

The best things in life are free but data sharing is not: A multi-institutional study of the realities of academic data sharing

*Presented to IASSIST 2024
Halifax, Nova Scotia*

**/ ASSOCIATION
OF RESEARCH
LIBRARIES /**

Research Team

- **Jake Carlson, University at Buffalo**
- 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
- **Mikala Narlock, Data Curation Network**
- Jonathan Petters, Virginia Tech
- Shawna Taylor, Association of Research Libraries
- Cynthia Hudson Vitale (PI), Association of Research Libraries

Session Overview

- Phase 1 – DMS Expenses Data Collection & Analysis
 - Background & Methodology
 - Institution Results
 - DMS Opportunities
 - Research Results
- Phase 2 RADS – Collaborations and Implications
 - Data Curation Network
 - Joining institutions to RADS

RADS Background and Methodology

Phase 1: Retrospective Study, 2021-2023



UNIVERSITY OF MINNESOTA



Cornell University



Funded by the National Science Foundation (NSF) EAGER grant #2135874: Completing the Lifecycle: Developing Evidence Based Models of Research Data Sharing, 2021-2023



Study Participants



Administrators

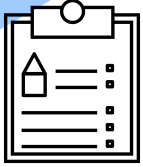
- Units that support any data sharing activities
- Expenditure and salary knowledge
- Fiscal year 2021/2022



Researchers

- Funded between 2013-2022
- Funded by: DOE, NIH, NSF

Research Methods



Surveys

Administrators – **Fiscal Year 21/22**

- Out of 27 DMS activities, which data sharing activities their unit supports
- Personnel cost = % effort × salary
- Infrastructure costs

Researchers – **During the grant period**

- Out of 27 activities, data sharing activities were done during their grant period
- Personnel cost = % effort × salary
- Infrastructure costs



Interviews

RADS DMS Activities & Phases

Phase	Number of Activities*
Planning Design and Start Up	9
Data Collection, Storage, and Management	4
Making Data Broadly Available	8/9
Data Retention, Including Preservation and Long-Term Access	4
Project Closeout and Compliance	2

*27 activities for researchers; 28 for institutions/administrators

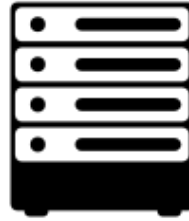
Service Categories



Research
Offices
(RSCH)



Libraries +
Archives
(LIB)



IT +
Compute
(IT)



Institutes
+ Centers
(IC)



Libraries



Libraries provide support for public access to research data throughout the grant lifecycle.



Libraries and data sharing

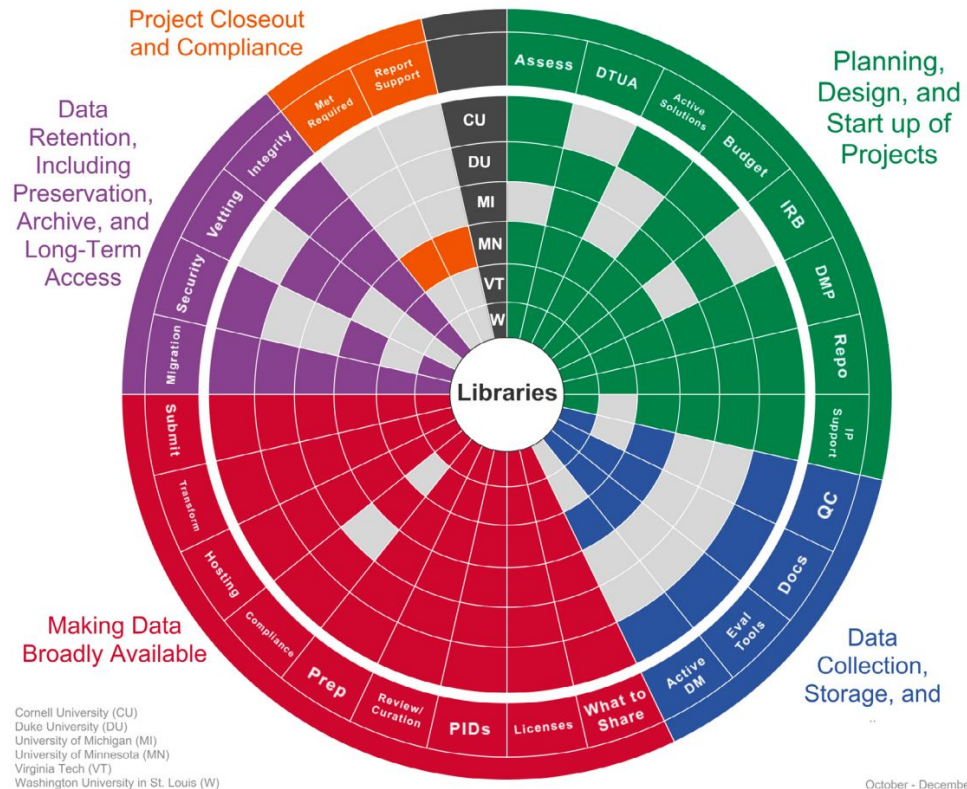
Broad support for:

- Planning, Design, and Start Up of Projects
- Making Data Broadly Available
- Data Retention, Including Preservation, Archive, and Long-Term Access

Support is less prevalent for:

- Data Collection, Storage, and Management
- Project Closeout

**Libraries – Services & Infrastructure for Public Access to
Research Data ([LINK](#))**



October - December 2022

<https://tabsoft.co/3yCyU5x>

Opportunities for Libraries

- **Providing guidance on data licensing**
 - (66% of researchers reported “not doing”)
- **Assigning persistent identifiers (DOI, ROR, ORCID)**
 - (only 20% of researchers report internal assistance)
- **Consulting on selecting data for publishing and data publishing in general**

Interested in more opportunities? <https://bit.ly/rdm-ops>

The page features four decorative hexagonal shapes at the corners. The top-left and bottom-right corners have a cyan hexagon overlapping a light blue one. The top-right and bottom-left corners have a medium blue hexagon overlapping a light blue one. The word "Considerations" is centered in a bold, dark blue font.

Considerations



Data sharing instruction

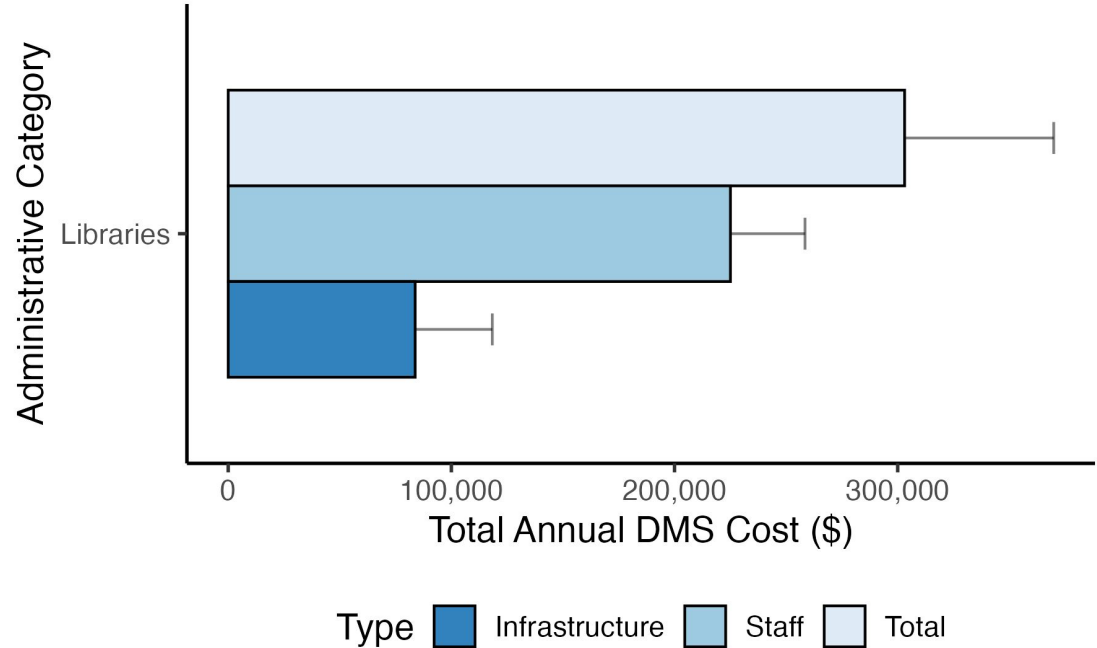


The slide features a white background with decorative elements in the corners. Each corner contains a cluster of overlapping hexagons in various shades of blue, including light blue, medium blue, and a vibrant cyan. The central text is a large, bold, dark blue sans-serif font.

