and Domain Specific Data Sharing Policies*

- **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
- **NIMH Data Sharing Policy**
  - Expects all raw and analyzed data from NIMH-funded human subjects research 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
- **HEAL Initiative Public Access and Data Sharing Policy**
  - Expects a Public Access and Data Sharing Plan from HEAL funding with proposed process for making and their underlying primary data publicly available immediately to the extent possible

*Non-comprehensive list of NIH data sharing policies

NIH 2015 Intramural Human Data Sharing Policy - Manual Chapter 3016

- **Applies to all IRP studies generating human data, including the Clinical Center and ICs, undergoing scientific review after Oct. 1, 2015**

- **Requires a Data Sharing Plan for any research involving human data that should include:**
  - a **commitment to share, at a minimum, the data underlying any publications resulting from the research** or an explanation of why sharing is not possible
  - timeline for making data publicly accessible, in general, **no later than time of publication of main findings**
  - any intellectual property issues (e.g., patent filings) or contractual obligations that would **preclude sharing and secondary research** with the data

- **Scientific Director or their designee approves all Data Sharing Plans**

- **All IRP-supported clinical investigators are expected to develop protocols and consent processes/forms to enable broad data sharing for secondary research**

- **Investigators are encouraged to deposit data in publicly accessible research repositories for sharing to the extent feasible and appropriate**

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)*

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)

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**

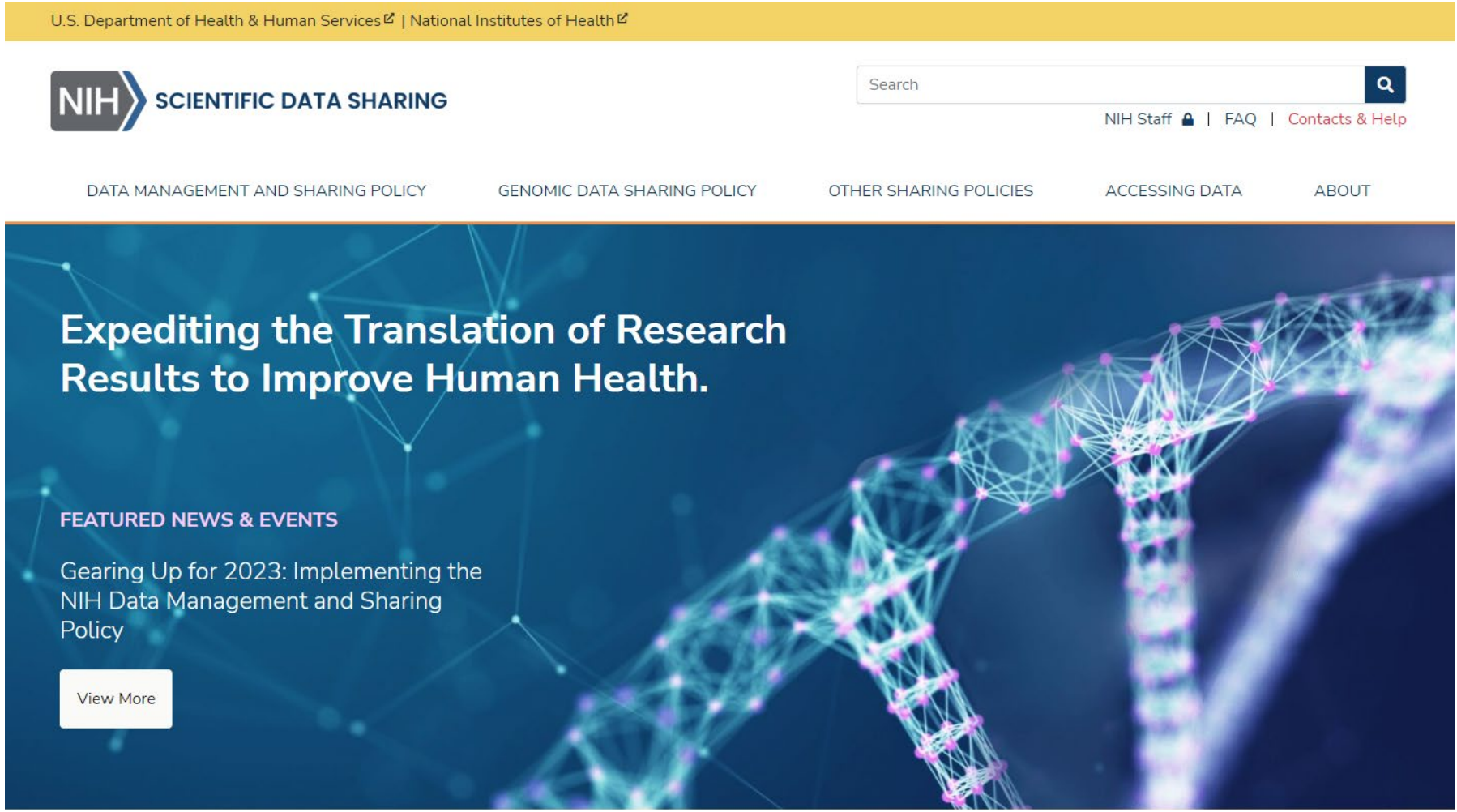
  - **Responsibly implemented**
    - Plans should outline protection of privacy, rights, and confidentiality
    - Abide by existing laws, regulations, and policies
  - **Prospectively planned for at all stages of the research process**

