NIH Data Management and Sharing (DMS) Policy

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

Agenda

1. Overview of the NIH DMS Policy
2. Writing a DMS Plan
3. Budgeting for a DMS Plan
4. Diving into DMS Budgeting
5. Considerations for Human Subjects
6. Compliance
7. Resources

Overview and Impact

Stephen Dewhurst, PhD
Interim Vice President for Research

NIH DMS Policy: Goals

Goal: To make the results and outputs of NIH-funded research available to the public through effective and efficient data management and data sharing practices.

Why? Data sharing enables researchers to:

- Rigorously test the validity of research findings,
- Strengthen analyses through combined datasets,
- Reuse hard-to-generate data,

Good data management practices are foundational to effective data sharing.

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

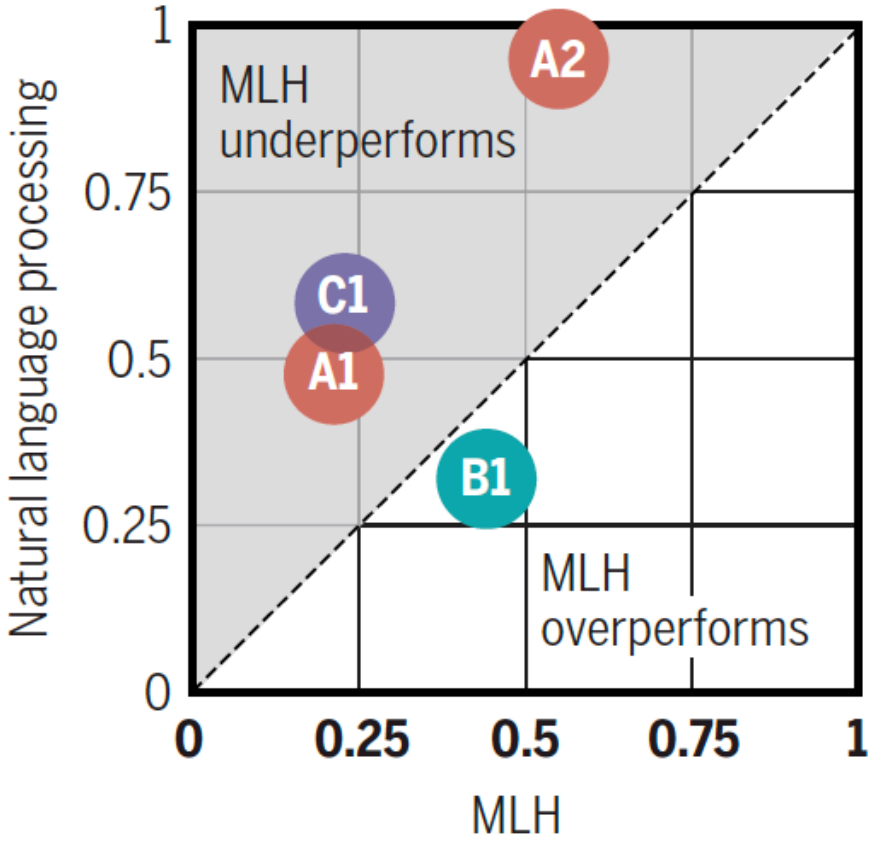
Example of Why This Matters: Reproducibility in Machine Learning for Health (MLH)

Evaluation metrics

A **Technical reproducibility**
- 1 Code available
- 2 Public dataset

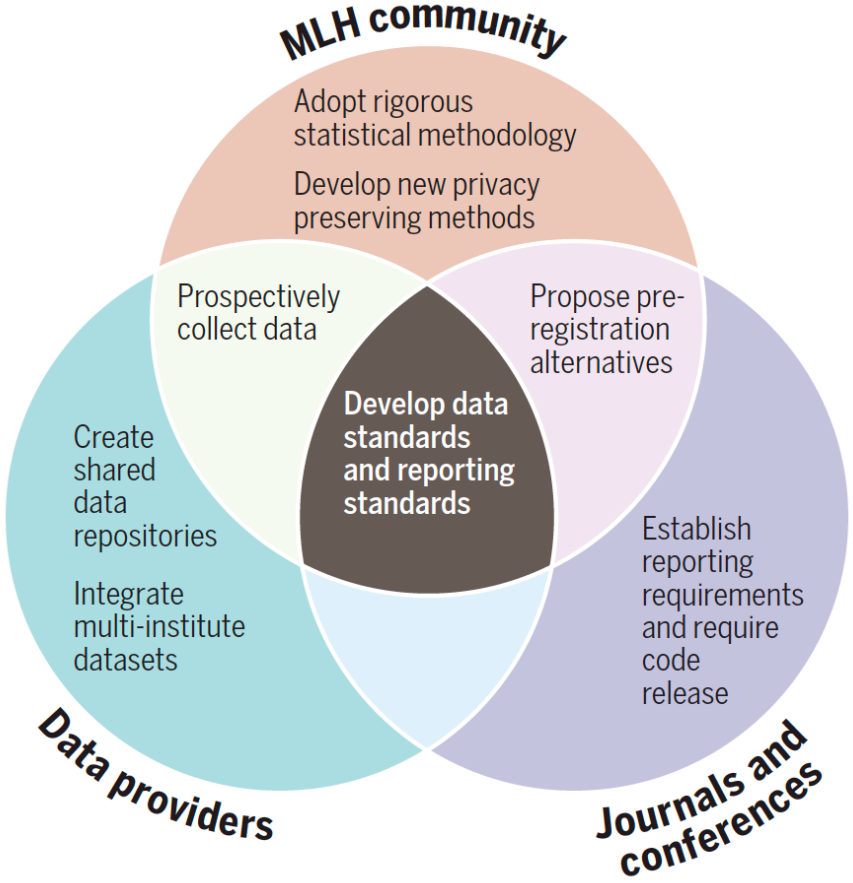
B **Statistical reproducibility**
- 1 Variance reported

