Meeting NIH Data Management Plan Requirements

Jennifer Darragh
Senior Research Data Management Consultant
CDVS/Duke University Libraries

William Krenzer, Ph.D.
Research Project Manager
Duke Office of Scientific Integrity

Duke University Libraries

Center for Data and Visualization Sciences

askdata@duke.edu
https://library.duke.edu/data
@duke_data

CDVS provides multiple programs supporting data driven research

Data Science

GIS & Mapping

Data Visualization

Data Management

Data Management Plans

"A data management plan (also referred to as a data sharing plan) is a document that describes how you will treat your data during a project and what **happens with the data after the project ends**."

Michener, W. K. (2015). Ten Simple Rules for Creating a Good Data Management Plan. PLOS Computational Biology, 11(10), e1004525. https://doi.org/10.1371/journal.pcbi.1004525

NIH Policy Requirements Reminders

- Researchers planning to generate scientific data are **required** to submit a plan
- Upon approval of a plan, it is the expectation that the researchers **and institutions** implement data management and sharing practices as described
- 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
- The plan will become a Term and Condition of the Notice of Award

https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.ht

NIH Policy Reminders

- NIH strongly encourages researchers to plan for how data management and **sharing** will be addressed in the **informed consent** process
- NIH strongly encourages the use of **established repositories** to the extent possible for preserving and sharing scientific data

https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html

Allowable costs

- Curating data and developing support documentation
- Personnel costs required to perform the types of data management and sharing activities described in the supplemental information
- Fees for long-term data preservation and sharing - funds for these activities must be spent during the performance period, even for scientific data and metadata preserved and shared beyond the award period
- **NEW**: There no longer is a separate line item specifically for data management.

Unallowable Costs

- Facilities and Administrative costs
- Costs associated with the routine conduct of research
- Data access fees

NIH 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

https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-014.html

Terms to Know

- Metadata
- Standards
- Curation
- Repositories
- Digital preservation
- Data Sharing

Metadata & Standards

Structured information that is necessary to find, access, understand, and reuse the data.

Commonly used data formats

Many different standards

Curation

Data curation is “the active and ongoing management of data through its lifecycle of interest and usefulness to scholarship, science, and education, which includes appraisal and selection, representation, and organization of these data for **access and use over time**.”

Shreeves, S. L., & Cragin, M. H. (2008). Introduction: Institutional repositories: Current state and future. Library Trends, 57(2), 89–97. http://doi.org/10.1353/lib.0.0037

- Are my variables and values clearly labeled and defined?
- Have I adequately explained my methods (collection, analysis, cleaning)?
- Have I formatted my data in a way that it is interoperable across platforms?
- Have I logically arranged my files and explained how they relate?
- Have I commented my code to explain what it does?

Repositories

Repositories proliferate with different access options, funding models, etc.

- Funder-based repositories
- Institutional repositories (Duke RDR)
- Domain or Society repositories
- Generalist repositories (e.g., Dryad)

Preparing Data for Publishing - March 7 (Zoom) March 19 (In-person)

Digital Preservation

"*Digital information lasts forever—or five years, whichever comes first.*" - J. Rothenberg

- Format migrations, checksums and fixity
- Redundant backups and long-term access options
- Disaster and contingency planning
- Retention and transfer

A Tool to Help

https://dmptool.org/

Implementing Your DMP

- A DMP is a **living document**, revisit it often and submit any changes during regular reporting intervals
- Make the plan **actionable**
  - **Develop** a RACI Matrix (Roles & Responsibilities)
  - **Train** research team members on general data management best practices and standard operating **procedures**
  - **Write** data sharing into associated forms and protocols (i.e., consent forms, IRB protocols, contracts, etc.)

LabArchives ERN

Data AND Documentation files on electronic notebook pages

LabArchives is a web-based application developed to store, share, and manage your research data and documentation in a centrally accessible, electronic format.

