SRA INTERNATIONAL
ANNUAL MEETING
CHICAGO 2024
OCTOBER 26-30

How It Started and How It's Going: A Retrospective on New Data Management and Sharing Requirements and Case Studies for Emerging Best Practices

Alison Krauss, Huron Consulting Group
David Stewart, University of Houston

Today's Objectives

- Provide context and background for current data management and sharing policies
- Provide an overview of the current policy landscape
- Provide an overview of challenges and emerging best practices

Background

Data Management and Sharing Overview

How is data management and sharing defined?

"The process of validating, organizing, protecting, maintaining, and processing scientific data to ensure the accessibility, reliability, and quality of the data for its users."[1]

Why is it important?

To promote rigorous and reproducible research and bolster public trust and confidence in research.

[1] NIH Data Management and Sharing Policy

Distrust in Research

27% of Americans say they have little to no confidence that scientists act in the public interest.[1]

23% of people do not trust scientists to tell the truth.[2]

Research Misconduct – HHS reports around 13 research misconduct findings annually. In 45 trainee cases, 72% of supervisors hadn't reviewed the source data.[3]

Reproducibility crisis in scientific research

[1] Pew Research Center Trust in Scientists Report 2023
[2] Edelman Trust Barometer 2024
[3] Wright, D. E., Titus, S. L., Cornelison, J. B.. (2008). Mentoring and Research Misconduct: An Analysis of Research Mentoring i n Closed ORI Cases. Science and Engineering Ethics, 14, 323-336.

SRA INTERNATIONAL

Research Misconduct Case Studies

Falsification of Research Data

According to the U.S. Office of Research Integrity (ORI), Richard Eckert, the chair of the Department of Biochemistry and Molecular Biology at the University of Maryland, Baltimore, faked data in 13 published papers and two grant applications.

Case Summary: Eckert, Richard L | ORI - The Office of Research Integrity (hhs.gov)

False Reporting

ORI found that Bret Rutherford, formerly a Research Psychiatrist at NYSPI, committed research misconduct in funded studies. He falsely reported that all human subjects met the criteria for late-life depression studies in five published papers.

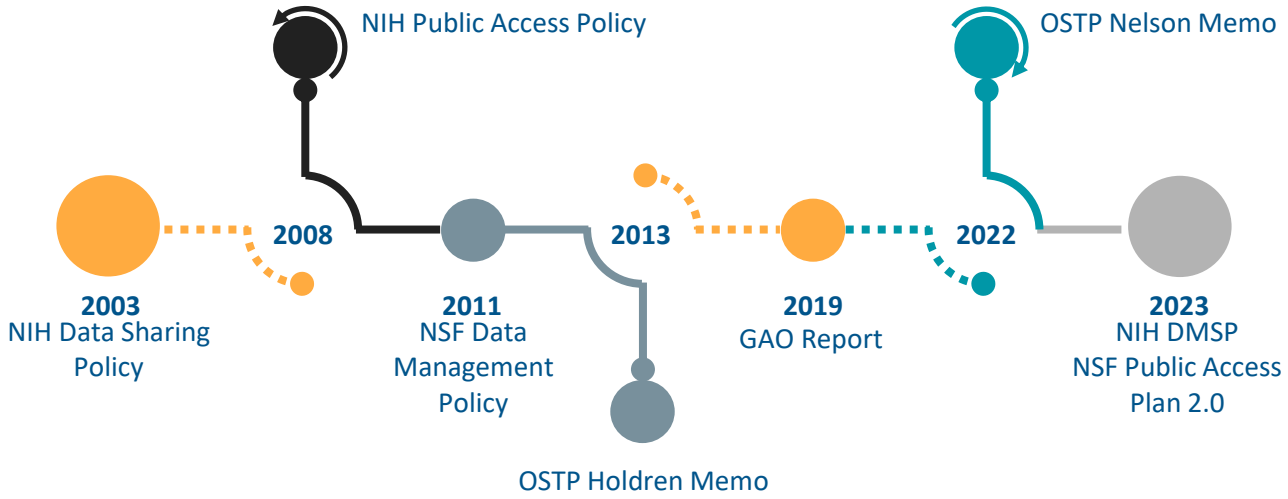
Case Summary: Rutherford, Bret | ORI - The Office of Research Integrity

Plagiarism

ORI determined that Dr. Arunoday K. Bhan, a Research Fellow in the Dept. of Pediatrics at Boston Children's Hospital and Harvard Medical School, falsified, fabricated, and/or plagiarized data in one published paper and two grant applications.

