Uncovering the costs of academic data sharing: From real data to tips for future planning with ARL and DCN

July 31, 2024

NNLM National Center for Data Services

ASSOCIATION OF RESEARCH LIBRARIES

DATA CURATION NETWORK

Presenters

- **Shawna Taylor**, Project Manager, Open Science, Association of Research Libraries
- **Wendy Kozlowski**, Director, Research Data and Open Scholarship, Cornell University Library
- **Alicia Hofelich Mohr, Ph.D.**, Data Management Research Associate, the College of Liberal Arts, University of Minnesota
- **Mikala Narlock**, Director, Data Curation Network

Research was made possible in part by the Institute of Museum and Library Services, LG-254930-OLS-23, and the National Science Foundation, #213587.

RADS Research Team

- **Jake Carlson**, University at Buffalo (formerly University of Michigan)
- Lizhao Ge, George Washington University/Association of Research Libraries
- **Joel Herndon**, Duke University
- **Alicia Hofelich Mohr**, University of Minnesota
- Lisa Johnston, University Wisconsin, Madison
- Wendy Kozlowski, Cornell University
- **Jennifer Moore**, Washington University in St. Louis
- Jonathan Petters, Virginia Tech
- **Shawna Taylor**, Association of Research Libraries
- **Cynthia Hudson Vitale** (PI), Association of Research Libraries

✚ **Mikala Narlock** (phase two)

Learning Objectives & Agenda

1. Identify RADS Data Management and Sharing (DMS) Activities
2. Understand Real DMS Expenses Reported from NIH Researchers
3. Learn Practical Tips for Future DMS Planning
    a. Understand how the RADS DMS Activities can bridge the connection between budgeting and action

Realities of Academic Data Sharing Initiative

image courtesy of Canva

Phase 1: 2021-2023
Phase 2: 2023-2026

ARL project, in collaboration with the DCN

Phase 1: Research conducted across six universities

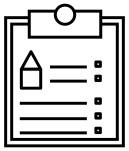
ASSOCIATION OF RESEARCH LIBRARIES

Who

Administrators
N = 69

Research PIs
N = 255

How



Surveys

Interviews

What

→ What **does your unit support** in terms of data management & sharing (DMS) activities? (Admins)

→ What **DMS activities did you do** and did you get support? (Researchers)

→ **How much does it cost**?

- Personnel cost: % effort X salary
- Infrastructure costs

Final NIH Policy for Data Management and Sharing

Notice Number:

NOT-OD-21-013

Key Dates

Release Date:

October 29, 2020

Effective Date:

January 25, 2023

Allowable Costs

Reasonable, allowable costs may be included in NIH budget requests for:

- 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 established repositories, such as data deposit fees.

Unallowable Costs

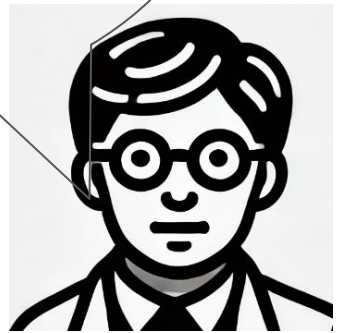
- **Budget requests must NOT include:**
  - Infrastructure costs that are included in institutional overhead (for instance, Facilities and Administrative costs)
  - Costs associated with the routine conduct of research, including costs associated with collecting or gaining access to research data.
  - Costs that are double charged or inconsistently charged as both direct and indirect costs

Many researchers' DMS budget under the new DMS Policy:

"$0 Illustration," generated by OpenAI's DALL·E 3, July 2024.

I already do these things…why do I need to budget for them separately?

Data sharing is new to me…won't salaries for DMS just eat into the budget? I need to focus on the science…



"Researchers, icon style," generated by OpenAI's DALL·E 3, July 2024.