Financial Considerations

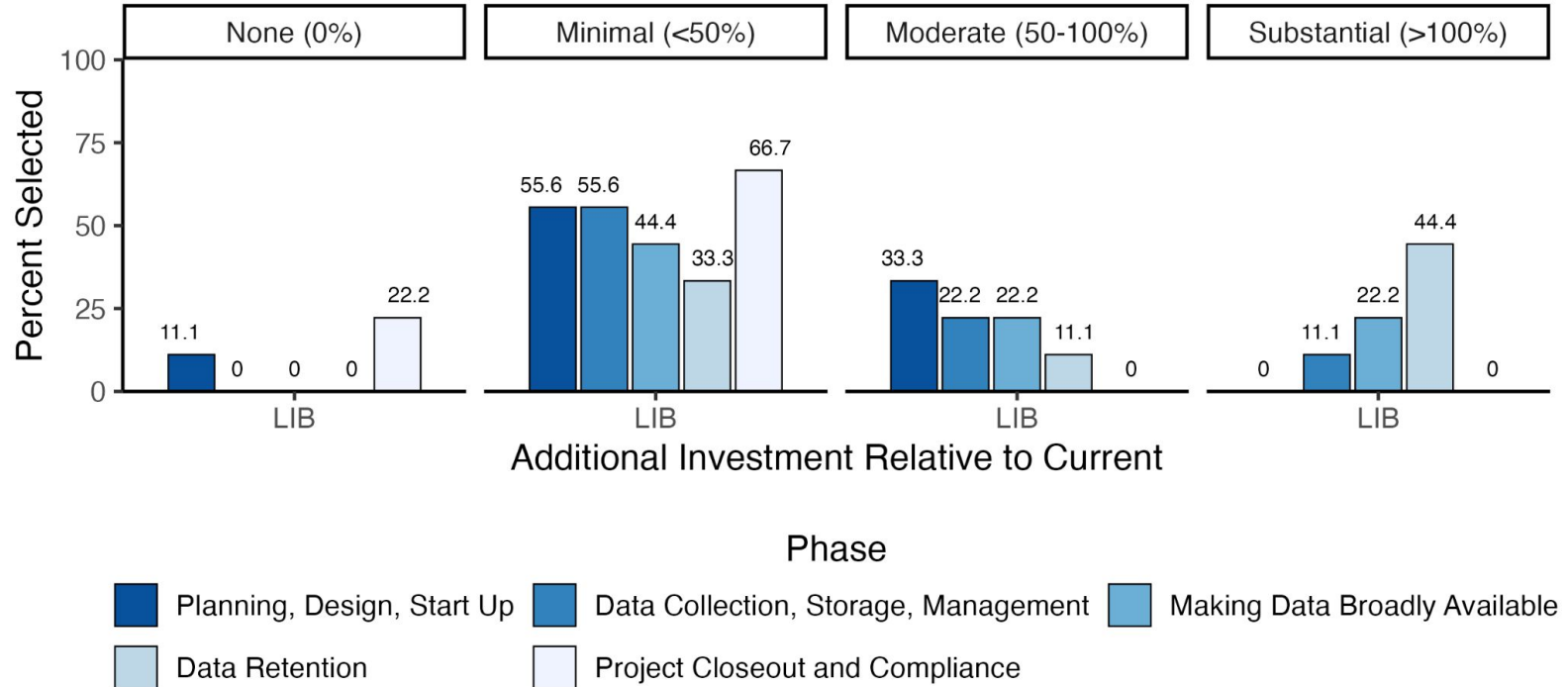
Libraries Expenses

- Libraries have the largest total data sharing expenses
- Staffing was largest expense for libraries



Mean + SE

Libraries future investments in RDM





**IT, Research Offices,
Specialized Institutes &
Centers**

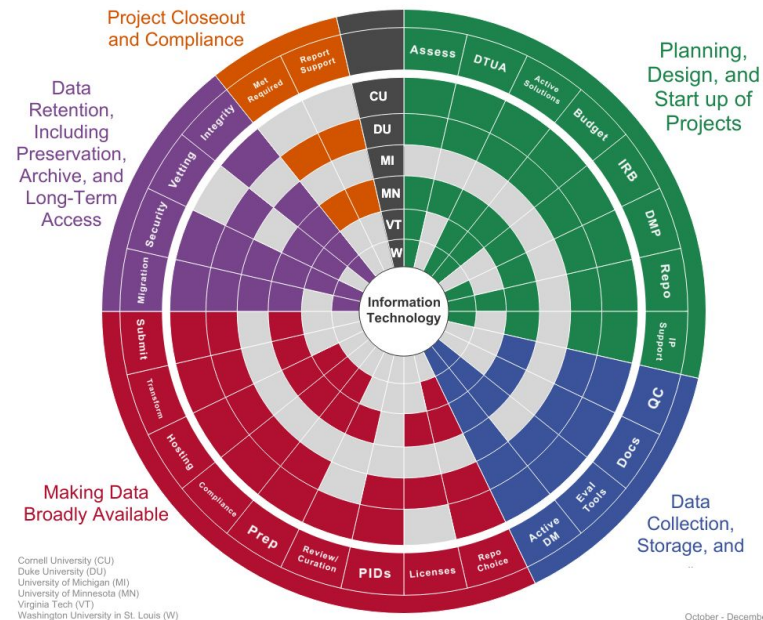
Information Technology Offices - Services & Infrastructure for Public Access to Research Data (LINK)

IT Offices provide support for public access to research data across most of the phases

- 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

Less support provided in

- Project Closeout and Compliance



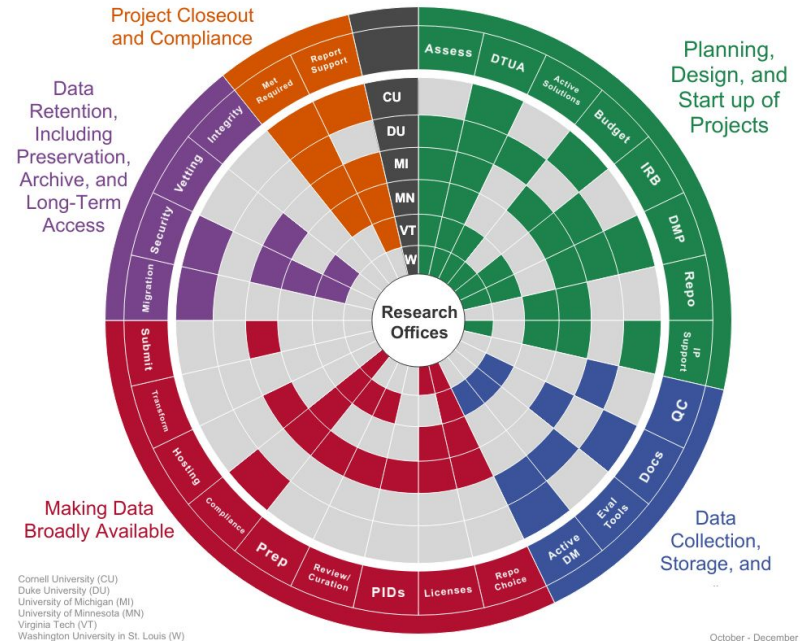
Research Offices - Services & Infrastructure for Public Access to Research Data ([LINK](#))

Research Offices provide support for public access to research data primarily in the beginning and end stages.