Case Summary: Bhan, Arunoday K. | ORI - The Office of Research Integrity

Policy and Regulatory Timeline



Current Policy Landscape

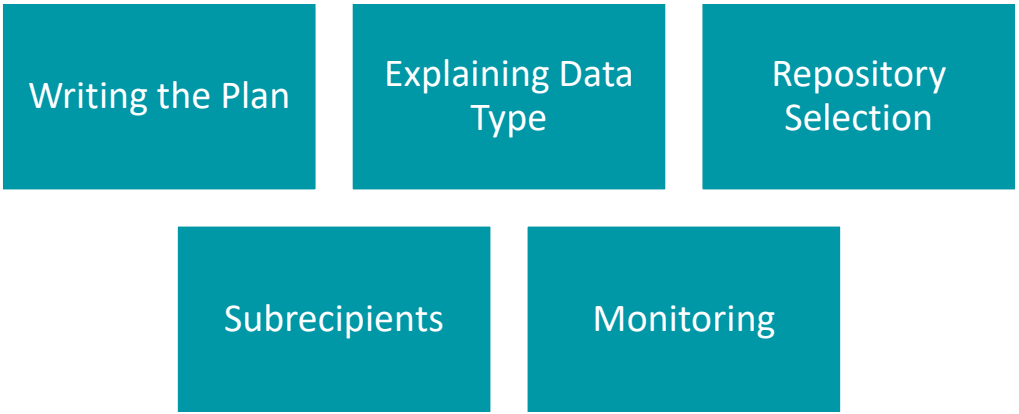
The National Institute of Standards and Technology has released the NIST Research Data Framework (RDaF) Version 2.0 as a tool for agencies and organizations improve research data management capacity and capabilities.

NIST ODI

Challenges and Best Practices

Plan Elements

Challenges and Emerging Best Practices



Writing the DMS Plan

NIH

- NIH has numerous example plans on it's "Writing a Data Management & Sharing Plan" website
- The same website provides a detailed description of what is expected for each element of the plan.

NSF

- Similar elements as required for NIH but limited to 2 pages.
- Directorate-specific guidance

Other Agencies

- Other agencies generally follow the same guidance as NIH.

DMP Tool

- Log in with individual or institutional license
- Provides a sponsor-specific template
  - Currently has 3 templates for NIH, including FDP alpha and bravo
- Provide example language to help you write the content
- Provide instructions from the sponsor
- Provides additional helpful guidance from your institution (provided by your data management librarian) and from the DMP Tool
- Allows all collaborators to work on the plan at the same time

Element 1: Data Type

Be sure to include all the following information:

- Type of data (images, measurements, sequencing, gene expression, etc.)
- Number of subjects/observations and whether combined
- File types and expected file sizes
- Metadata to be collected
- Which data will be preserved, and which will be shared

Best practices

- Create a chart with data type, number of subjects/observations, file type, and est. file sizes.
- Create a chart to indicate which metadata is to be associated with the data type.

Repository Selection

Recurring issues with repository selection[1]:

- Not using an established repository or not committing to one
- Planning to share only through “publication” and “conferences”
- Naming a repository that isn’t appropriate
- Not indicating which data goes to which repository when multiple repositories are listed
- Overreliance on generalist repositories