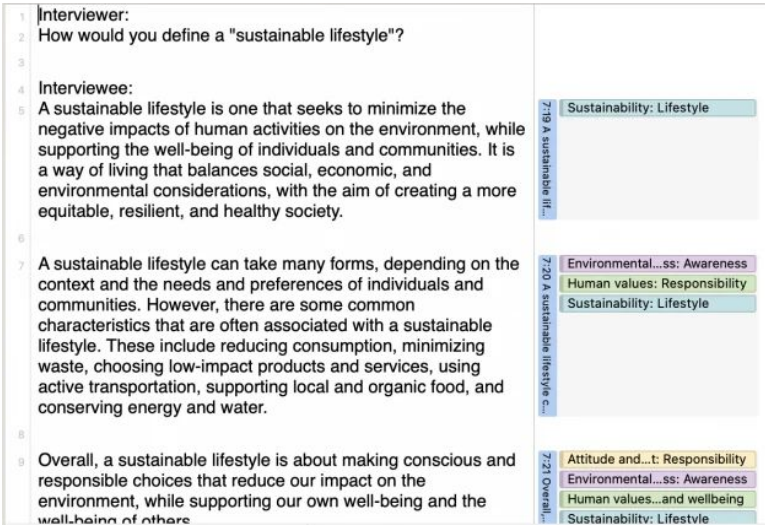
Data Management and Sharing Activities

Coding

Need: A shared vocabulary around and definitions for typical activities done by researchers in preparation for sharing research data

```
241 ```{r}
242 atotal_inst_unit %>%
243   pivot_longer(cols = Total_expense:Infrastructure
244                names_to = c("Type", ".value"),
245                names_sep = "_") %>%
246   mutate(SE = SD/sqrt(N)) %>%
247   filter(Type == "Total") %>%
248   ggplot(aes(x=Category, y=expense)) +
249   geom_errorbar(aes(ymin = expense-SE, ymax = expe
    = .5, position = position_dodge(width = .75)) +
250   geom_col(fill="#3182bd",
251            position = position_dodge(width=.75),
252            color="black",
253            width=.75) +
254   scale_fill_manual(values = arlcol3)+
255   scale_y_continuous(labels = scales::comma) +
```

Computer Scripts



Qualitative Data Classification

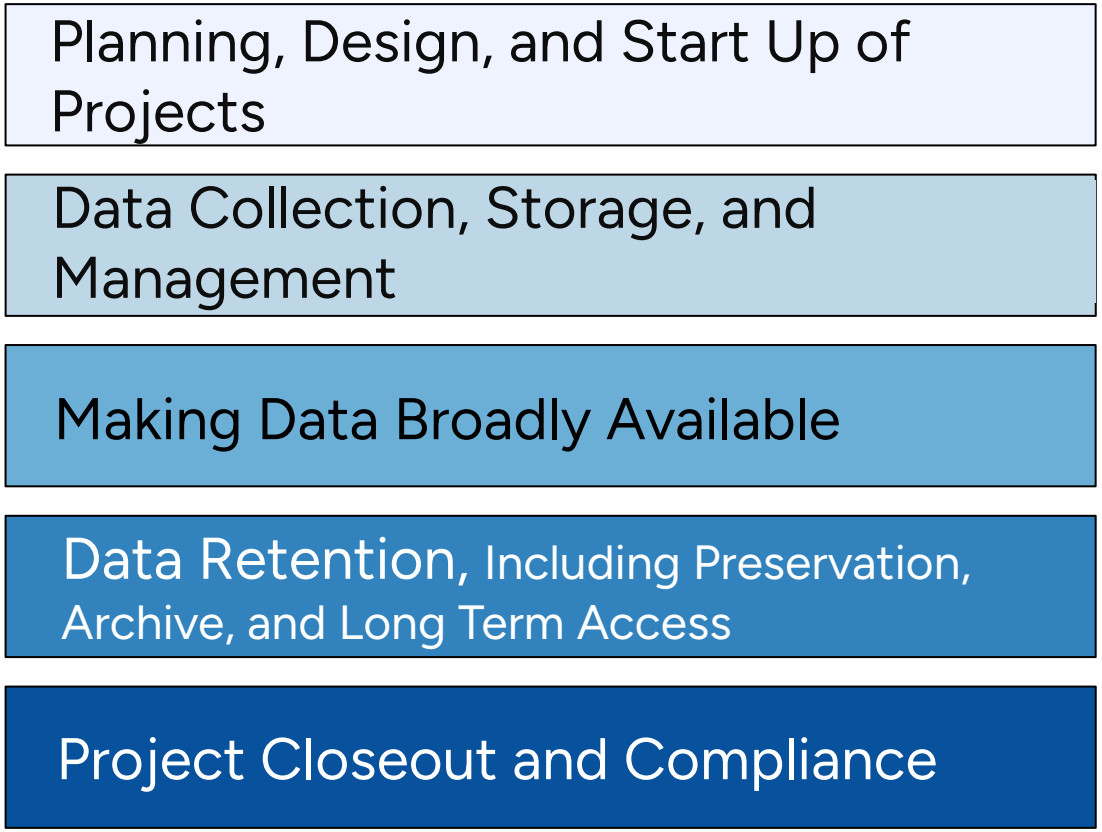
Data Management and Sharing Activities

Process:

- ⏩ Identified activities done in support of making data publicly accessible
- ⏩ In collaboration with (COGR) Council on Governmental Relations (Roles and Responsibilities Matrix)
- ⏩ Initial list included some exemplar tasks to further clarify activities
- ⏩ Refined based on community feedback

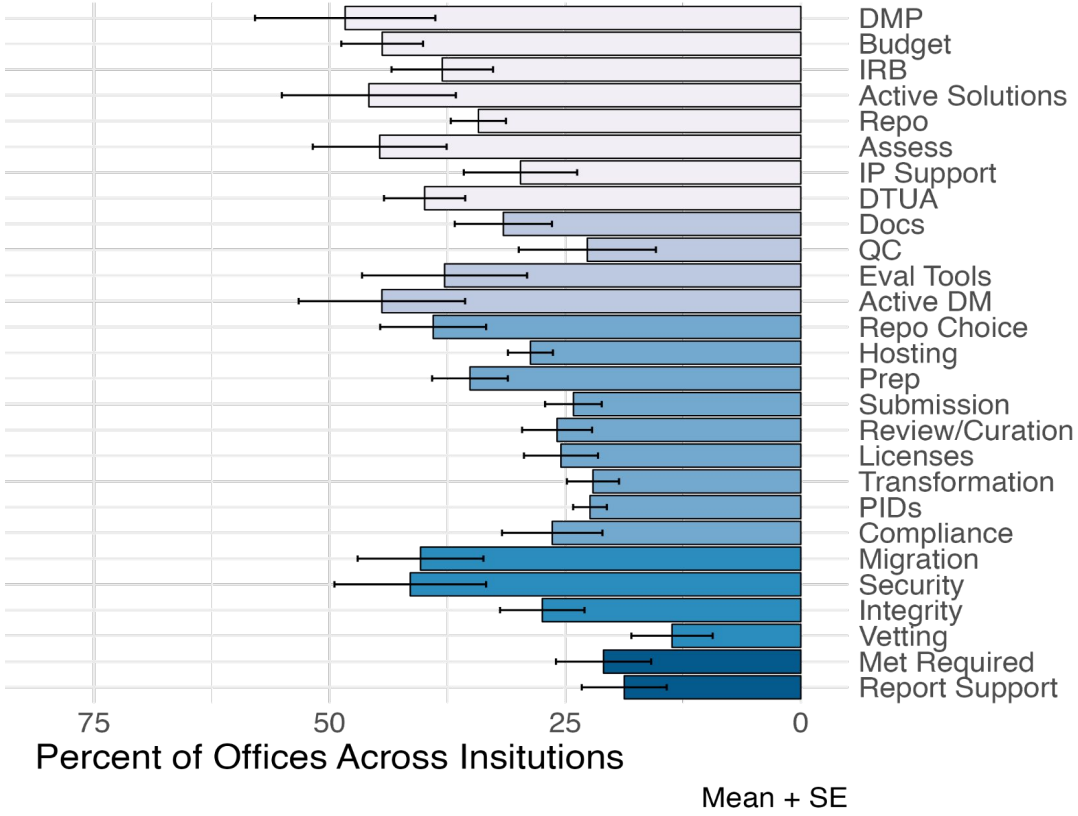
Researchers	Institutions
Preparing data management plans (DMPs) or data management and sharing plans (DMSPs), which includes determining who has responsibility for various research data related activities, and communicating these plans to the research team	Consulting on or preparing data management plans (DMPs) or data management and sharing plans (DMSPs), including considerations for ethical and privacy-related practices
Identifying data management and sharing costs to be included in grant budgets	Consulting on data management and sharing costs and expenses to be included in grant budgets
Preparing institutional review board (IRB) protocols and informed consent language for data sharing	Reviewing of institutional review board (IRB) protocols and informed consent language for data sharing
Determining storage solutions for active research data	Developing, building, providing, or recommending storage solutions for active research data
Selecting an appropriate repository (or repositories) for making research data broadly available	Developing, building, providing, or recommending appropriate repository (or repositories) for making research data broadly available (this includes advising on, managing, hosting, running, or otherwise facilitating access to a repository)