LabArchives is provided to everyone at Duke for no cost.

REDCap

- Web-based interface for database build, data entry, reporting
- HIPAA-compliant and secure – data are stored on DHTS servers behind Duke firewall
- Automated data export (.csv, SPSS, SAS, Stata, R)
- Can be used for electronic consent (eConsent)
- Regular office hours

https://redcap.duke.edu/redcap/

Duke Supported Repositories

- Duke Research Data Repository
- Qualitative Data Repository
- ICPSR
- Vivli

Duke Research Data Repository

- Provides curation, open access, and preservation of research data produced by the Duke community
- Data management consultants work with researchers to organize their data, and create appropriate metadata to ensure the data meets **FAIR principles** (Findable, Accessible, Interoperable, and Reusable).
- 300 GB storage available for all Duke projects, for larger datasets, contact datamanagement@duke.edu
- Boilerplate language for data management plans, and advisement on data sharing language for consent forms and IRB protocols

Qualitative Data Repository (QDR)

- Provides curation, open and mitigated access, and preservation of **qualitative** research data.
- Duke paid membership (there may be a nominal charge depending on complexity of the potential deposit and help needed)
- Provides guidance on sharing human participant data (including templates if you plan to use QDR)
- Must be social science related or follow social science methods (interviews, transcripts, annotated images, etc.)

ICPSR

- Provides curation, open and mitigated access, and preservation of **quantitative** research data.
- Duke paid membership (there may be some additional charges so contact them early)
- Offers deductive disclosure risk assessment and de-identification assistance (for a fee - we can connect you)
- Offers virtual data enclave for restricted data use as well as other secure dissemination methods depending on sensitivity
- Hosts several thematic archives and funder-supported sub-repositories

Vivli

- Provides curation, mitigated access, and preservation of **clinical research data**.
- Duke is a member and there is no charge for deposit
- Offers de-identification assistance for a fee (otherwise data is expected to be de-identified)
- All data requests are subject to a review
- Offers a secure analysis environment (virtual enclave)
- We are working with the Medical Library and DCRI on setting up a workflow to help researchers deposit with Vivli

ON AND AFTER OCTOBER 5TH 2023:

- The Data Management and Sharing Plan
  - ~2 pages (more if needed) describing the data, how it will be managed, and how it will be shared
  - Investigators responsible named
  - Repositories named, how long data is available

- The Budget
  - ~~Line Item~~
  - **Justification**
    - <0.5 pages describing the costs associated with the DMSP

No longer required to add up all DMSP costs – include in appropriate category:

- Personnel
- Equipment
- Supplies
- Other expenses

ALLOWABLE COSTS FOR **DATA MANAGEMENT AND SHARING**

Generally, ….

Policy
Additional costs required to make data easily accessible and interpretable by others

ALLOWABLE COSTS FOR **DATA MANAGEMENT AND SHARING**

Policy States	This is....	This is NOT

Curating data
Developing supporting documentation
Preparing metadata to foster discoverability, interpretation, and reuse
Formatting data according to accepted community standards, or for transmission to and storage at a selected repository for long-term preservation and access
	• Time-costs (effort)
	  • Organizing data for team as well as upload
	  • Maintaining accurate libraries and README files
	  • Reviewing files for documentation and quality
	  • Time spent writing comments in analysis code
	  • Metadata (standard operating procedures, instructions on interpreting readouts of raw data
	  • Documentation of variable names and file names)
	  • Time spent adjusting file names, content, structure, or documentation for specific repositories
		Costs associated with acquiring data, such as from secondary data sources (such as Medicaid data)

ALLOWABLE COSTS FOR **DATA MANAGEMENT AND SHARING**

Policy States	This is....	This is NOT
De-identifying data	• This is sometimes included in repository fees or can be outsourced within Duke, typically requires an honest broker (ex: DOCR, Vivli) OR • Time costs (effort) of a team member	