- Planning, Design, and Start Up of Projects
- Project Closeout and Compliance

Less support provided in

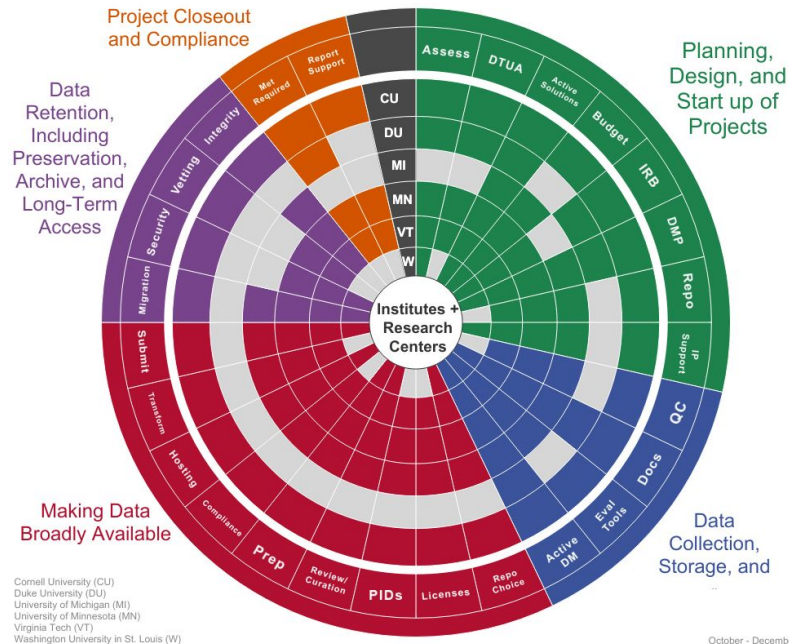
- Making Data Broadly Available



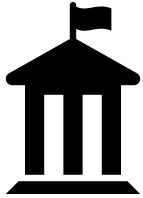
Institutes & Research Centers - Services & Infrastructure for Public Access to Research Data (LINK)

Research Centers provide support for public access to research data in all of the phases.

But individual centers focus on their members, not the institution as a whole.



Administration / Service Units are still adjusting



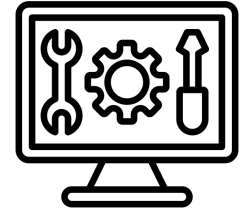
Much of the support for data management and sharing provided by the institution is:

- Center the needs of the institution, rather than the researcher.
- Generally, more focused on minimizing risk, than on sharing data.

Libraries, and IT Units to varying extents, provide researcher focused data services, but many are not taking advantage of the services, or know they exist.

Opportunities for Underutilized / Underdeveloped Services

- For IT Departments
 - Data security services
 - Creating quality control mechanisms or procedures for infrastructure
- For Research Offices
 - Ensuring funding agency requirements for data sharing have been met



Created by SITI NURHAYATI
from Noun Project

IT department by SITI
NURHAYATI from [Noun Project](#)
(CC BY 3.0)



Created by Pike Picture
from Noun Project

Workplace research by Pike Picture
from [Noun Project](#) (CC BY 3.0)

Opportunities for Underutilized / Underdeveloped Services

- For Research Institutes & Specialized Centers
 - May not be available to provide outside services
 - May serve as pilots or models for providing support services



Created by Studio 365
from Noun Project

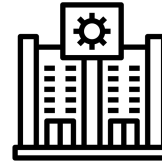
Research center by
Studio 365 from
[Noun Project](#) (CC
BY 3.0)

Opportunities for Cross Campus Collaboration

- Developing recommendations, policies and practices for deaccessioning / removing research data at the institution
- Identifying and budgeting for the costs of data management and sharing
- Training / Education



Created by SITI NURHAYATI
from Noun Project



Created by Studio 365
from Noun Project



Created by Pike Picture
from Noun Project

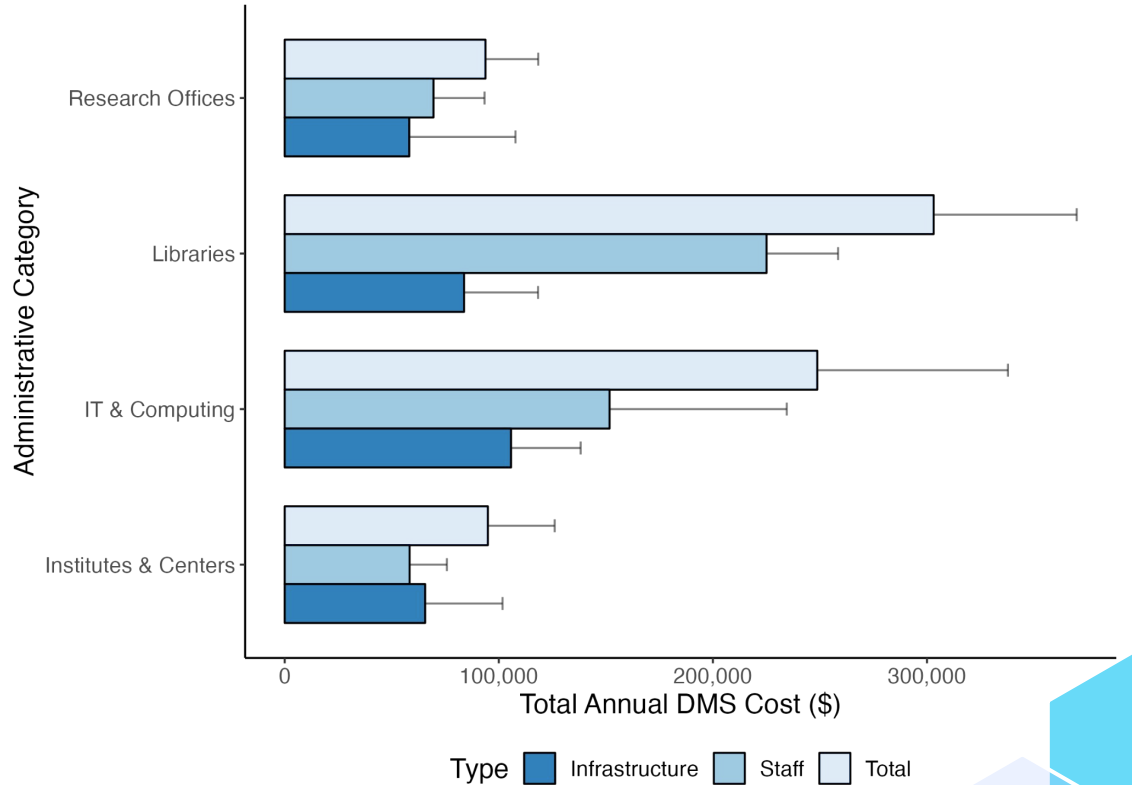


Created by M. Oki Orlando
from Noun Project




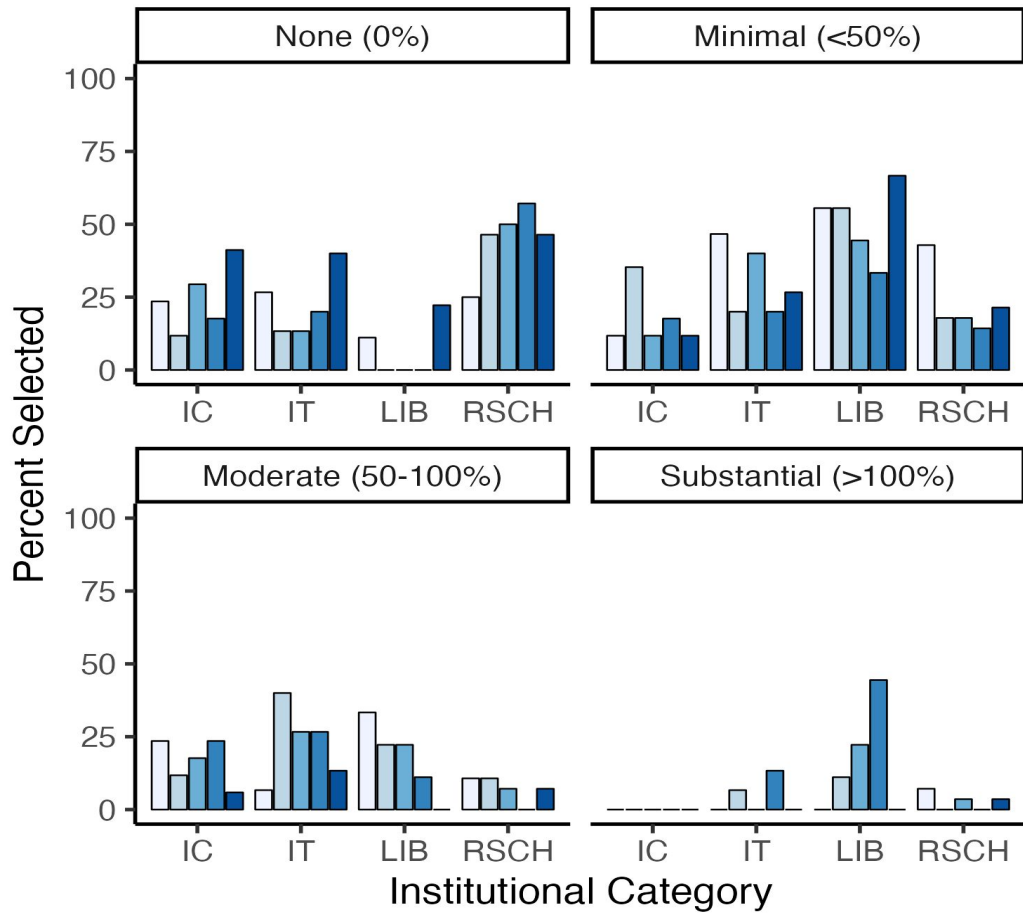
Administrator Expenses

- Averaged across responses within institutional service areas
- Libraries and IT faced largest total DMS expenses
- Staffing was largest expense for libraries



Future DMS Investments within 5yrs by Category

-  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





Researchers

How do researchers engage with Data Management & Sharing?