C **Conceptual reproducibility (replicability)**
- 1 Multiple datasets



McDermott, MBA et al. Sci. Trans. Med. 13, eabb1655 (2021); https://www.science.org/doi/10.1126/scitranslmed.abb1655

Improving Reproducibility in MLH



McDermott, MBA et al. Sci. Trans. Med. 13, eabb1655 (2021); https://www.science.org/doi/10.1126/scitranslmed.abb1655

NIH DMS Policy: Scope

The DMS Policy takes effect 1/25/2023:

- **Applies to ALL new NIH research grants**, but not to training/fellowship, infrastructure & other non-research awards

Requires:

- Submission of a DMS Plan outlining how scientific data and any accompanying metadata will be managed and shared, *taking into account any potential restrictions or limitations*.
- Compliance with the plan.

https://sharing.nih.gov/data-management-and-sharing-policy/about-data-management-and-sharing-policies

NIH DMS Policy: Details

- Shared scientific data made accessible ASAP – no later than the time of an associated publication or end of performance period, whichever comes first.
- DMS plans will be reviewed by NIH program staff (not peers).
- Peer reviewers will only consider a proposal if data management budget is acceptable.
- *Costs for DMS can be included in the budget (discussed later in this presentation)*

NIH DMS Policy: Definition of Data

Definition of Scientific Data: The 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.

https://sharing.nih.gov/data-management-and-sharing-policy/about-data-management-and-sharing-policies

Why?

Open sharing of research data is essential not only to ensuring ongoing public support for the work we do (and that the public pays for), but also to ensuring that our work as researchers is of the **greatest benefit possible to the people it is intended to serve**.

Writing a NIH Data Management Sharing Plan

Daniel Castillo, MLIS, EdD
Head of Scholarly Communications and Research Initiatives
University of Rochester Medical Center, Edward G. Miner Library

Agenda

- Scientific Data and Data Sharing Plans
- NIH Six Required Elements
- Library Resources and Services

DMS Policy

- Set of requirements and guidelines related to the management and sharing of research data
- Applies to work funded wholly or partially by the National Institutes of Health (NIH)
- **January 25th, 2023**

DMSP

A Data Management and Sharing Plan (DMSP) is a **two-page** document included in applicable NIH proposals that prospectively outlines how a research team intends to **manage, preserve, and share the scientific data** associated with the proposed work.

What is Scientific Data?

It includes...

- Any data 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

It does not include...

- Laboratory notebooks
- Preliminary analyses
- Compared case report forms
- Drafts of scientific papers
- Plans for future research
- Peer reviews
- Communications with colleagues
- Physical objects such as laboratory specimens

NIH. 2022. Research covered under the data management & sharing policy.

Justifiable Reasons for Limiting Sharing of Data

It is understood that some data cannot be shared openly or even at all, due to certain ethical, legal, and social issues.

Reasons include...

- Informed consent will not permit or will limit the scope or extent of sharing and future research use
- Existing consent prohibits sharing or limits the scope or extent of sharing and future research use
- Privacy or safety of research participants would be compromised and protective measures such as de-identification and Certificates of Confidentiality would be insufficient.
- Explicit federal, state, local, or Tribal law, regulation, or policy prohibits disclosure
- Restrictions imposed by existing or anticipated agreements
- Datasets cannot practically be digitized with reasonable efforts

NIH. 2022. Frequently asked questions (FAQs).

However, you must include a rationale for why scientific data will not be shared.

When Does it Need to be Shared?