Data Management and Sharing Phases

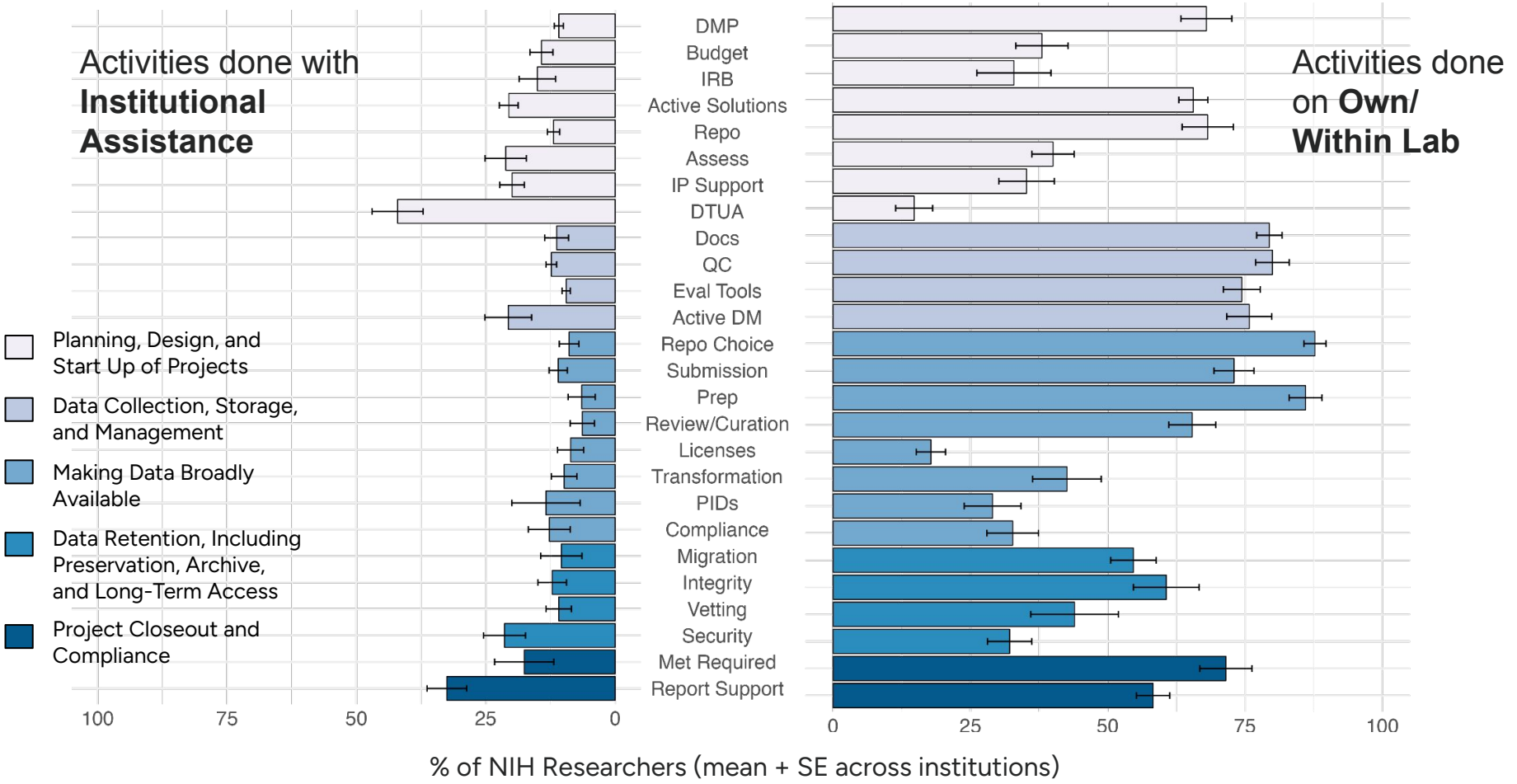


Institutions Support Most Data Management and Sharing (DMS) Activities

- Planning, Design, and Start Up of Projects
- Data Collection, Storage, and Management
- Making Data Broadly Available
- Data Retention, Including Preservation, Archive, and Long-Term Access
- Project Closeout and Compliance