Potential Limitations on Sharing

- **Data Management and Sharing Plans should 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

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



Elements of a Data Management and Sharing Plan (Intramural Template)

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

A. Data standards for clinical protocols – CDEs

B. Data standards for all plans

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

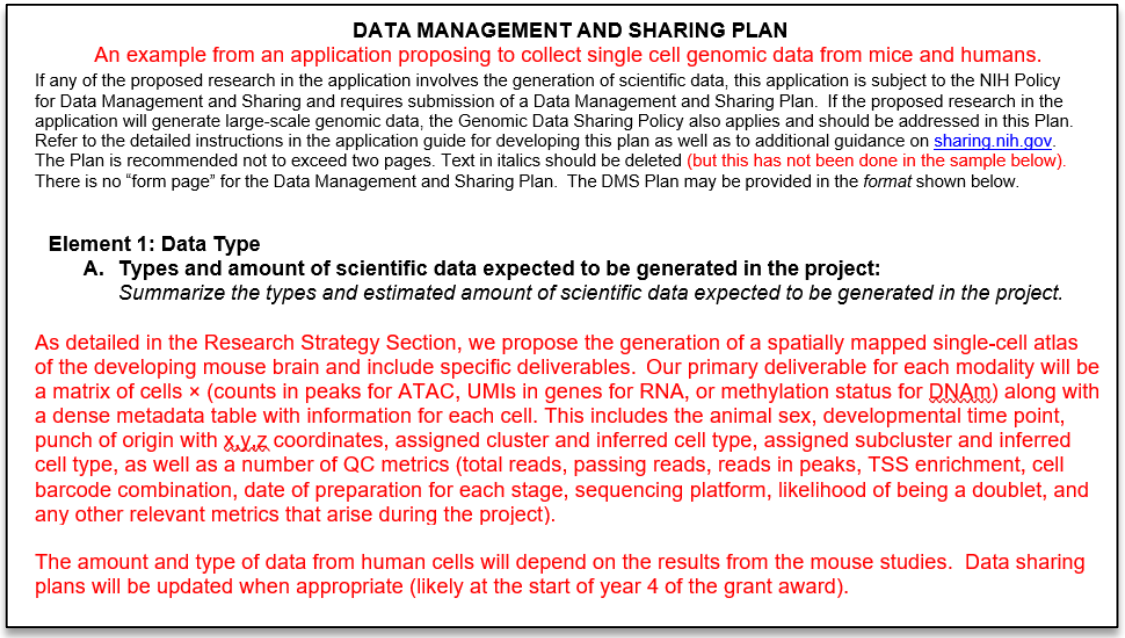
Element 6: Oversight of Data Management and Sharing

Element 7: Other Elements (if applicable)

See Writing a Data Management & Sharing Plan for details and IRP DMS Plan template

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. 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