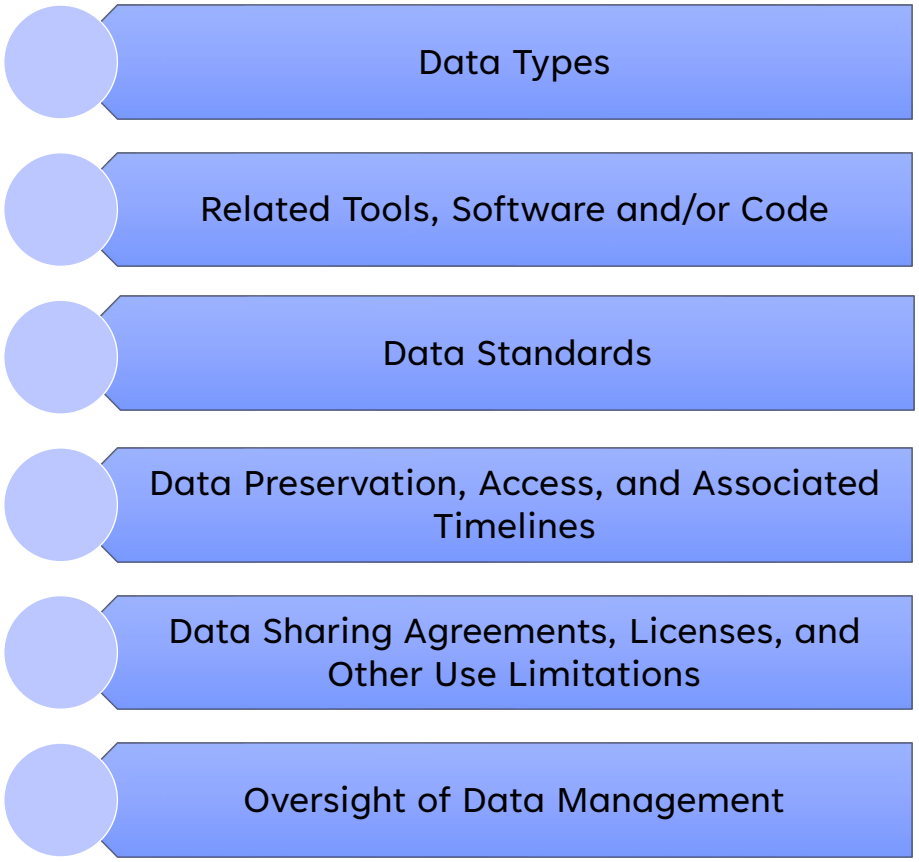
According to the DMS Policy, scientific data needs to be shared by the **earlier** of two timepoints:

The time of an associated publication

OR

The end of the performance period (unless the grant enters into a no-cost extension)

DMSPs for NIH Require 6 Elements



NIH. (2020). *Supplemental information to the NIH policy for data management and sharing: Elements of an NIH data management and sharing plan.*

Data Type

- Summary of the **types and estimated amount** of scientific data.
- Descriptions may include data **modality**, level of **aggregation**, and degree of data **processing.**
- Description of **which scientific data** from the project will be **preserved and shared**. Remember, NIH does not anticipate for all scientific data to be shared. You must **provide a rationale** for these decisions.
- A brief listing of the **metadata**, other relevant data, and any associated **documentation** that will be made accessible to ensure the scientific data is understandable and interpretable.

NIH. (2020). *Supplemental information to the NIH policy for data management and sharing: Elements of an NIH data management and sharing plan.*

Data Type

This project will produce imaging data generated/obtained from ultrasonography. Data will be collected from 10 of research participants/specimens/experiments, generating 1 datasets totaling approximately half a GB [amount of data] in size. The following data files will be used or produced in the course of the project: DICOM. Raw data will be transformed by filtering and enhancing the images and the subsequent processed dataset used for statistical analysis. To protect research participant identities, individual data will be made available for sharing.

Adapted from NIH. (2022). 2023 NIH Data Management and Sharing Policy.

Data Type

Based on ethical, legal, and technical considerations, the following data produced in the course of the project will be preserved and shared: All data produced in the course of the project will be preserved and shared.

To facilitate interpretation of the data, metadata documentation will be shared and associated with the relevant datasets.

Adapted from NIH. (2022). 2023 NIH Data Management and Sharing Policy.

Related Tools, Software and/or Code

- List the names of **tool(s) and software** needed to access or manipulate shared scientific data to support replication or reuse.
- If necessary, specify **how these tools can be accessed**, and if the tools are likely to remain available.

NIH. (2020). *Supplemental information to the NIH policy for data management and sharing: Elements of an NIH data management and sharing plan.*

Related Tools, Software and/or Code

Imaging data will be made available in DICOM format, which requires the use of MicroDicom to be accessed and manipulated. The MicroDicom tool, which can be used to view the images is available free of charge through https://www.microdicom.com/.

Adapted from NIH. (2022). 2023 NIH Data Management and Sharing Policy.