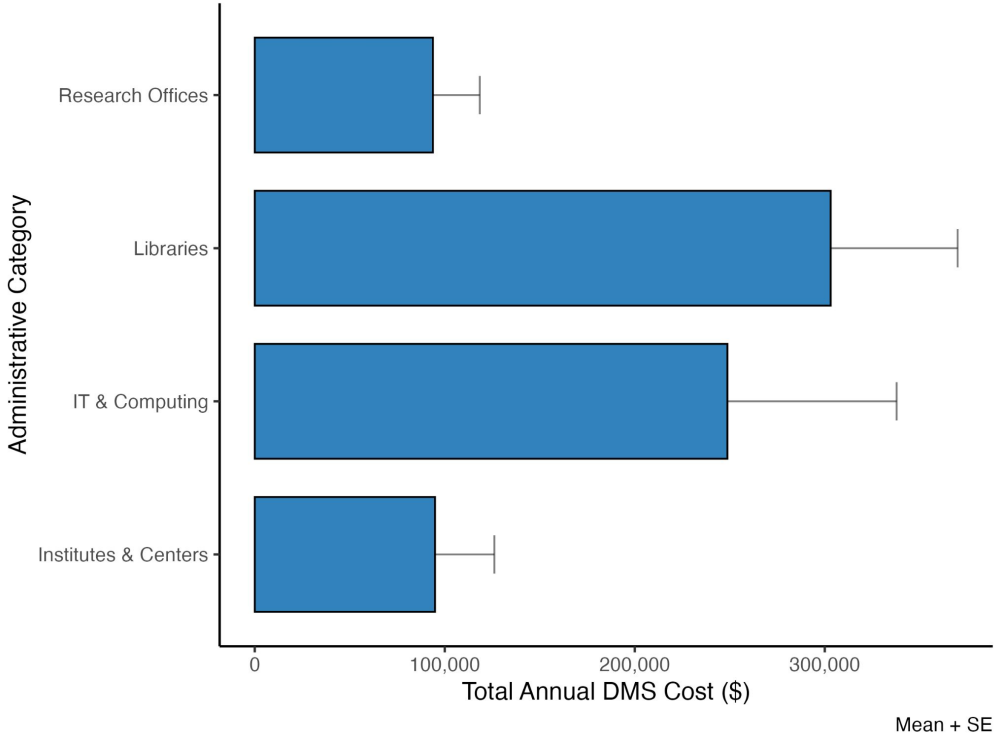


Most researchers do most data sharing activities on their own



Institutional offices are investing in DMS

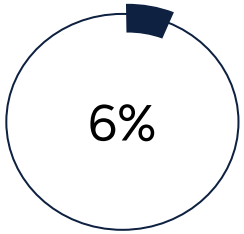
- Offices where researchers may find internal support
- Libraries and IT faced largest total DMS expenses



Average Researcher Expenses

$29,800

Average per funded project cost to researchers



On average, 6% of grant award went to DMS

Researcher Expenses

Researchers were grouped by their total award amount

- Average NIH Award ~ $2,730,000; n = 56
- Average NSF Award ~ $428,000; n = 32

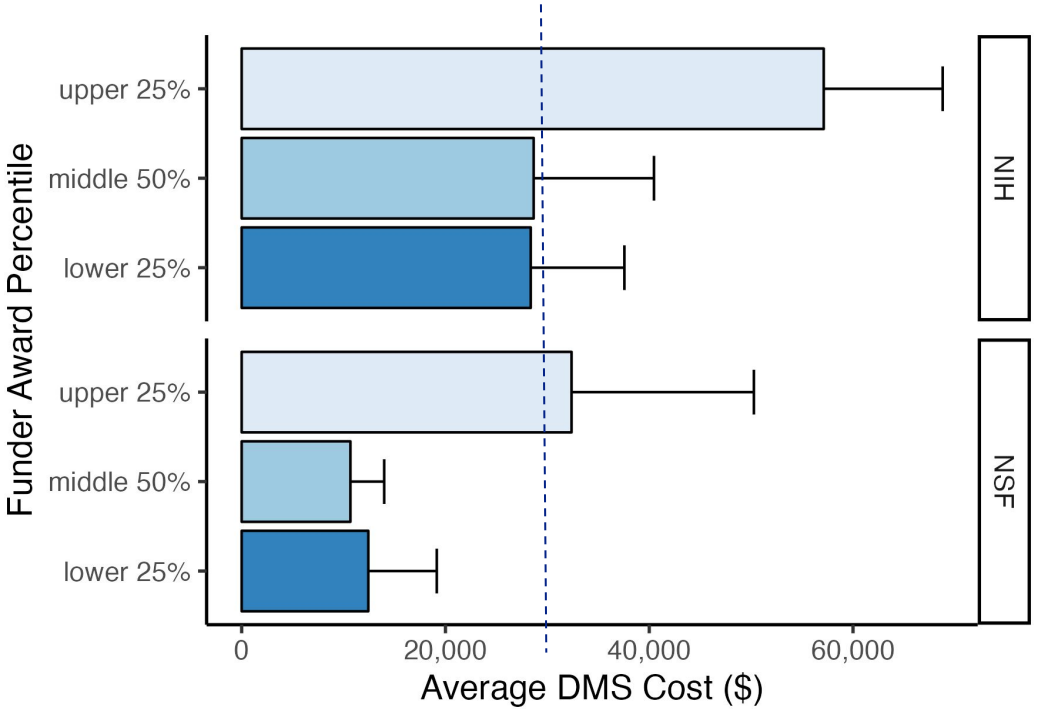
	NIH			NSF		
Percentile Group	**Award min**	**Award max**	**N**	**Award min**	**Award max**	**N**
Lower 25th	$8,000	$421,000	14	$15,000	$138,000	5
Middle 50th	$422,000	$1,910,000	27	$150,000	$470,000	16
Upper 75th	$2,020,000	$123,000,000	15	$474,000	$8,200,000	11