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

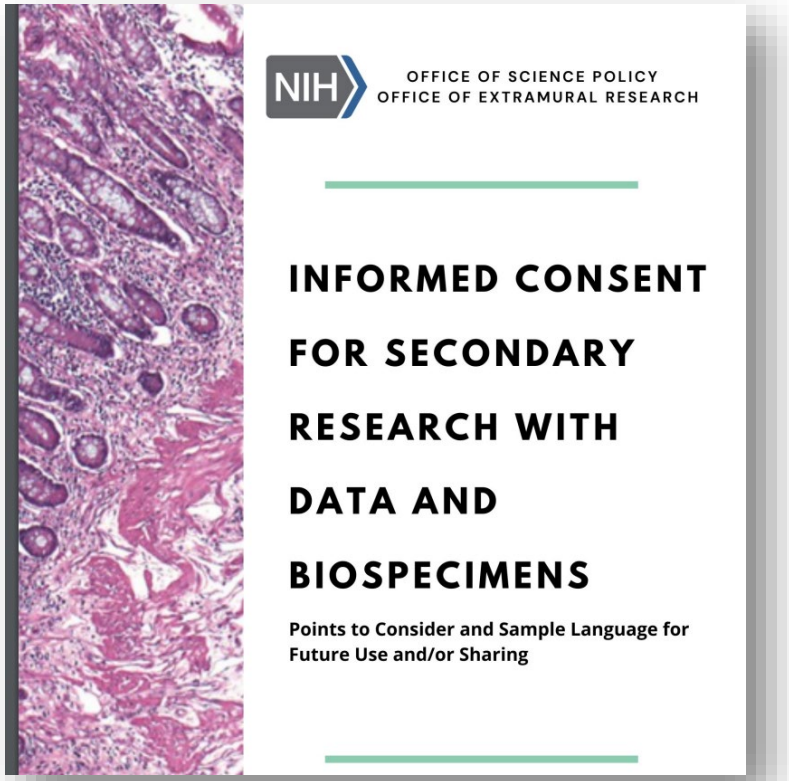
Generalist Repository Ecosystem Initiative

- **Introduction to Generalist Repositories for NIH Data Sharing**
  September 15 at 3pm ET / Noon P
- **Meet the GREI Generalist Repositories**
  October 12 at 1pm ET / 10am PT
- **How to include generalist repositories in your NIH data management and sharing plans**
  November 10 at 3pm ET / Noon P
- **Best practices for sharing data in a generalist repository: Metadata, data preparation, and reporting**
  December 8 at 3pm ET / Noon PT

- **Virtual Generalist Repository Ecosystem Initiative (GREI) Workshop**
  January 24-25 at 11am ET

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: Protecting Privacy When Sharing Human Research Participant Data

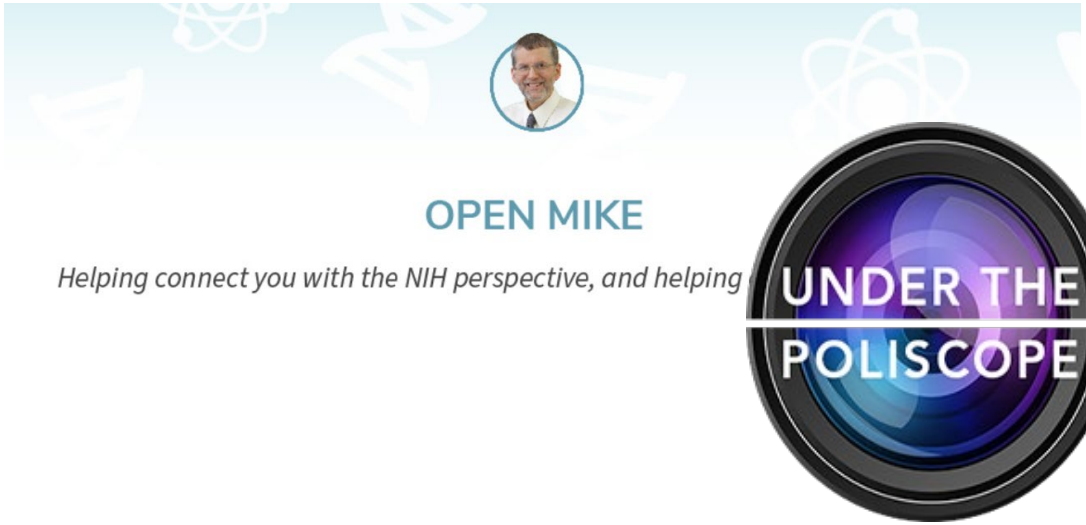
- Provides a basic **framework for considering how to protect privacy** when sharing data from human participants
- Not intended as a guide for regulatory compliance
- **Broadly applicable** to different research contexts
- Establishes shared principles, provides best practices, and offers considerations 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