Standards

- Indication of what **standards** will be applied to the scientific data and associated metadata:
  - Data formats
  - Data dictionaries
  - Data identifiers
  - Definitions
  - Unique Identifiers
  - Other data documentation
- Plan may indicate that **no consensus** data standards exist for the scientific data and metadata

NIH. (2020). *Supplemental information to the NIH policy for data management and sharing: Elements of an NIH data management and sharing plan.*

Standards

To facilitate their efficient use, all of our data and materials will be structured and described using the following standards: Network Common Data Format.

Alternatively

Formal standards for DICOM data have not yet been widely adopted. However, our data and other materials will be structured and described according to best practices.

Adapted from NIH. (2022). 2023 NIH Data Management and Sharing Policy.

Data Preservation, Access, and Associated Timelines

- Name of **repository(ies)** where data will be archived
- How the scientific data will be findable and identifiable (i.e., **persistent unique identifier**)
- **When** the scientific data will be made available and for **how long**
- Identify any differences in timelines for different subsets of scientific data to be shared
- Scientific data should be shared as soon as possible, and **no later** than time of an **associated publication** or end of the **performance period**

NIH. (2020). *Supplemental information to the NIH policy for data management and sharing: Elements of an NIH data management and sharing plan.*

Data Preservation, Access, and Associated Timelines

All dataset(s) that can be shared will be deposited in Figshare.

The Fighsare Repository provides metadata, DOI persistent identifiers, and long-term access. This repository is supported by University of Rochester and dataset(s) are available under a Creative Commons CC0 license.

Data will be made available as soon as possible or at the time of associated publication, whichever comes first.

Adapted from NIH. (2022). 2023 NIH Data Management and Sharing Policy.

Access, Distribution, or Reuse Considerations

- Describe any **factors** affecting subsequent access, distribution or reuse of scientific data related to:
  - **Informed consent**
  - Privacy and confidentiality protections consistent with applicable federal, Tribal, state, and local **laws, regulations, and policies**
    - De-identification
    - Certificates of Confidentiality
    - Other Protective Measures
  - Note whether **access** to scientific data derived from humans **will be controlled** (i.e., made available by a data repository only after approval).
  - Note any other considerations that may limit the extent of data sharing

NIH. (2020). *Supplemental information to the NIH policy for data management and sharing: Elements of an NIH data management and sharing plan.*

Access, Distribution, or Reuse Considerations

There are no factors that will affect subsequent access, distribution or reuse of the scientific data. All data will be de-identified and we will not place any access controls over the data.

Oversight of Data Management and Sharing

- Indicate how compliance in the Plan will be...
  - Monitored and managed
  - Frequency of oversight
  - By whom (e.g., titles, roles)

NIH. (2020). *Supplemental information to the NIH policy for data management and sharing: Elements of an NIH data management and sharing plan.*

Oversight of Data Management and Sharing

The PI on this project will be responsible for all aspects of data management and sharing, including collecting, analyzing, and describing the data, as well as uploading the data to Figshare.

Alternatively

Daniel Castillo, the lab manager for the radiology imaging lab, will be responsible for all data management and sharing, including monitoring adherence to NIH DMS policy. Data will be collected and analyzed by Ehsan Moghadam, post-doc, and Heather Owen, post-doc. There is a budget request for a part-time employee who will be assisting with de-identifying collected data and developing the metadata for sharing.

Adapted from Samuel J. Wood Library. (2022). Oversight of Data Management and Sharing.

New UR Libraries DMS Service

Templates Created by the NIH

Template Name	Download	Organization name	Last Updated	Funder Links	Create a new plan	Sample Plans (if available)
NIH-GDS: Genomic Data Sharing		National Institutes of Health (nih.gov)	09-20-2022	NIH Genomic Data Sharing Policy [PDF] NIH Public Access Plan [PDF] NIH GDS Policies and Supplemental Information Data Submission and Release Expectations Developing Genomic Data Sharing Plans	Requires login	
NIH-GEN DMSP (Forthcoming 2023)		National Institutes of Health (nih.gov)	11-14-2022	Final NIH Policy for Data Management and Sharing Research Covered Under the Data Management & Sharing Policy	Requires login	Elements of an NIH Data Management and Sharing Plan Selecting a Repository for Data Resulting from NIH-Supported Research Allowable Costs for Data Management and Sharing
NIH-GEN: Generic (Current until 2023)		National Institutes of Health (nih.gov)	10-25-2021	NIH Data Sharing Policy and Implementation Guidance NIH Public Access Plan [PDF] NIH Data Sharing Policies by Program	Requires login	NIH Examples of Data Sharing Plans

DMPTool. 2022. Funder requirements. California Digital Library.

Data Management Guide

https://libguides.urmc.rochester.edu/datamanagement

- Detailed outline including six required elements
- Templates
- Contact information

Data Management Librarians

Ehsan Moghadam
University of Rochester
Medical Center
Edward G, Miner Library

Ehsan_Moghadam@URMC.Rochester.edu

Heather Owen
University of Rochester.
River Campus Libraries

howen@library.rochester.edu

dataservices@library.rochester.edu

Thank you!