Research PIs

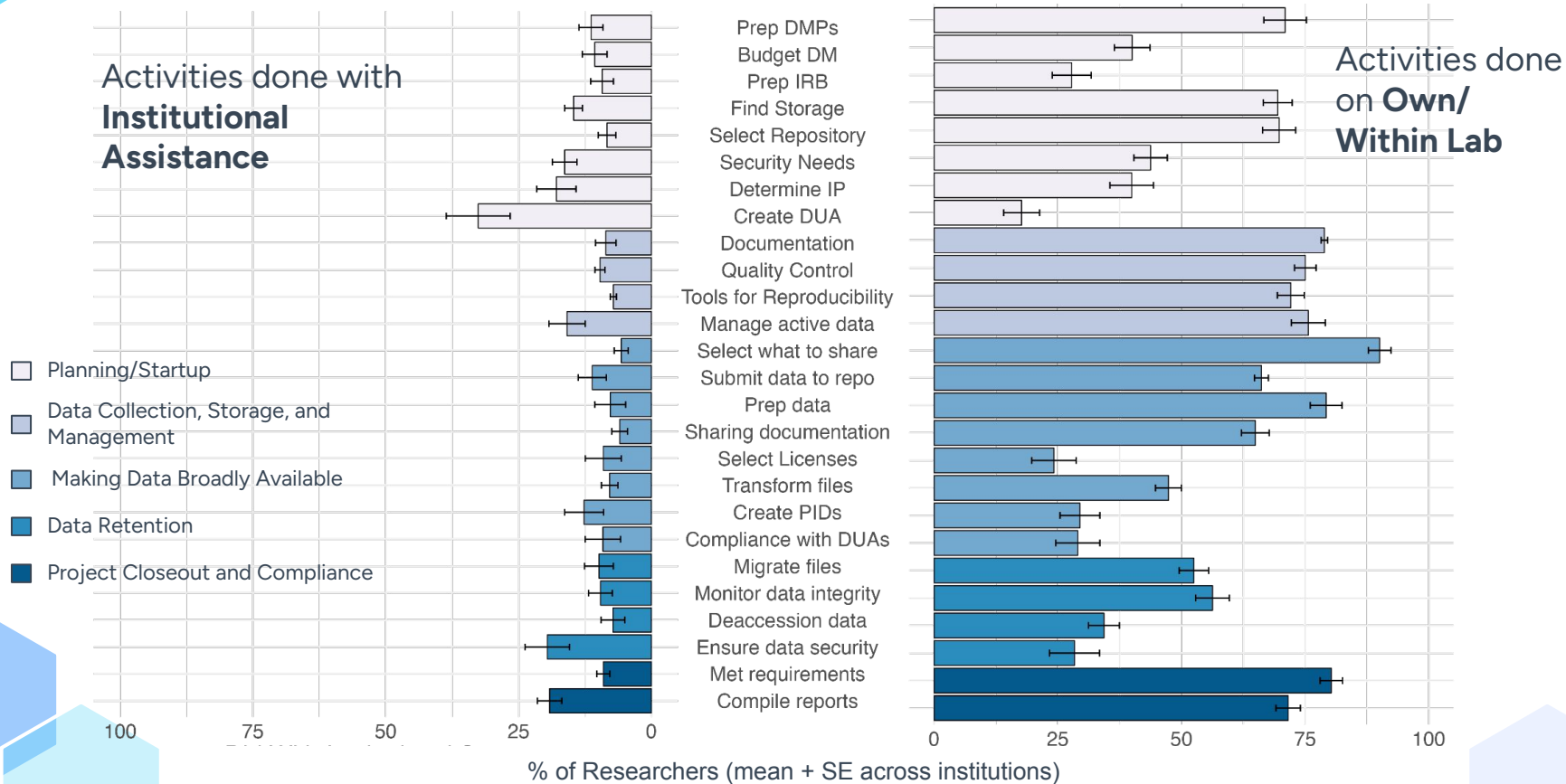
N = 255 (8.4%);
91 with complete
expense data

NSF, NIH, DOE
Funded Grant

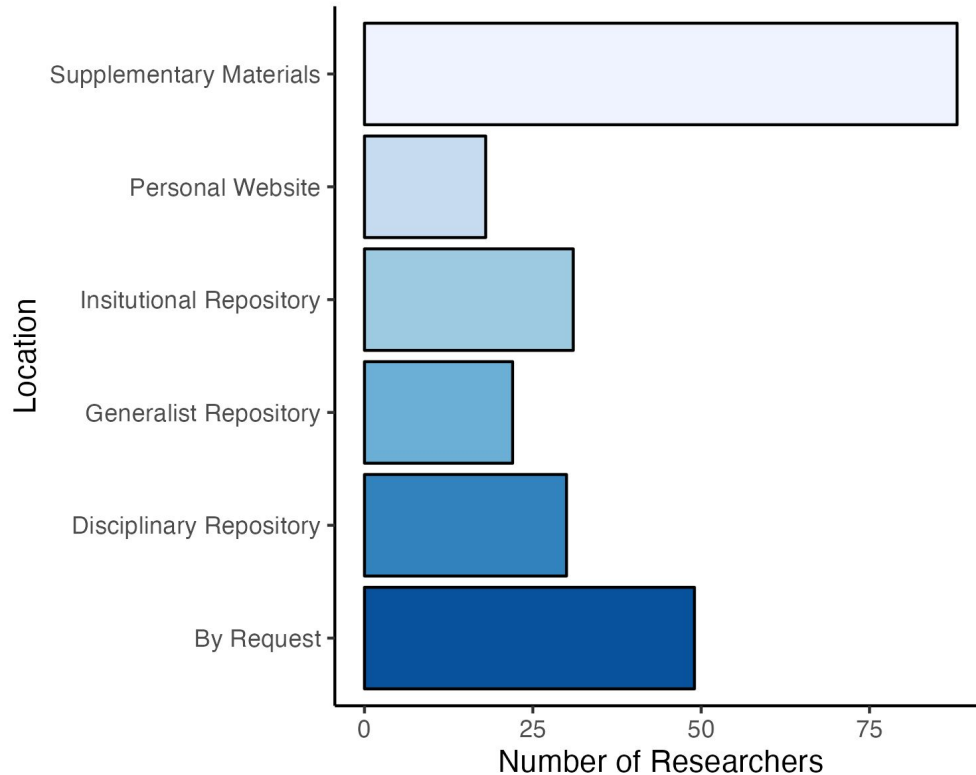


- What data management and sharing activities did you do and did you get support?
- Where was your data shared?
- How did you decide where to share?
- How much did it cost?
 - Personnel cost: % effort X salary
 - Infrastructure costs
- Would you share again?

Most researchers do data management & sharing on their own



Where did they share data and why?



57% Selected a single category;
30% selected two
13% selected three or more

Top 5 Influences: Where did researchers share?

1. Easiest/quickest option – 71.4%
2. Personal experience – 67.9%
3. Least expensive option – 56.0%
4. Funder recommendation – 44.0%
5. Journal/publisher recommendation – 43.5%

...

