

How to prepare a budget for your data management plan

DataWorks! Salon December 15, 2022

Julie Goldman, Research Data Services Librarian, Harvard Library

2023 NIH Data Management and Sharing Policy

Submission of a Data Management and Sharing Plan (DMSP)

Request funds for data management and sharing



Compliance with the awardee's plan as approved by NIH ICO

Budgeting for Data Management & Sharing

- Reasonable, allowable costs:
 - Curating data
 - Developing supporting documentation
 - Formatting data according to community standards or repository
 - De-identifying data
 - Preparing metadata to foster discoverability, interpretation, and reuse
 - Local data management considerations
 - Preserving and sharing data through established repositories
- All allowable costs submitted in budget requests must be incurred during the performance period

Budgeting for Data Management & Sharing

- Unallowable costs:
 - Infrastructure costs that are included in institutional overhead
 - Facilities and Administrative costs: facilities operation and maintenance costs, depreciation, and administrative expenses
 - The organization is responsible for the management and accounting of costs consistently
 - Costs associated with the routine conduct of research
 - Costs associated with collecting or gaining access to research data
 - Costs that are double charged or inconsistently charged as both direct and indirect costs

Budget Pages

- R&R Detailed Budget
 - Requests for direct costs to support the DMS activities must be reflected as a single line item under Other Costs titled “Data Management and Sharing Costs”
 - NIH instructs that personnel costs related to DMS activities should be included in this Other Costs line item rather than the Personnel section
 - It is suggested that all effort for the PI(s) and Senior/Key Personnel must remain in the Personnel section
- Modular Budget
 - Costs associated with DMS activities referenced in the Additional Narrative Justification

R&R Budget Justification/ Additional Narrative Justification

- A summary of the DMSP and a description of the associated requested costs
- Brief summary of type and amount of scientific data to be preserved and shared and the name of the established repository(ies) where they will be preserved and shared
- Indicate general cost categories such as curating data and developing documentation, preserving and sharing data through established repositories, etc., including an amount for each category and a brief explanation
- The recommended length is no more than half a page

Data Management and Sharing Plan as an Indicator of Allowable Costs

1. **Data Types:** Costs for data collection, acquisition, storage
2. **Related Tools, Software, Code:** Costs for data access, security
3. **Common Data Standards:** Costs for data documentation
4. **Data Preservation, Access, Timelines:** Costs for data preservation
5. **Access, Distribution, Reuse:** Costs for data availability, reuse
6. **Oversight of Data Management:** Costs for operationalizing

Budget Categories

Category: Preparing

Activity: Make a Data Management Plan

- Make a DMP before you start creating data to help make decisions about managing your data.
- Check if there is a department within your organization to support data management planning.

Costs: 2 hrs to 2 days, depending on the complexity of your project and how much you have prepared ahead of time

*** Start early! There may be some standard language you can develop for use across projects*

Category: Data Collection

Activity: Acquiring External datasets

- Do you plan to use existing (commercial or open) data?
- Your library may be able to help you acquire a license to a crucial database.
- In research data repositories, data can be available at no or low costs.

Costs: No cost to \$200-300K or more, depending on the repository and whether the data is obtained from proprietary sources

*** Confirm if the dataset can be capitalized or if it should be included in the MTDC base for F&A recovery*

Budget Categories

Category: Data Documentation

Activity: Data description and metadata

- Is data in a spreadsheet or database clearly described with variable and value labels, code descriptions, etc.?
- Are validated questionnaires and standard coding used?

Costs: 4 hrs per single experiment (120 measurements) filling in 60 required metadata fields, with assistance of a data manager (salary range \$80,000-\$100,000)