Roadmap for 2023 and Beyond



- **OSP *Under the Poliscope* and *Open Mike* blogs** provide a general roadmap for what to expect
- **Out now!**
  - NIH 2-part 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
- **Ongoing in 2023 and beyond:**
  - Additional FAQs and guidance
  - Ongoing assessment of the Policy for short- and long-term goals
  - Incentives for data sharing

IRP DMS Working Groups

WG 1 (Sept. – Dec. 2021): Bob Balaban, Andy Baxevanis, Brian Brooks, Luigi Ferrucci, Michael Gottesman, Richard Wyatt, Chuck Dearolf

WG 2 (May 12, 2022 – June 29, 2022): Chuck Dearolf (chair), Andy Baxevanis, Yang Fann, Luigi Ferrucci, John Gallin, Jonathan Green, Janice Lee, Glenn Merlino, Charles Rotimi, Michael Gottesman, Richard Wyatt

Requirement

- IRP investigators will:
  - Prospectively plan for the managing and sharing of scientific data
  - Submit a DMS plan
  - Comply with the approved plan

Scope for the IRP

- Applies to all IRP research associated with a ZIA and/or a clinical protocol that will be conducted on or after January 25, 2023
- Plans will include the material requested in the IRP DMS template
- Research subject to the NIH Genomic Data Sharing policy and the 2015 IRP Human Data Sharing policy will use the IRP DMS plan template

Guiding Principles

- We will make a good faith effort to comply, recognizing that there will be a learning curve
- Maximize the sharing of scientific data, when possible
- Justify if data cannot be shared or if sharing is limited
- Use existing repositories to share data, when possible
- The Policy leaves many decisions to the investigator and the IC approver (SD, CD, or designee)

Two Pathways for DMS Plan Submission and Review

- All plans will be entered as text into the NIDB
  - Pathway 1 - For research associated with a clinical protocol
  - Pathway 2 - Other research that is associated with a ZIA

For research associated with a clinical protocol

- The new plan template must be used on/after January 25, 2023
- Plans will be entered as text into the NIDB, then exported
- Submit the plan along with other protocol materials into the PROTECT system as part of the IC Initial Scientific Review
- It is not sufficient to indicate that the data will be deposited in ClinicalTrials.Gov - While this meets the requirements of FDAAA, it is not sufficient for the new DMS policy
- Plans will be reviewed by the IC Scientific Review Committee
- For protocols submitted prior to January 25, 2023, submit a DMS plan at the time of the quadrennial review

For ZIA Research Not Associated with a Protocol (1)

- Plans will be entered into the NIDB
- Plans for ongoing projects must be submitted by Jan 25, 2023
- The SDs or their designees will review and approve/disapprove the plans

For ZIA Research Not Associated with a Protocol (2)

- The same DMS plan can be used by an investigator who has multiple ZIAs with overlapping projects, provided the plan covers all the research to be conducted
- In the future, new and updated DMS plans must be entered by the time of the annual reports process
- To aid compliance/monitoring, future years will include a question on how the investigator complied with the plan during the annual reporting period

Resources

- NIH resources:
  - NIH Data Sharing website, Sharing.NIH.gov
  - Email questions to Sharing@nih.gov
- OIR Sourcebook, https://oir.nih.gov/sourcebook/intramural-program-oversight/intramural-data-sharing/2023-nih-data-management-sharing-policy
- NIH Library offers one-on-one or group counseling, courses, and other support
- Sample DMS Plans – NIH, Indiana Univ., DMP Tool, repository samples
- IC Data Science offices