10. Library recommendation – 10.7%
11. Research office recommendation – 9.5%
12. Campus IT recommendation – 3.0%

Researcher Expenses

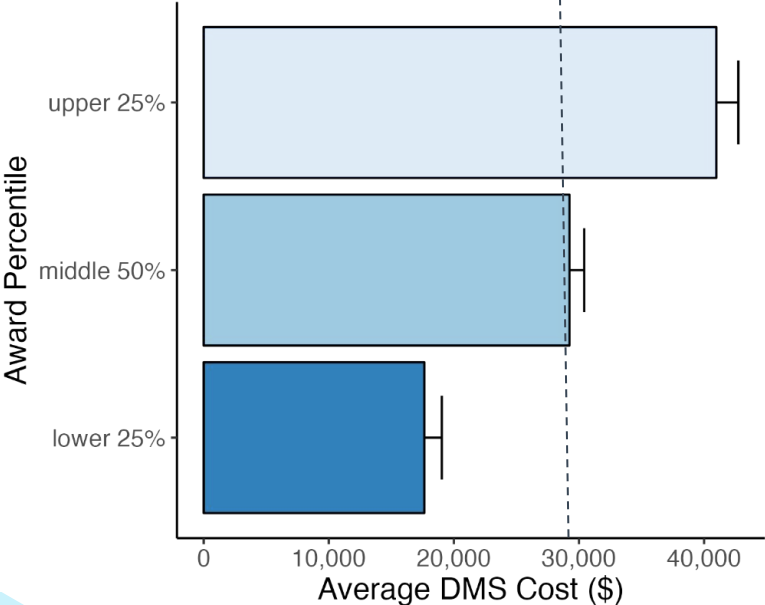
Researchers were grouped by their total award amount

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

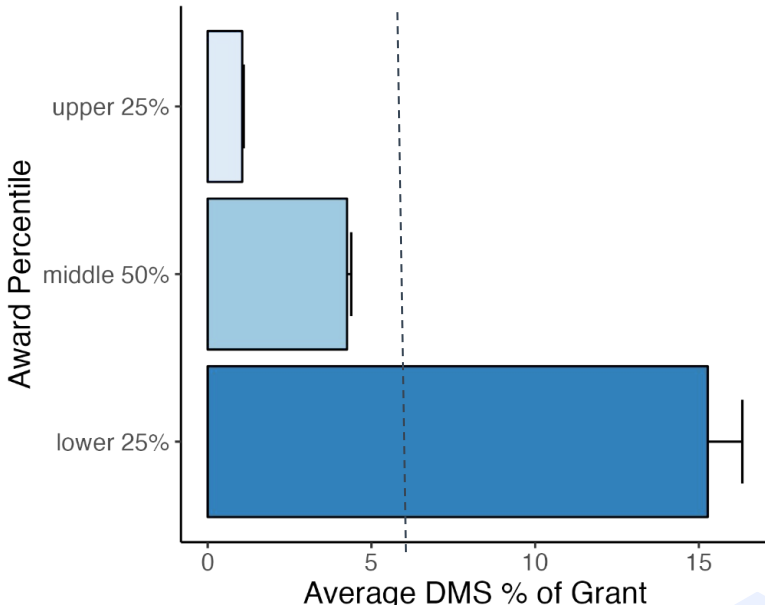
Percentile Group	Award min	Award max	NIH	NSF	DOE	Total N
Lower 25th	\$8,000	\$310,906	7	13	0	20
Middle 50th	\$326,386	\$1,382,409	26	18	3	47
Upper 75th	\$1,467,763	\$122,910,010	23	1	0	24