*** If data description is carried out as part of data creation, data input has lower or no additional cost*

Category: Data Sharing & Reuse

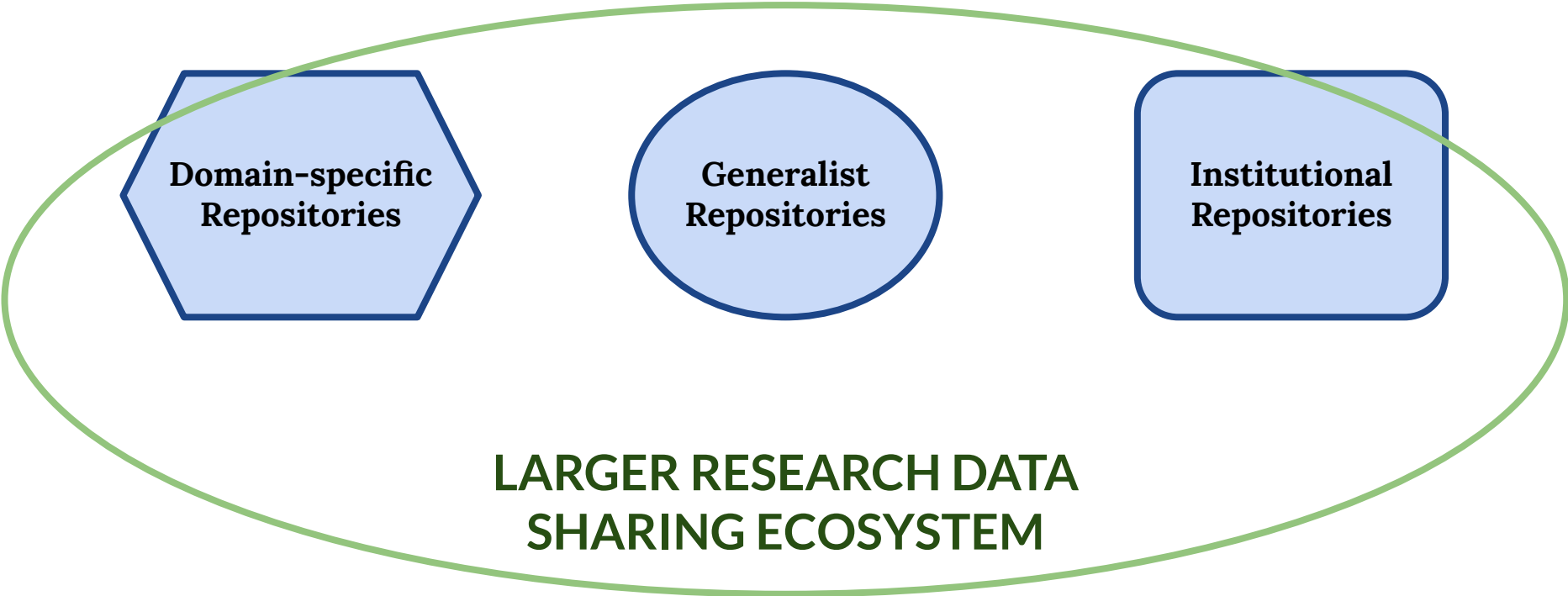
Activity: Data sharing

- Will you deposit data in a repository?
- Find out what the cost are of data deposit and/or long-term storage per year cost in time and effort needed to prepare data for sharing & preservation

Costs: Dryad charges \$120 per data submission to cover costs of curating and preserving data, plus \$50 for each additional 10GB (over 50GB)

***Completing a data deposit form (e.g. Zenodo a free repository) may take 15 min to 4 hrs*

Data Repositories



The diagram features a large, light green oval that encloses the central content. At the top center of the oval is the title 'Data Repositories'. Below the title, three light blue shapes are arranged horizontally: a hexagon on the left, a circle in the middle, and a rounded rectangle on the right. Each shape contains text describing a type of repository. At the bottom center of the oval, the text 'LARGER RESEARCH DATA SHARING ECOSYSTEM' is written in a bold, dark green font.

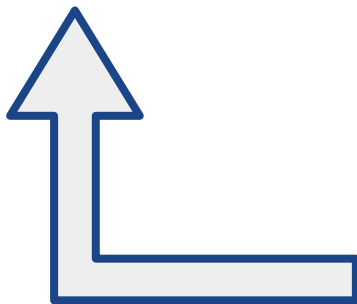
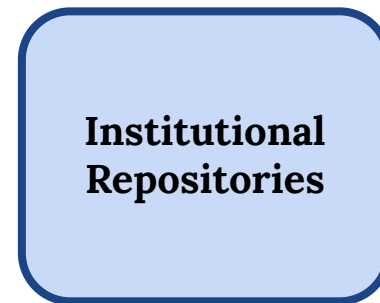
**Domain-specific
Repositories**