Budgeting for a DMS Plan

Brenda Kavanaugh
Associate Director and NIH Liaison
Office of Research & Project Administration

What about DMSP Costs?

Funds may be requested for data management and sharing in the budget and budget justification

Allowable Costs

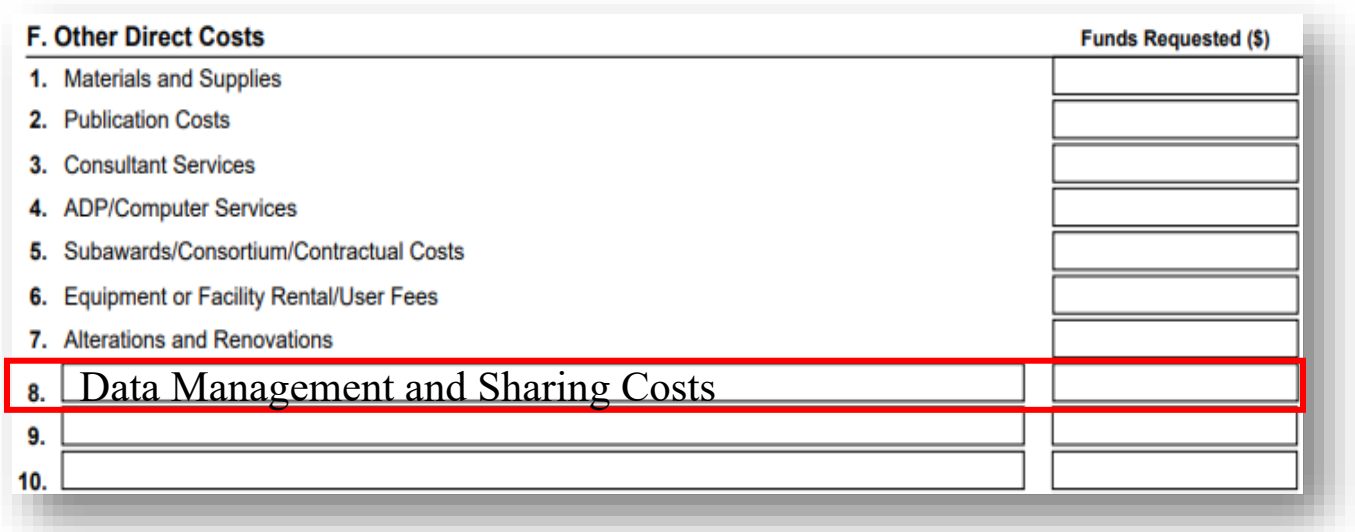
- Curating data (including applicable personnel costs)
- Developing supporting documentation
- Formatting data
- De-identifying data
- Preparing metadata
- Unique and specialized information infrastructure
  Preserving and sharing data
- Multiple repositories
- Must be incurred during the performance period

Unallowable Costs

- Infrastructure costs that are included in institutional F & A
- Costs associated with the routine conduct of research
- Costs that are double/inconsistently charged as both direct and indirect

How to request

- Line item in the R & R detailed budget (Section F. Other Direct Costs)
  - Do not combine with other costs, if none enter $0