Researcher Expenses

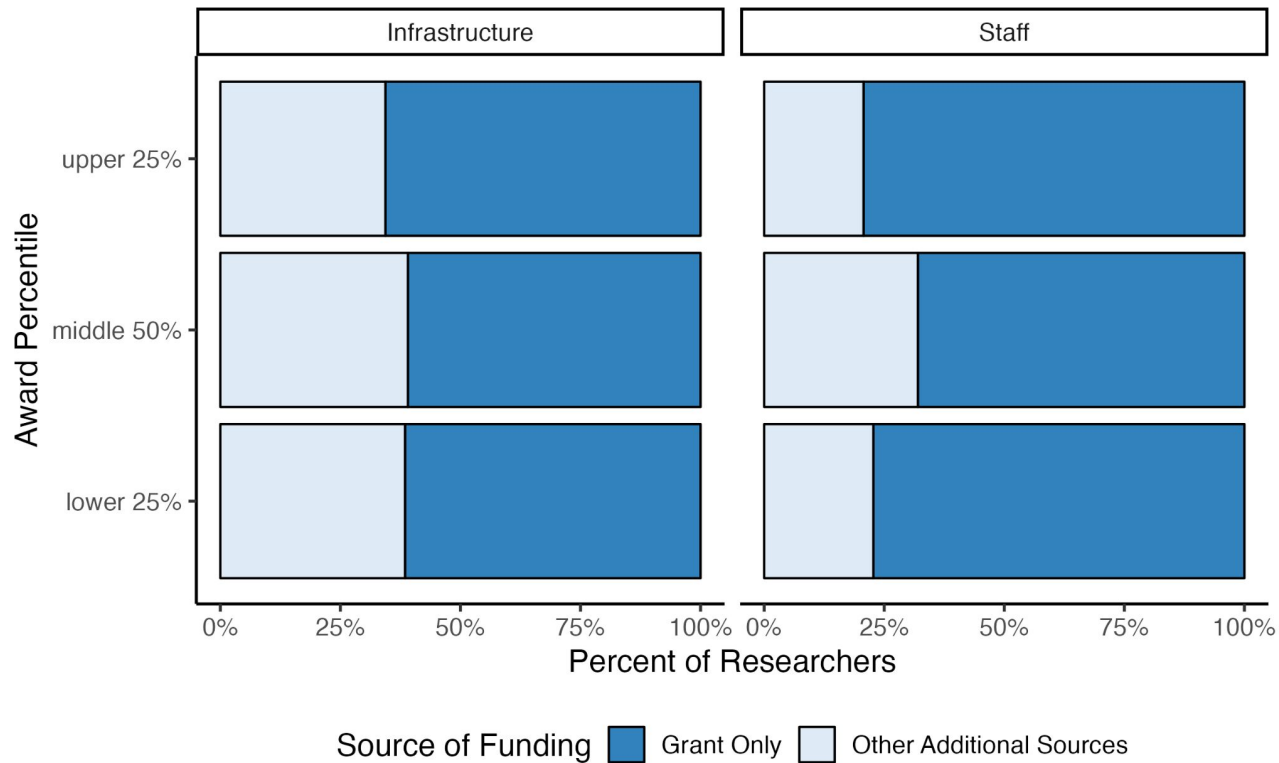
Overall DMS Expense
~ \$29,800



DMS Expense as Percent of Award
~6%

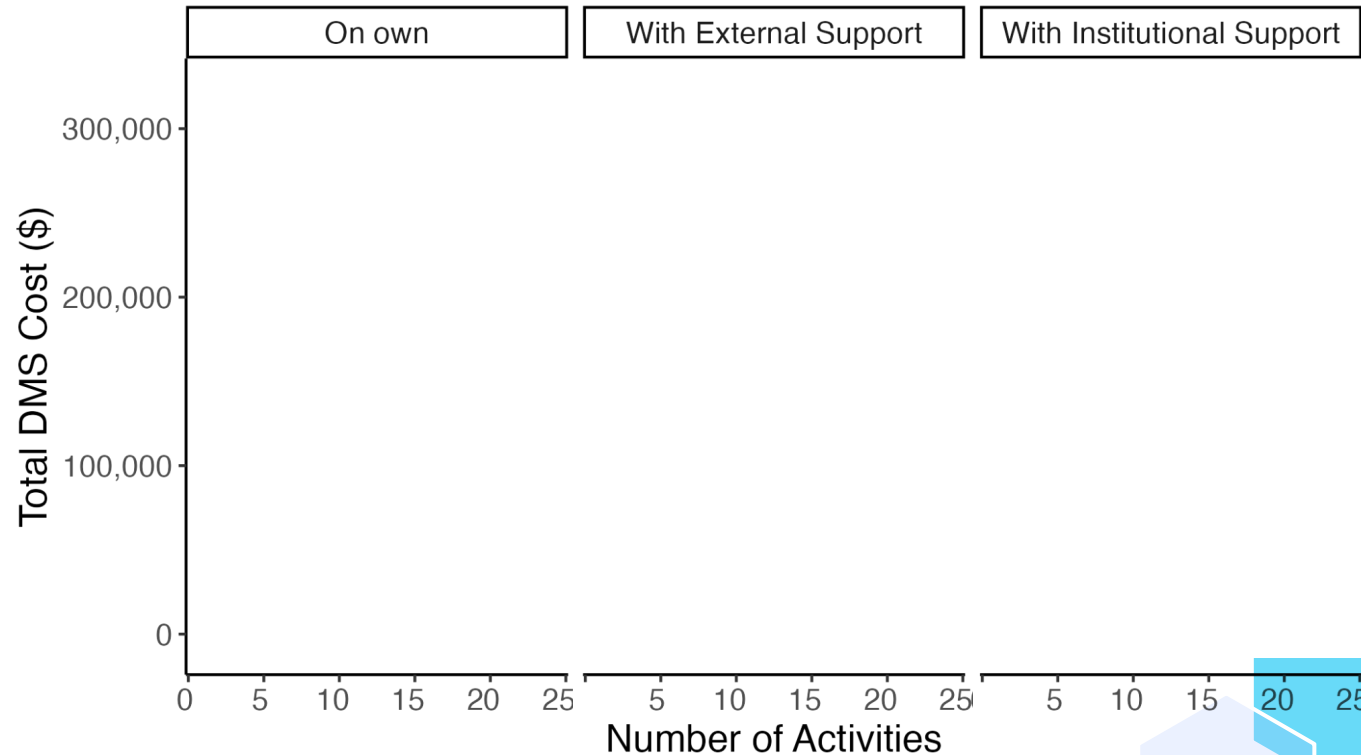


How are these expenses covered?



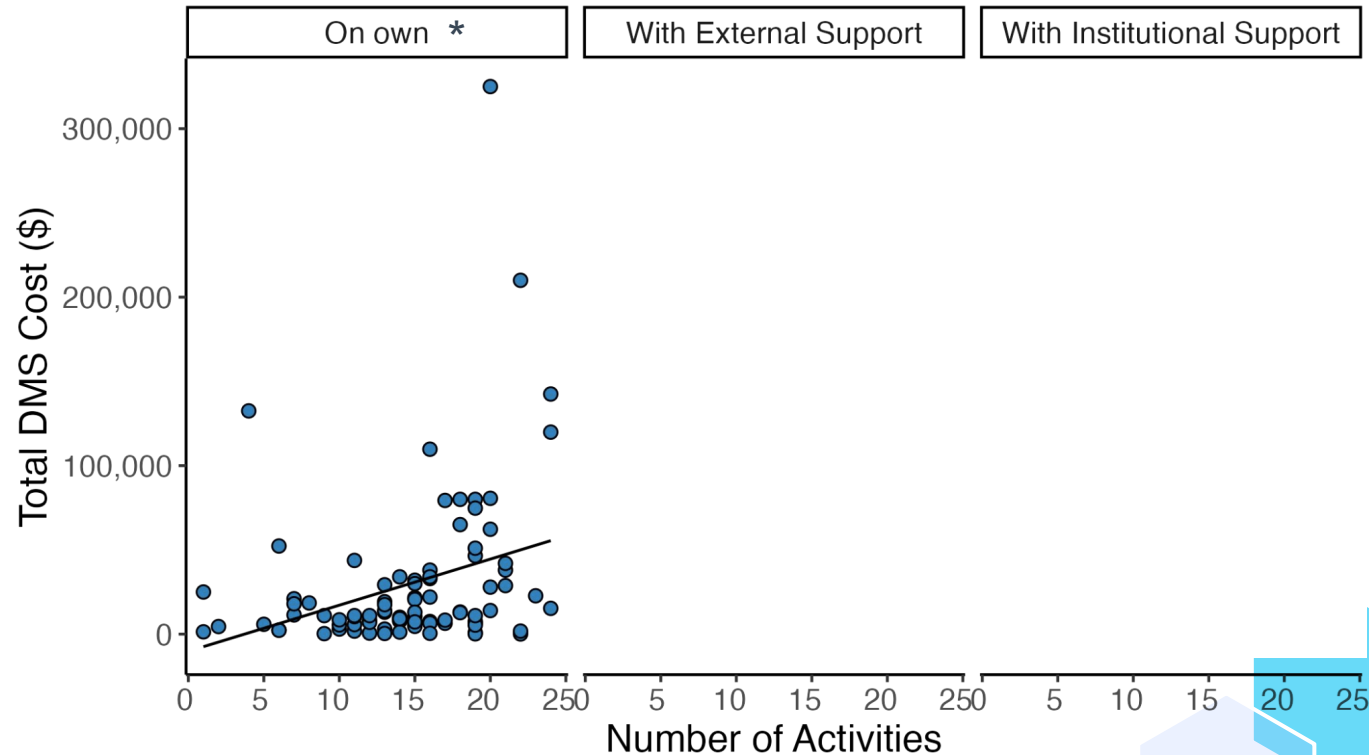
How do Researcher Expenses Relate to Use of Institutional Services?

Does the number of activities done with (or without) support correlate with total cost for data management and sharing?



How do Researcher Expenses Relate to Use of Institutional Services?

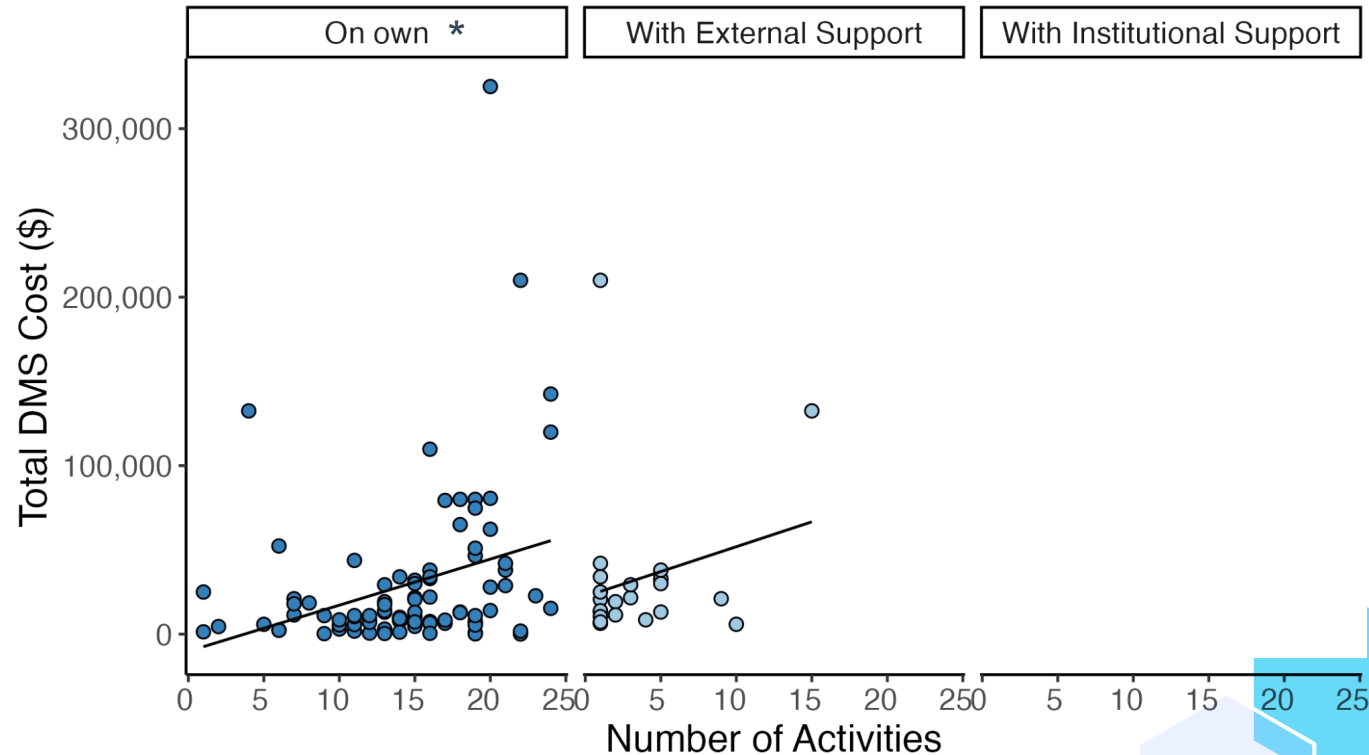
Doing DMS activities on their own was associated with greater DMS expense



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

How do Researcher Expenses Relate to Use of Institutional Services?