[1] FDP May 2024

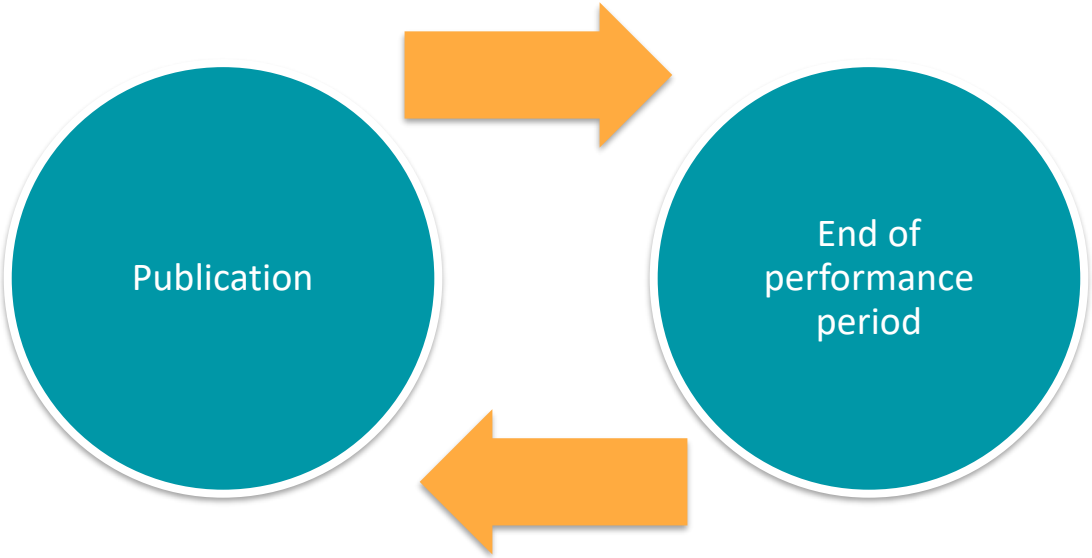
Repository Selection

Most PIs should already be aware of relevant repositories for their research. If they are unsure, there are several options available:

- Repositories that are discipline or data-type specific should be given preference.
- There are 147 NIH-supported repositories
- Some NOFOs specify a required repository
- Your data management librarian
- Generalist or Institutional

Timeliness of Data Sharing (Public Release)

Whichever comes first:



Subrecipients

Subrecipients must agree to share all data related to the project

- All documentation that supports the research outcomes described in the progress report
- Frequency: no less than once per year (align with timing of the RPPR)
- Data sharing should incorporate principles for protection of participant privacy

Special requirement for foreign subawards:

- Access to all lab notebooks
- Access to all data (usually shared via cloud services)
- Access to all documentation
- Frequency: provide to prime recipient no less than once every year (aligned with timing of the RPPR)

Protecting Participant Privacy (Element 5)

What protections are needed?

Will sharing be restricted through controlled access?

Informed consent that includes broad data sharing.

De-identification of data

Tiered access (open, registered user, controlled)

Committee to oversee sharing of datasets and issue of DSUAs

Limitations defined in Data Sharing and Use Agreements

- When sharing would compromise the privacy or safety of participants
- When limitations are explicitly described in informed consent documents

Oversight and Monitoring (Element 6)

How the institution will ensure adherence to the plan

- Meetings to discuss status of DMS activities
- Ensure repository data matches the experiment
- Check that repository DOIs are included in publications

Who will be responsible for implementation and oversight

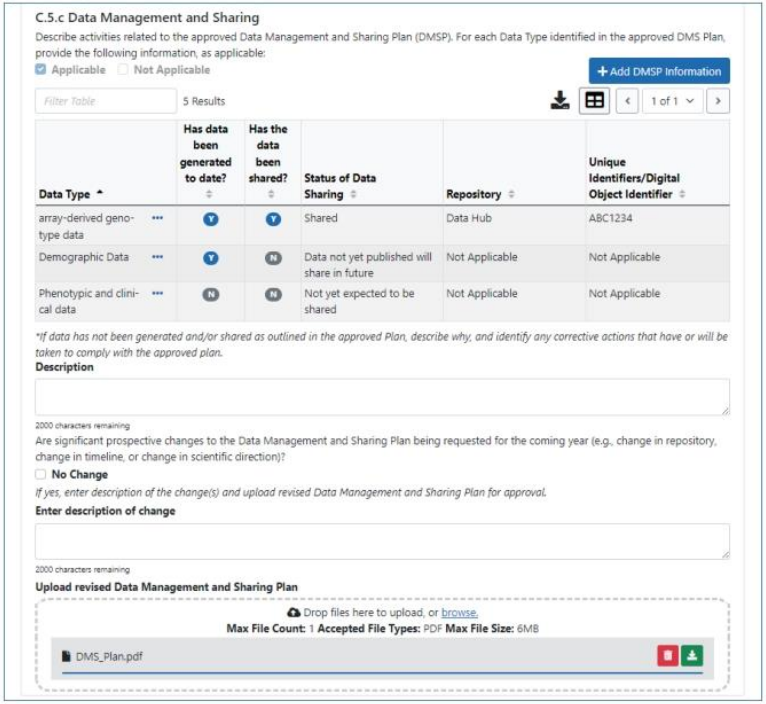
- The PI
- Data manager
- Office of Sponsored Projects (Compliance)