Average DMS expenditures by funder and percentile of total awarded grant amount

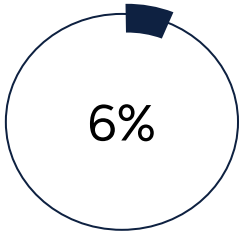
$29,800

Average NIH-only
~ $36,200

Average NSF-only
~ $18,400

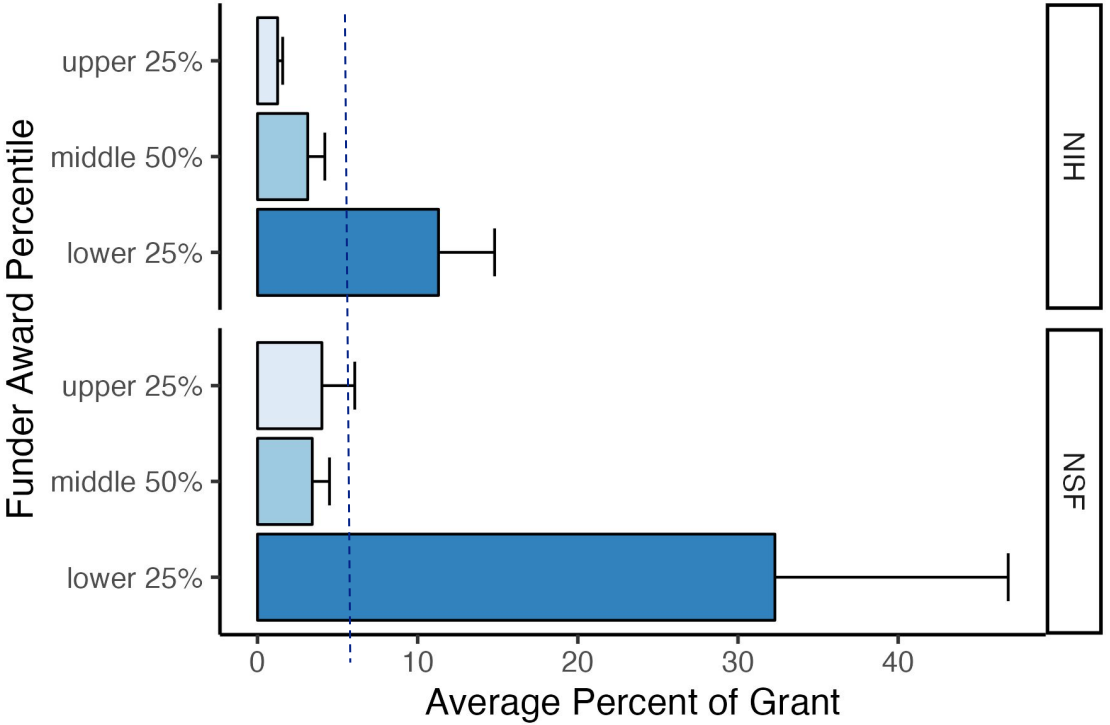


Average DMS expenditures by funder of total awarded grant amount