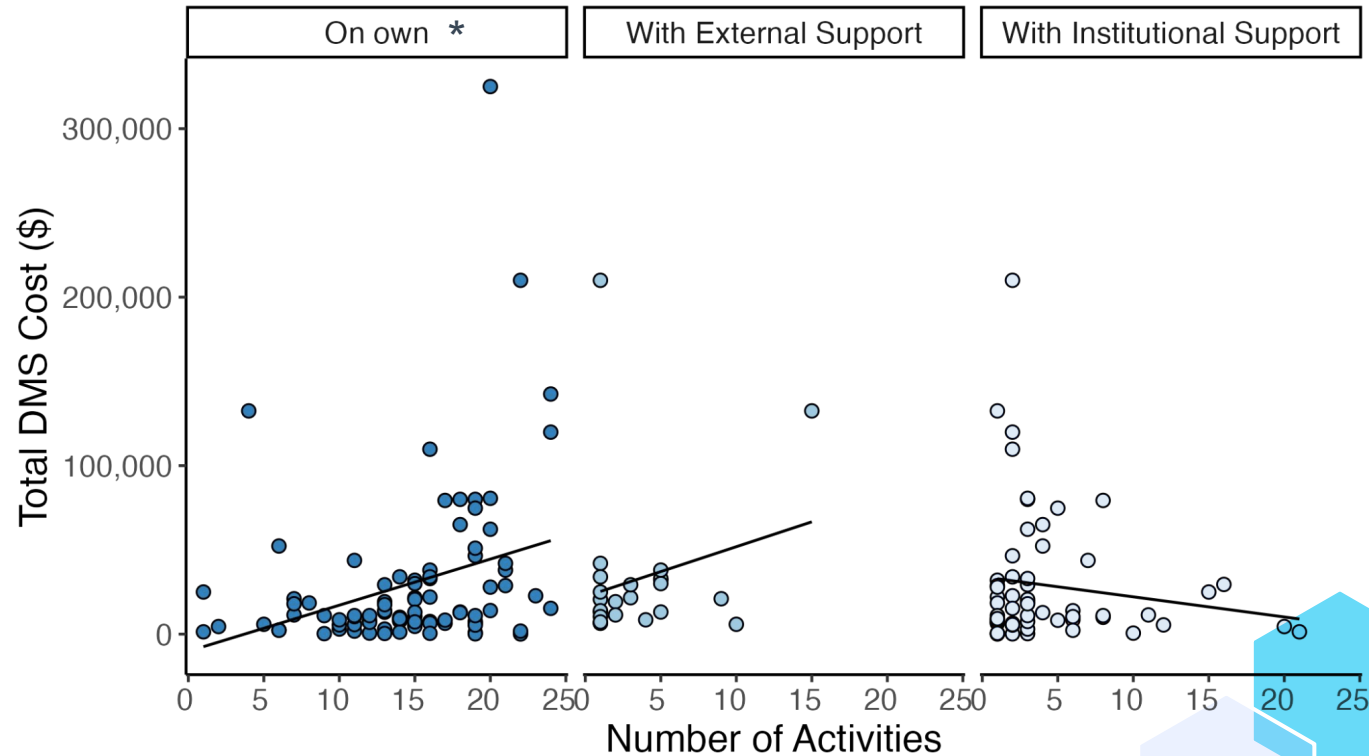
Doing DMS activities on their own was associated with greater DMS expense



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

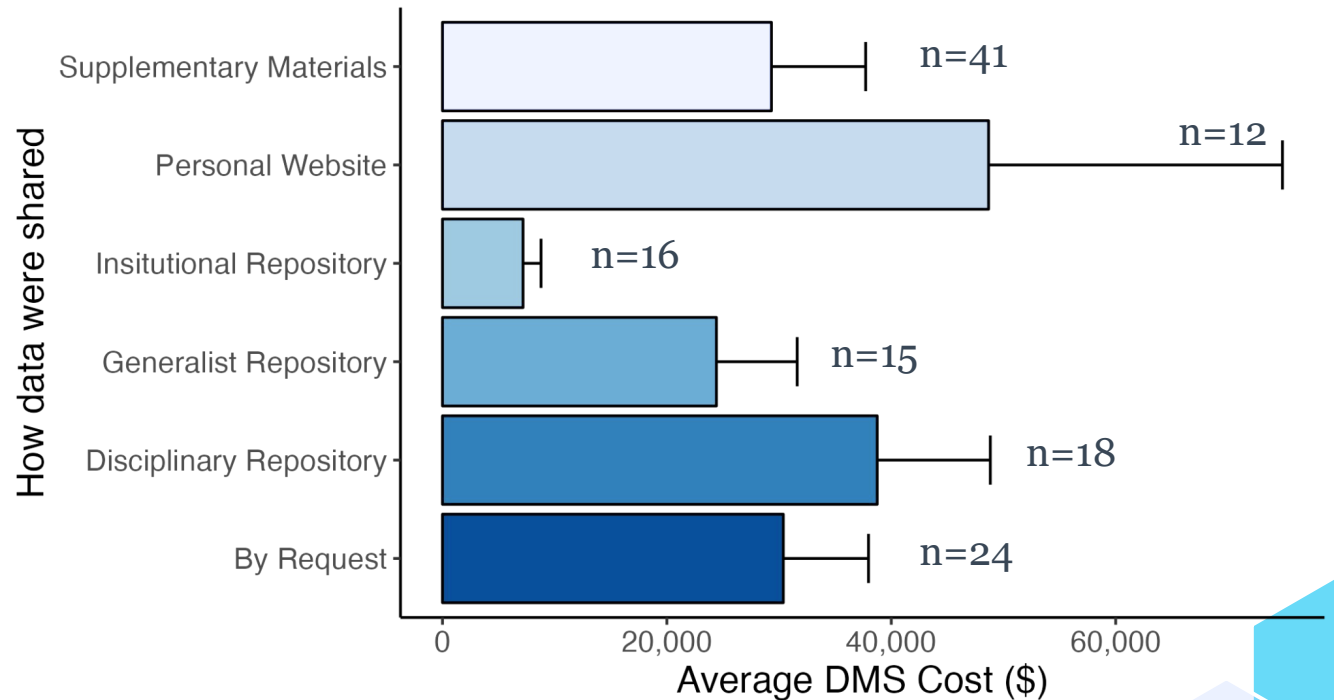
How do Researcher Expenses Relate to Use of Institutional Services?

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



How do Researcher Expenses Relate to Use of Institutional Services?

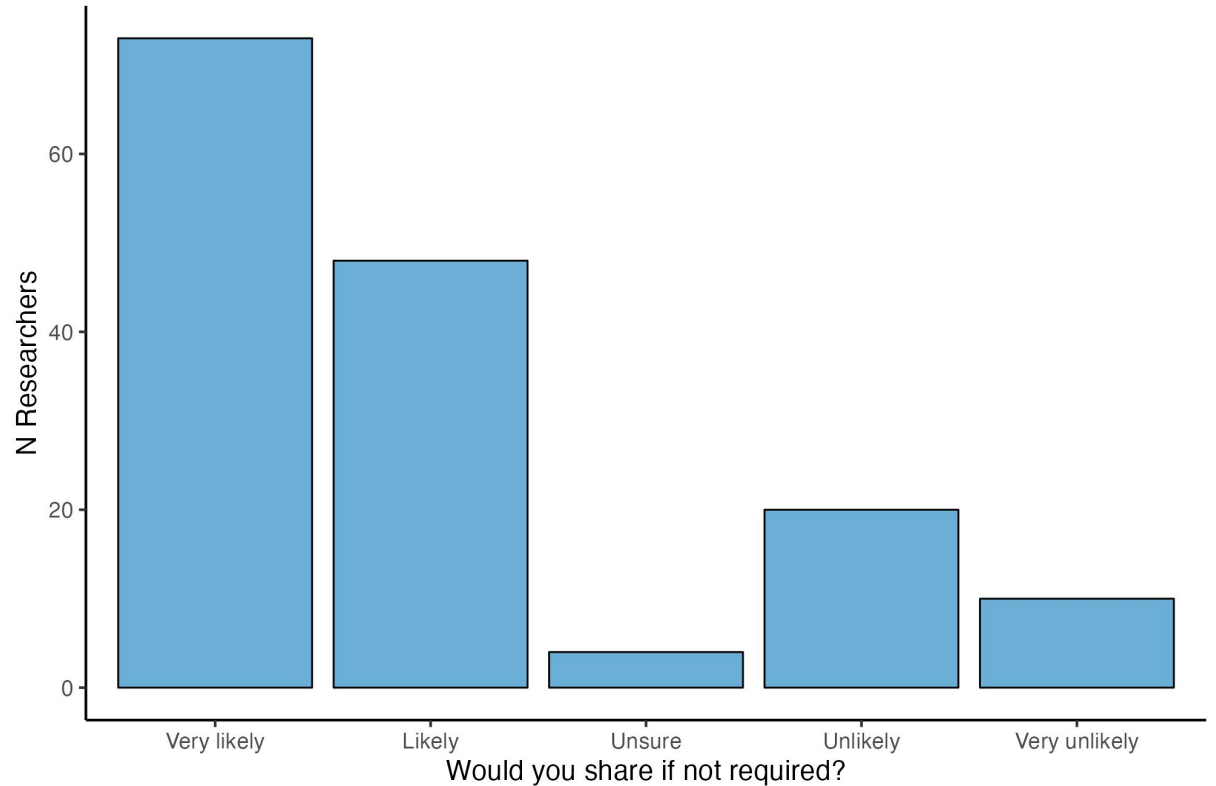
Sharing in an Institutional Repository (IR) was associated with lower DMS Expenses