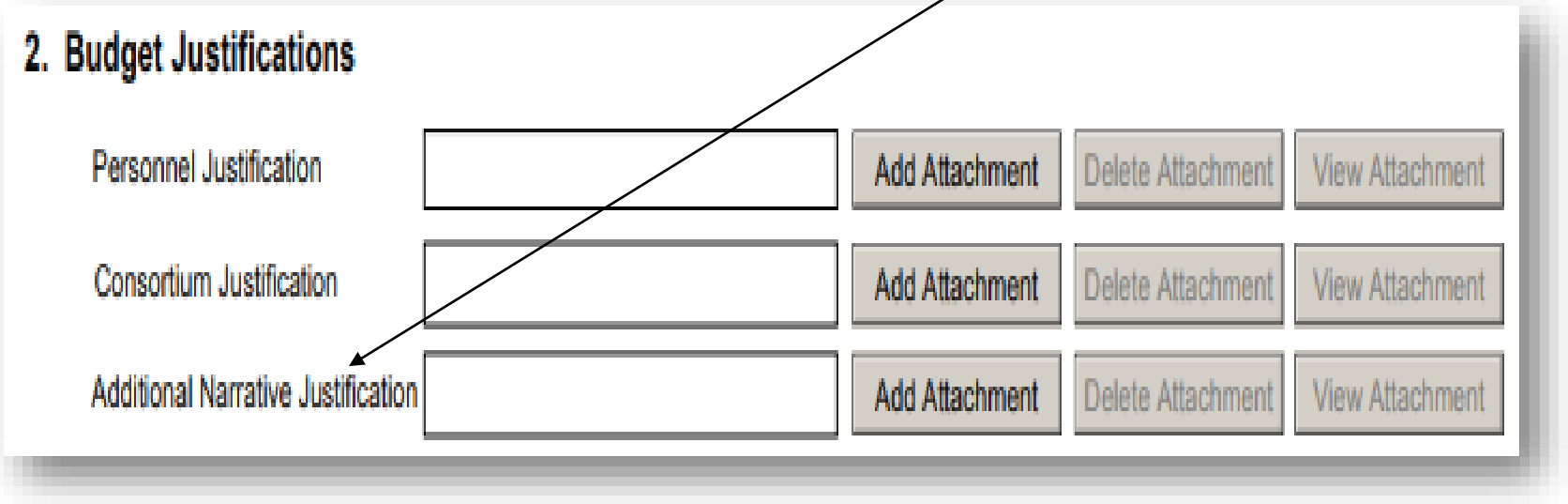


F. Other Direct Costs	Funds Requested ($)
1. Materials and Supplies	
2. Publication Costs	
3. Consultant Services	
4. ADP/Computer Services	
5. Subawards/Consortium/Contractual Costs	
6. Equipment or Facility Rental/User Fees	
7. Alterations and Renovations	
8. Data Management and Sharing Costs	
9.	
10.	

- Budget Justification -Brief summary of the DMS Plan and describe the requested costs

How to request

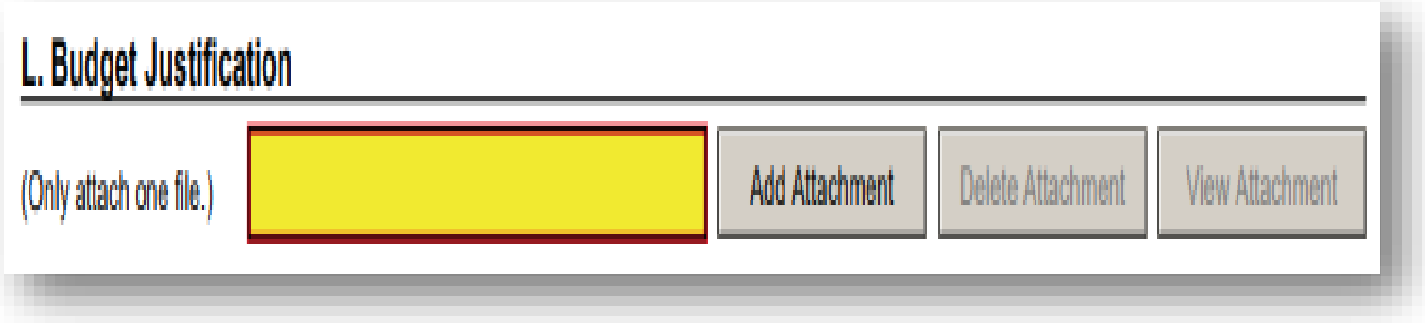
- Modular budget – additional narrative justification



Justifying the costs

- Brief summary of type/amount of data preserved/shared
- Name of the established repository(ies) to be used, and general cost categories
- Recommended length-no more than half a page
- Label as "Data Management and Sharing Justification“
- Genomic Data Sharing - include requested costs for genomic data management and sharing

Justifying the costs



Budget Assessment

- Peer Reviewers may provide comments on the reasonableness
- Will not impact the score

Other Considerations

- Potential for additional costs if IC requests revision to the DMSP
- Some data may require storage beyond the period of performance
- Consider subrecipient costs

Diving into DMS Budgeting

Martin Zand, MD, PhD
Sr. Associate Dean, Clinical Research
School of Medicine and Dentistry
Co-Director
Clinical & Translational Science Institute

NIH DATA SHARING PLAN – DATA AND COST ESTIMATES

- Data must be stored for 5 years the grant ends
  - Data storage costs cumulative (store all data generated cumulatively)
  - Must add costs for "tail" over lifetime of the grant
- Budget for a lab data manager
  - You are managing the data anyway, need to use this in budget justification
- If you have complex data, use a robust data management system
  - Electronic Lab Notebook OK for simple data (NOT PHI, Images)
  - System needs to store metadata (instrument settings, etc.)
  - Know the metadata required by an NIH repository for deposit
  - Robust systems (e.g. BioLaboratory Server – CTSI, PAX, etc.) have better organization
- Do not underbudget personnel
  - Custom data services (Lab database, de-identification, etc.

NIH DATA SHARING PLAN – ALLOWABLE COSTS FOR BUDGETS

- Curating data
- Developing supporting documentation
- Formatting data according to accepted community standards, or for transmission to and storage at a selected repository for long-term preservation and access
- De-identifying data
- Preparing metadata to foster discoverability, interpretation, and reuse
- 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).
- Preserving and sharing data through each established repositories, such as data deposit fees.
- All allowable costs submitted in budget requests must be incurred during the performance period, even for scientific data and metadata preserved and shared beyond the award period.