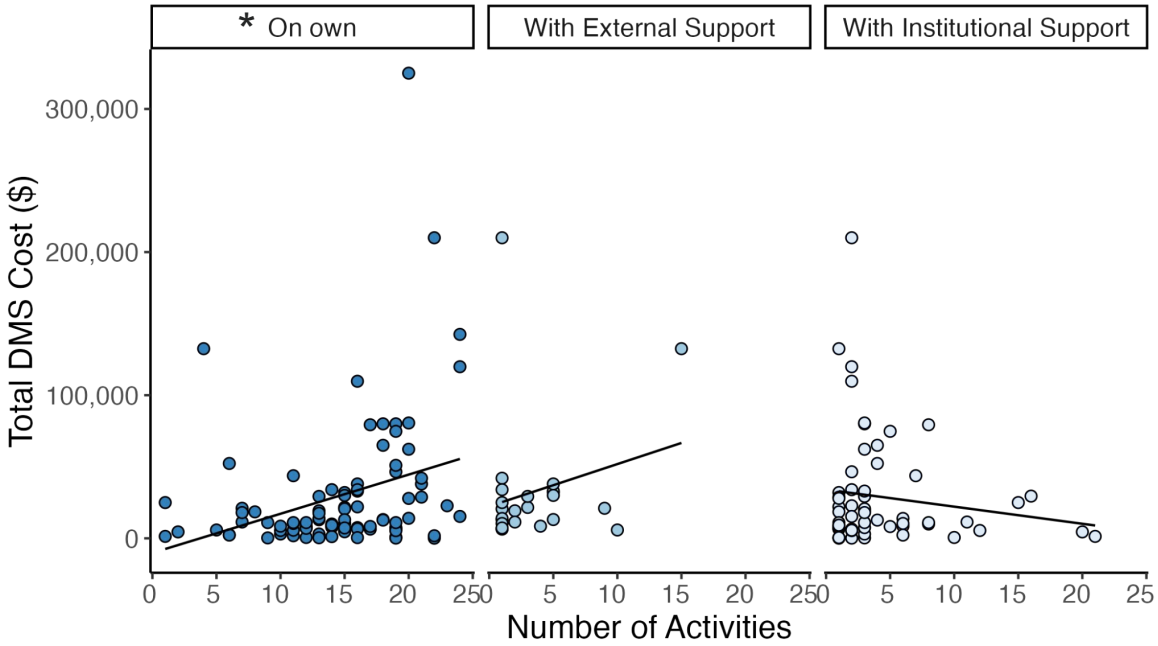
Average NIH-only
~ 4.7%

Average NSF-only
~ 8.1%



How do Researcher Expenses Relate to Use of Institutional Services?

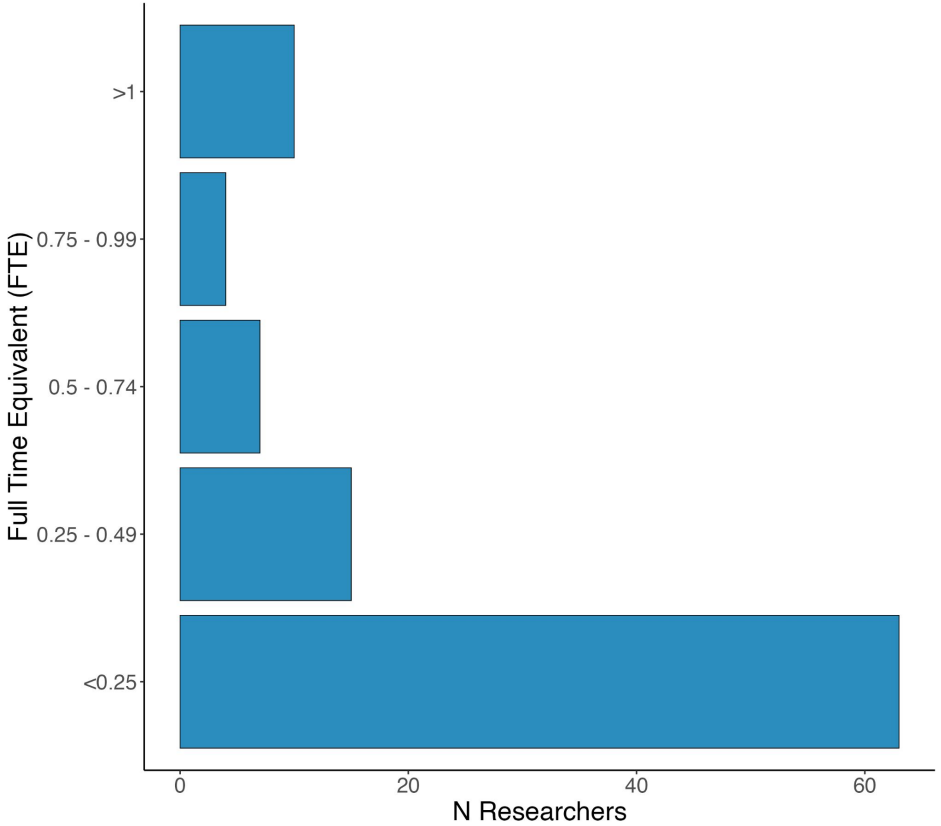
Doing DMS activities on their own was associated with greater DMS expense