Frequency of oversight

- Quarterly group meetings
- Annual progress reports

Requirements for DMS reporting (Annual RPPR)

- What data has been generated?
- Status of sharing
- If shared, in what repository and under what DOI?
- If data has not been generated or shared, what corrective actions have or will be taken?
- If significant changes to the DMS Plan are anticipated, describe those changes and provide a revised DMS Plan



C.5.c Data Management and Sharing

Describe activities related to the approved Data Management and Sharing Plan (DMSP). For each Data Type identified in the approved DMS Plan, provide the following information, as applicable:

Applicable Not Applicable

+ Add DMSP Information

Filter Table 5 Results 1 of 1

Data Type	Has data been generated to date?	Has the data been shared?	Status of Data Sharing	Repository	Unique Identifiers/Digital Object Identifier
array-derived genotype data	Y	Y	Shared	Data Hub	ABC1234
Demographic Data	Y	N	Data not yet published will share in future	Not Applicable	Not Applicable
Phenotypic and clinical data	N	N	Not yet expected to be shared	Not Applicable	Not Applicable

If data has not been generated and/or shared as outlined in the approved Plan, describe why, and identify any corrective actions that have or will be taken to comply with the approved plan.

Description

2000 characters remaining

Are significant prospective changes to the Data Management and Sharing Plan being requested for the coming year (e.g., change in repository, change in timeline, or change in scientific direction)?

No Change

If yes, enter description of the change(s) and upload revised Data Management and Sharing Plan for approval.

Enter description of change

2000 characters remaining

Upload revised Data Management and Sharing Plan

Drop files here to upload, or browse.

Max File Count: 1 **Accepted File Types:** PDF **Max File Size:** 6MB

DMS_Plan.pdf

Challenges and Best Practices

Budgeting

Budgeting

The NIH maintains a site to assist with budgeting for DMSP costs. This includes allowable and unallowable costs.

Allowable	Unallowable
• Effort and resources to curate, format, and de-identify data, as well as preparing associate metadata • Development of support documentation • Infrastructure/storage costs • Repository fees	• Routine research costs • Direct charging for costs covered under F&A/double charging

Budgeting

In NICHD's review of DMSP budget justifications, they found:

- Justifications missing or not following SF424 instructions
- Zero or low budgets, even with robust DMS plans
- Disconnect between staff described in the DMS plan and the DMS budget
- High costs of data management, with little to no sharing

Budgeting

- Some PIs may be reluctant to include these costs. Encourage PIs to estimate the time they and research staff will expend meeting the collection, curation, and documentation steps detailed in the plan. *This is especially important if de-identification of human subjects data is needed.*
- Costs must occur during the award period; this can present a challenge to institutions if storage/repository fees will be needed after project end.

Budgeting

Other considerations for developing a budget for DMSP costs:

- A very thorough tool is available from the National Academies for determining numerous drivers of these costs .
- Some services such as de-identifying, developing metadata tags, and QA/QC of data sets can be provided by third parties. This cost be included in Other Directs, or could also help the PI estimate the costs for doing this work internally.

Trajectory of Regulations

Trajectory of Federal Rules and Regulations

NIH Monitoring

- Review of RPPR DMSP

NSF Public Access Policy 2.0

- January 2025 - Section 3: Increasing Equitable Access to Federally Funded Research Results
- January 2027 – Section 4: Ensuring Scientific and Research Integrity

OSTP Nelson Memo

- All federal research funding agencies have until December 31, 2024, to publish their updated public access policies.
- Agency plans and guidance can be found at science.gov
- Agencies must implement their public access plans by December 31, 2025

Discussion / Questions

Alison Krauss, akrauss@hcg.com
David Stewart, dstewart@uh.edu