**Generalist
Repositories**

**Institutional
Repositories**

**LARGER RESEARCH DATA
SHARING ECOSYSTEM**

Disciplinary Repositories



NIMH
National Institute
of Mental Health

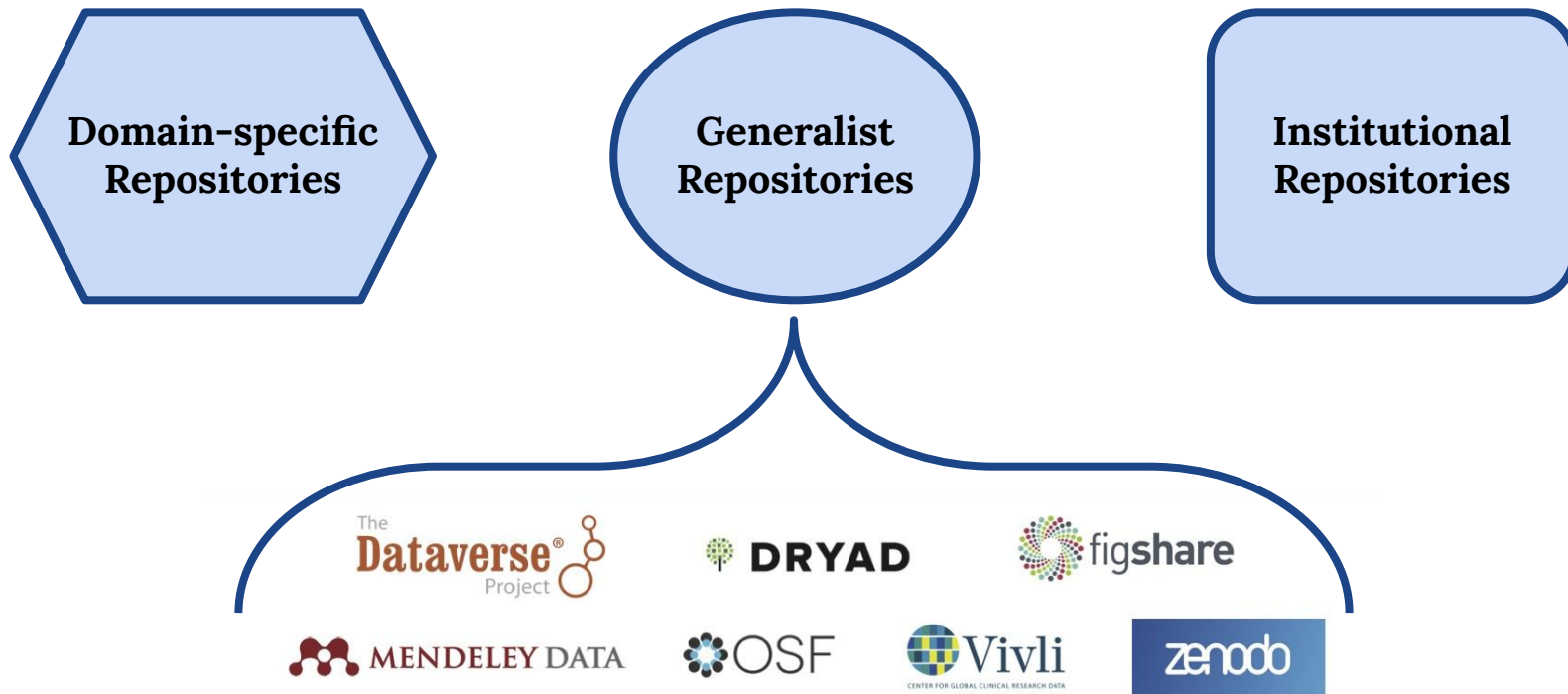
NDA Data Submission Cost Estimation Tool

The screenshot shows an Excel spreadsheet with the following content:

Questions	Sample Answers	Comments
1. How many subjects will be included in the research project?	300	NDA expects to receive data for all subject from whom data were collected including affected individuals, controls, parents, and siblings
2. How many sites will collect data from subjects in this project?	2	
3. How many times will data from this project be submitted to NDA?	11	Standard NDA data submissions occur biannually. Submissions are expected every 6 months over the life of the grant. Please refer to the NIMH Data Archive Data Sharing Regimen site (https://nda.nih.gov/nda/sharing-regimen.html) for specific data repository submission cycle dates.
4. How many data structures will be submitted?	12	Data Structures refers to each measure/data collection instrument within a research project's data dictionary, which will be formatted to NDA standards. These should be individually listed in the Resource Sharing Plan.
5. How many unique experiments (e.g., omics, EEG, eye tracking, fMRI) will be conducted in the study?	5	Investigators must create an NDA Experiment that describes the parameters for each omics assay or neurosignal recording experiment. NDA Experiments need not be created for structural MRI data collection.
6. How many publications analyzing human subjects data are expected for this project?	3	Researchers are expected to share analyzed data using the NDA Study feature which links data submitted to NDA to published results. See https://nda.nih.gov/nda/manuscript-preparation.html for more information
7. What is the hourly rate charged for the Principal Investigator?	\$100	Hourly rate in US Dollars
8. Will the project have a Data Manager?	Yes	Click the cell and use the drop-down arrow to select "Yes" or "No"
9. What is the hourly rate charged for the Data Manager?	\$45	Hourly rate in US Dollars
Total Cost:		\$13,130

Source: <https://nda.nih.gov/contribute/contribute-data.html#cost>

NIH Generalist Repository Ecosystem Initiative



Link:

Data Curation Services: Harvard Dataverse

Service Name	Description	Service Components	Costs	
<i>Data Curation Services</i>			Harvard	Non-Harvard
Free consultation and assessment	Triage level + Office hours Follow-up with recipients of support letters and consultation clients	Demonstrations of software Consult on projects size and scope and fit for Harvard Dataverse	Up to 3 hours	Up to 2 hours
Extended consultation services		<ul style="list-style-type: none"> • Demonstrations of software • Deidentification consult • Dataverse organization consult • File organization consult • Replication verification consult 	\$100/hour	\$200/hour
Dataverse set-up services and dataset and data file ingest	Establish data curation infrastructure and organization enabling data owners to self-curate their data.	<ul style="list-style-type: none"> • Extended Consultation • Dataverse creation and customization • Dataset creation • Data file ingest • Metadata enrichment of dataset and data files • Documentation for use of final product 	\$2000 per "collection" plus \$100 minimum per dataset	\$4000 per "collection" plus \$200 minimum per dataset
Ongoing dataverse administration and curation services	Ongoing maintenance of Dataverse and datasets Continuous updates of dataverse and datasets as needed to reflect software upgrades and sharing standards	<ul style="list-style-type: none"> • Extended consultation services • Dataverse set-up services and dataset and data file ingest • Maintenance of dataverses and datasets within Harvard Dataverse 	\$10,000/year plus \$100 minimum per dataset	\$20,000/year plus \$200 minimum per datasets
Custom services for existing dataverse	One time curation services This includes: <ul style="list-style-type: none"> • file-upload • metadata enhancement (bi-directional linking), • linking affiliated datasets • Documentation • Digitization of audio/paper/content for sharing purposes 		\$100/hour	\$200/hour

Source: <https://support.dataverse.harvard.edu/curation-services>

Typical Questions

- Where are you planning to store the data while the project is active? Is there an associated fee?
- Is there a fee associated with tools or software you are planning to use to collect or analyze the data?
- Do you have subrecipients?
 - If yes, will they be responsible and need to budget for data management and sharing activities?
 - If yes, who will be responsible for depositing into each selected repository or be responsible for covering the associated fees?
- Is there a deposit fee for any of the repositories you are planning to use? If yes, will they allow you to pre-pay prior to the end of the period of performance?
- Will you need dedicated staff time or the services of a core or vendor to support any of the following data management and sharing activities to meet repository requirements?
 - Data curation?
 - Developing supporting documentation?
 - Formatting data according to accepted community standards or for transmission and storage at selected repository?
 - Preparing metadata?
 - De-identifying data?

Resources for Budgeting

- NIH [Allowable Costs for Data Management and Sharing](#)
- National Academies' report on [Life-Cycle Decisions for Biomedical Data: The Challenge of Forecasting Costs](#)
- OpenAIRE (EU) [Estimating Costs RDM Tool](#)
- Mons, Barend. 2020. "[Invest 5% of research funds in ensuring data are reusable](#)." *Nature* 578(7796): 491-491.
- Perry, Anja, and Sebastian Netscher. 2022. "[Measuring the time spent on data curation](#)." *Journal of Documentation* 78(7)

Thank You!

Ceilyn Boyd, Research Data Program Manager

Julie Goldman, Research Data Services Librarian

Katie Mika, Data Services Librarian

Melissa Korf, Senior Director, Research Contracts, Data & Security