* b = 2735.5, t(88) = 2.96, p = .004; adjusted R^2 = 0.08; all other p > .28

How are research teams staffing DMS work?

Most doing Data Management & Sharing on the research team did it less than 25% of their time



Community of 50+ data curators addressing sharing challenges

How are YOU using the RADS Expense Report?

- In DMP consultations

- In conversation with library administration to:
  - **Advocate for a full-time employee**
  - Consider sharing and preservation costs to researchers

- In conversation with organization administration to:
  - Inform our grant support model (e.g., percent of personnel costs)
  - Direct v. Indirect Costs

Some costs are "easier" to identify than others

Easier to identify/quantify

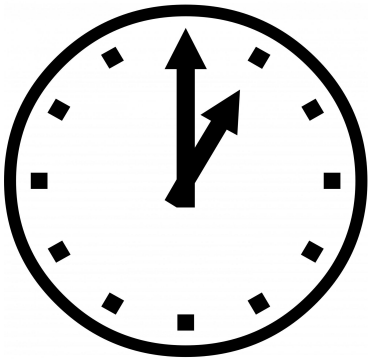
- Repository or curation fees
- Storage costs
- Salary for a Project Data Manager

Harder to identify/quantify

- Time doing "active" data management
- Institutional training/education
- Personal curation

How I talk to researchers about DMS budgeting

1. Designate a person on the team to have "primary" data management responsibility → % of their time
    a. How complex is the data/project?
    b. How large and disperse is the team?
2. Include professional data management support from institution → % of time (direct) or explicit mention in facility plan (indirect)
    a. Consultation, training, mentorship, doing of the work



How I talk to researchers about DMS budgeting

3. Determine if repository has a fee
    a. If yes → write in $
    b. If not → write in additional hours for team or professional to prep and make submission
4. Identify other hurdles for sharing (e.g., sensitive or restricted data; large data) and budget extra time or professional consultation for handling

Challenges to consider (and ask about)

1. What are the realities of post-award costs? How to plan ahead to ensure these are spent before final reports?
2. How to evaluate when data should be deaccessioned?
3. How can we shift the mindset from sharing being an “extra” to being a fundamental part of “doing research”?

Contact Info

Shawna Taylor - staylor@arl.org

Wendy Kozlowski - wak57@cornell.edu

Alicia Hofelich Mohr - hofelich@umn.edu

Mikala Narlock - mnarlock@umn.edu

Cynthia Hudson Vitale, PI - cvitale@arl.org

And follow our progress:

https://www.arl.org/realities-of-academic-data-sharing-rads-initiative/

RADS Resources & Project Results

RADS Data Management and Sharing Activities, v3:

Kozlowski, W., et al. “Realities of Academic Data Sharing (RADS) Initiative Public-Access Data Management and Sharing (DMS) Activities,” v3. November 6, 2023. https://doi.org/10.29242/radsdmsactivities2023.

Expense Report:

Hofelich Mohr, A., et al. *Making Research Data Publicly Accessible: Estimates of Institutional & Researcher Expense.* February 2024. https://doi.org/10.29242/report.radsexpense2024.

Gap Analysis Report:

Petters, J., et al. *Publicly Shared Data: A Gap Analysis of Researcher Actions and Institutional Support throughout the Data Life Cycle*. March 2024. https://doi.org/10.29242/report.radsgapanalysis2024.

Methodology Report:

Taylor, S., et al. *Realities of Academic Data Sharing (RADS) Initiative: Research Methodology 2022–2023 Surveys and Interviews.* January 2024. https://doi.org/10.29242/report.radsmethodology2023.