ALLOWABLE COSTS FOR **DATA MANAGEMENT AND SHARING**

Policy States	This is....	This is NOT
Local data management considerations, such as unique and specialized information infrastructure necessary to provide local management and preservation (for example, before deposit into an established repository).	Costs associated with PACE (may include costs of de-id)	Virtual machine fees, cost of storage solutions for large data sets (these costs belong elsewhere in the budget)

ALLOWABLE COSTS FOR **DATA MANAGEMENT AND SHARING**

Policy States	This is....	This is NOT
Preserving and sharing data through established repositories, such as data deposit fees.	Time costs (effort) or fees associated with de-id or curation, fees associated with some repositories including de-id, curation, or long-term storage fees	Cost of holding the same data within the lab per Duke's data retention policy

BUDGET JUSTIFICATION FOR DMSP

Should be included even if $0 budgeted, particularly to describe effort associated with activities that are allowable

- **Personnel Costs (included in personnel line item)**
  - List who has effort related to data management and calendar months of effort
- **Repository Costs**
  - Name repository(ies) and the fee for them (often there isn't one)
- **Infrastructure Costs**
  - Name specialized infrastructure (PACE, etc) per year and total

Rare is the instance that we would expect to see $0 and no personnel effort for DMS allowable costs because every proposal will have curation efforts to share the data at minimum

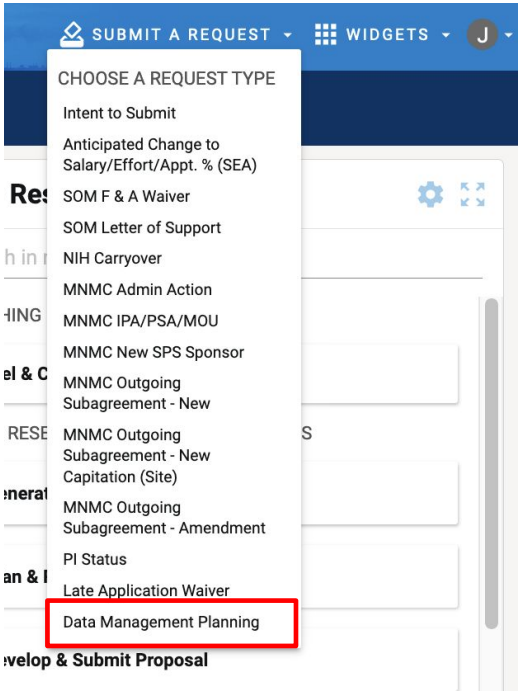
BUDGET JUSTIFICATION FOR DMSP

Example….

Cost Category	Specific Item	Cost
Specialized Environments	PACE	$2500 to set up, $500/years 2-4
Personnel	Curating Data	X% of data manager listed in key personnel effort in years 1-5
Personnel	Data De-Identification	$10,000 Vivli service OR $10,000 DOCR Honest Broker (fee if using PACE)
Repository	Vivli	$0 (free for Duke)

WHERE RESEARCHERS/RESEARCH ADMINISTRATORS FIND HELP

- ***Email*** researchdata@duke.edu with questions related to writing a data management and sharing plan OR building a budget
- ***Submit a request*** via MRH Submit a Request Data Management Planning
- ***Visit*** Office Hours on Thursdays at 9am, encourage your investigators to join offices hours
- ***Invite*** us to come and speak to your researchers on writing a data management and sharing plan and budgeting appropriately
- ***Explore*** the NIH Data Management FAQs and Budgeting for Data Management and Sharing page



RQMP YEAR 5 ATTESTATION

Resources

- Duke Data Management Guidance Document: For those without a funder requirement or to learn more tips/considerations (also available through the DMPTool)
- NIH Sponsored Repositories: List of repositories created by NIH for specific research areas or data types
- ICPSR Recommended Informed Consent Language for Data Sharing
- SecureIT: An OIT resource for identifying Duke secure storage solutions