NIH DATA SHARING PLAN – TEMPLATES

	A	B	C	D	E	M	N	O	P	Q	R	S	T	U	V	W	X	Y
1	**Local Storage**					**Cumulative Storage Needs Per Year (2)**						**Storage and Backup Cost Per Year**						
2	**Data Type**	**Size (GB)**	**Repository Cost/GB (1)**	**Local Cost/GB**	**Total Cost/GB**	**Y1**	**Y2**	**Y3**	**Y4**	**Y5**	**Y6-10**	**Y1**	**Y2**	**Y3**	**Y4**	**Y5**	**Y6-10**	**Total**
3						0	0	0	0	0	0	$ -	$ -	$ -	$ -	$ -	$ -	$ -
4	EHR Dataset	**100**	$ 0.20	$ 0.25	$ 0.45	100	100	200	200	200	200	$ 45	$ 45	$ 90	$ 90	$ 90	$ 450	$ 810
5					$ -	0	0	0	0	0	0	$ -	$ -	$ -	$ -	$ -	$ -	$ -
6	CODEX	10	$ 0.20	$ 0.25	$ 0.45	0	0	0	0	0	0	$ -	$ -	$ -	$ -	$ -	$ -	$ -
7	scRNAseq	0.75	$ 0.20	$ 0.25	$ 0.45	0	0	0	0	0	0	$ -	$ -	$ -	$ -	$ -	$ -	$ -
8	Bulk RNAseq	5	$ 0.20	$ 0.25	$ 0.45	100	300	800	1300	1800	1820	$ 45	$ 135	$ 360	$ 585	$ 810	$ 4,095	$ 6,030
9	Exome	15	$ 0.20	$ 0.25	$ 0.45	0	0	0	0	0	0	$ -	$ -	$ -	$ -	$ -	$ -	$ -
10	Miseq (microbiome)	0.5	$ 0.20	$ 0.25	$ 0.45	0	0	0	0	0	0	$ -	$ -	$ -	$ -	$ -	$ -	$ -
11	Whole genome 50x coverage	200	$ 0.20	$ 0.25	$ 0.45	0	2000	4000	6000	8000	8000	$ -	$ 900	$ 1,800	$ 2,700	$ 3,600	$ 18,000	$ 27,000
12	VCF variant call	1	$ 0.20	$ 0.25	$ 0.45	0	10	30	50	70	70	$ -	$ 5	$ 14	$ 23	$ 32	$ 158	$ 230
13	citeSeq	120	$ 0.20	$ 0.25	$ 0.45	0	0	0	0	0	0	$ -	$ -	$ -	$ -	$ -	$ -	$ -
14			$ 0.20	$ 0.25	$ 0.45	0	0	0	0	0	0	$ -	$ -	$ -	$ -	$ -	$ -	$ -
15	MRI (per image series)	0.5	$ 0.20	$ 0.25	$ 0.45	0	0	0	0	0	0	$ -	$ -	$ -	$ -	$ -	$ -	$ -
16	CT (per image series)	2	$ 0.20	$ 0.25	$ 0.45	0	0	0	0	0	0	$ -	$ -	$ -	$ -	$ -	$ -	$ -
17	Other DICOM (you will need to specify)	2	$ 0.20	$ 0.25	$ 0.45	0	0	0	0	0	0	$ -	$ -	$ -	$ -	$ -	$ -	$ -
18	Microscopy	5	$ 0.20	$ 0.25	$ 0.45	0	0	0	0	0	0	$ -	$ -	$ -	$ -	$ -	$ -	$ -
19			$ 0.20	$ 0.25	$ 0.45	0	0	0	0	0	0	$ -	$ -	$ -	$ -	$ -	$ -	$ -
20	Flow Cytometry	0.1	$ 0.20	$ 0.25	$ 0.45	5	13	21	29	37	37	$ 2	$ 6	$ 9	$ 13	$ 17	$ 83	$ 131
21	ELISPOT/Luminex	0.0001	$ 0.20	$ 0.25	$ 0.45	0.003	0.006	0.012	0.018	0.021	0.021	$ 0	$ 0	$ 0	$ 0	$ 0	$ 0	$ 0
22	Gel Images	0.0001	$ 0.20	$ 0.25	$ 0.45	0	0	0	0	0	0	$ -	$ -	$ -	$ -	$ -	$ -	$ -
23																		
24	**Total**					**105**	**2323**	**4851**	**7379**	**9907**	**9927**	**$ 47**	**$ 1,045**	**$ 2,183**	**$ 3,321**	**$ 4,458**	**$ 22,336**	**$ 33,390**
25																		
26						**Hours/year (2)**							**Data Handling Cost Per Year**						
27	**Data Services**				**Cost**	**Y1**	**Y2**	**Y3**	**Y4**	**Y5**	**Y6**	**Y1**	**Y2**	**Y3**	**Y4**	**Y5**	**Y6 (4)**	**Total**
28	Laboratory Data Manager (3)			Hourly	$ 25.00	208	416	416	416	416	104	$ 5,200	$ 10,400	$ 10,400	$ 10,400	$ 10,400	$ 2,600	**$ 49,400**
29	Data Life Cycle Management (fee for service database systems)			Hourly	$ 85.00	50	50	50	70	80	30	$ 4,250	$ 4,250	$ 4,250	$ 5,950	$ 6,800	$ 2,550	**$ 28,050**
30	Electronic Laboratory Notebook (Annual Cost}			Instance	$ 500.00	1	1	1	1	1	1	$ 500	$ 500	$ 500	$ 500	$ 500	$ 500	**$ 3,000**
31	Data Deidentification			Hourly	$ 85.00	30	20	30	30	30	-	$ 2,550	$ 1,700	$ 2,550	$ 2,550	$ 2,550	$ -	**$ 11,900**
32	Data Wrangling and Metadata Support for NIH data repositories			Hourly	$ 85.00	-	10	15	20	20	10	$ -	$ 850	$ 1,275	$ 1,700	$ 1,700	$ 850	**$ 6,375**
33	Data Deposit Fee (per deposit)			Instance	$ 300	-	1	3	3	3	-	$ -	$ 300	$ 900	$ 900	$ 900	$ -	**$ 3,000**
34																		
35	**Total**					**289**	**498**	**515**	**540**	**550**	**145**	**$ 12,500**	**$ 18,000**	**$ 19,875**	**$ 22,000**	**$ 22,850**	**$ 6,500**	**$ 101,725**
36																		
37	**(1) Figshare cost quoted. If using NIH repositories for data, cost is $0**																	
38	**(2) Tail for data storage is 5 years, and you will need to build this into your annual estimates (add 1/5 of tail to each year)**											**$ 12,547**	**$ 19,045**	**$ 22,058**	**$ 25,321**	**$ 27,308**	**$ 28,836**	**$ 135,115**
39	**(3) Estimate uses $50K/year salary and adds benefits for lab member**																	
40	**(4) Suggest you build 1 more year (NCE) into budget for tail metadata management**																	
41	**(5) You need to store all the data you accumulate during the project**																	

NIH DATA SHARING PLAN – TEMPLATES

- Use Cost Template For Budget Estimates
  - You MUST confirm personnel costs and hours needed if using other services (CIRC, CTSI, Biostatistics, etc.)
  - No different than any other budget for mice, reagents, etc.
  - Storage requirements add up, and we will be monitoring to meet storage needs.
  - If you have very large data sets (>1 TB), please contact your IT group (UIT, RAIT, CIRC) for storage estimates and plans

- *NEVER* put in a modular budget if not required by the award type
  - You limit yourself in the long run, and you will need to do the work anyway
  - More $$ for your research

- You will need to describe deidentification for Personal Health Information, if used
  - Should correspond to your budget
  - Get a consultation if this applies to you

Considerations for Human Subjects

Kelley O'Donoghue
Associate Vice President
Office for Human Subject Protection

Sharing Data Involving Human Subjects

- **Consent form** - include a description of any future sharing and how data will be used.

- **Assess the privacy and confidentiality protections needed for sharing data** - Unless a subject consents to sharing identifiable data, it should be shared only in a de-identified format.

- **Consider justifiable limitations to sharing data**
  - Data to be used only as described in the consent form
  - Consider when sharing may compromise subject privacy or safety, even if de-identified.

For more information, see NIH's supplement to the DMS policy Protecting Privacy When Sharing Human Research Participant Data and the webpage Sharing Data from Human Participants.

Compliance

Gunta Liders
Associate Vice President
Office of Research & Project Administration

Compliance

- The DMSP will become a term and condition of the award
- Compliance with the plan will be determined by the NIH IC
- NIH staff will monitor compliance with approved DMS plans during the annual RPPR process

Compliance

Failure to comply could result in additional terms or conditions on the grant or even termination of a grant – and could impact an investigator's ability to obtain future NIH funding.

Resources

Libby Reitz
Director for Research
Office of the Vice President for Research

Resources

- Office of the VP for Research website
  - Policy overview
  - How do I… / FAQs
  - Examples of DMS plans and budgets
  - Central link to other resources
- Miner Library and River Campus Libraries
  - Writing a DMS plan
  - Budgeting for data storage in Figshare

Resources

- Office for Human Subject Protection
  - Considerations for human subjects
- CIRC
  - Code storage (Gitlab)
  - Planning and storing large (>1TB) datasets
- Research & Academic IT
  - Planning and storage large (>1TB) datasets
- NIH website
  - Policy details, other resources, etc.

Thank You