Mean + SE

Would they share again?

About 80% of researchers were likely or very likely to share data again if they were not required to.



Current Collaborations and Implications



As professional data curators, data management experts, data repository administrators, disciplinary scientists and scholars we represent academic institutions and non-profit data repositories that steward research data for future use.



Example: Benefits at WashU

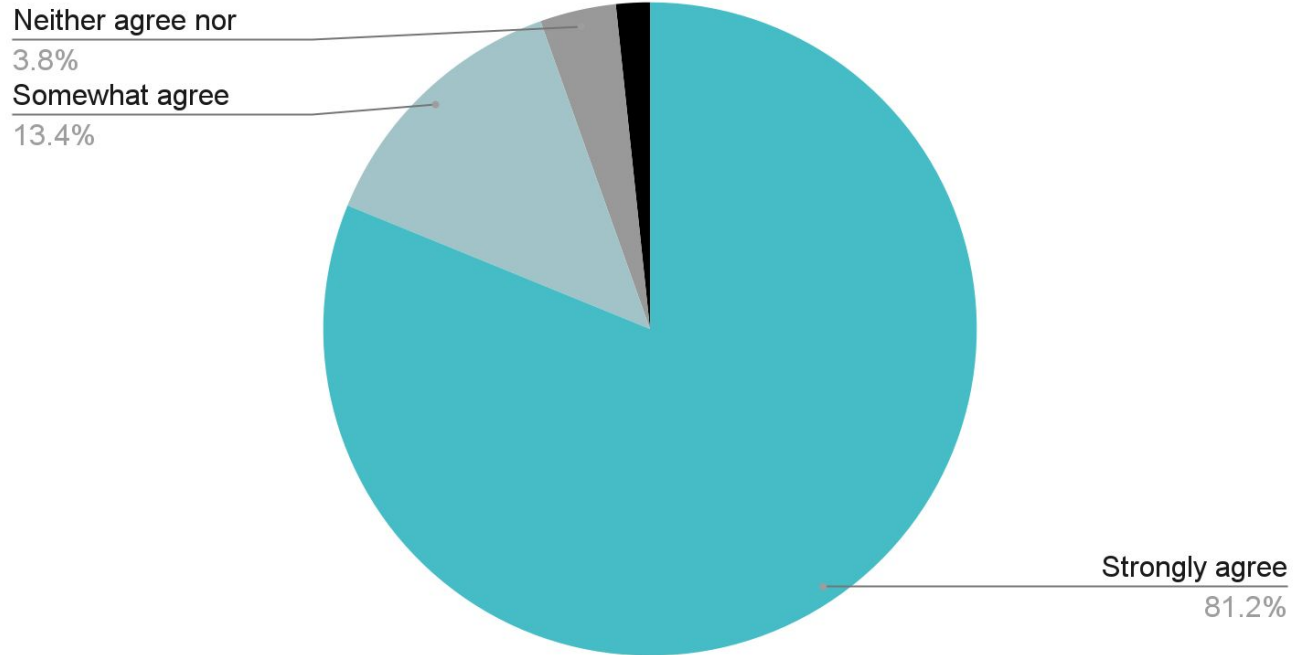
Metadata improvements:

- Insights to the quality of metadata in our repository
- Comparison with best practices and other institutions
- Opportunity to improve metadata and documentation
- Importance of mandatory use of PIDs

Visibility on campus:

- Made connections with research teams
- Elevated issue for Libraries administration
- Deepened relationships with research administrative staff
- Nationwide press exposure

Data curation by this repository adds value to the data sharing process



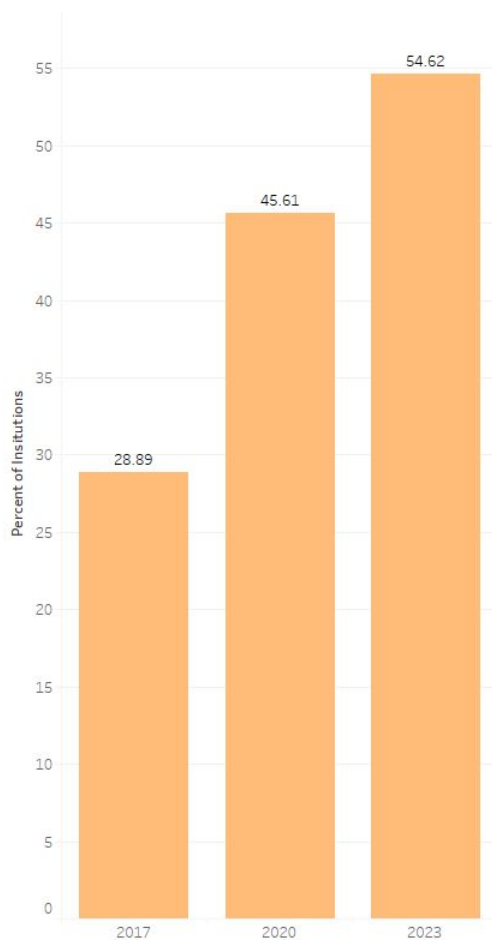
Institutional Infrastructure

Institutional Repositories and Institutional Data Repositories are important for researchers to share their data

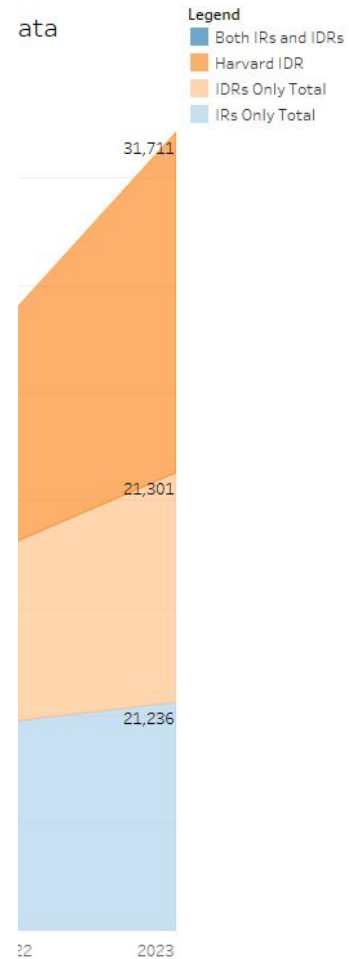
Total Count of Repositories (



Growth in Institutional Repositories (IRs) and Institutional Data Repositories (IDRs) from 2017 to 2023



ata



DCN ❤️ RADS

- Value of Curation
- Full-time staff member to support research
- Institutional repository advocacy



DCN/RADS ❤️ Joining Institutions

- Five institutions will join the DCN at no cost for 2 years to complete this work
- Have additional data on costs to institutions of different types
- DCN will gain invaluable insight into different needs for and benefits of our work

Contact us

Jake Carlson, jakecarl@buffalo.edu

Joel Herndon, joel.herndon@duke.edu

Alicia Hofelich Mohr, hofelich@umn.edu

Mikala Narlock, mnarlock@umn.edu

PI Cynthia Hudson Vitale, cvitale@arl.org

Project Manager Shawna Taylor, staylor@arl.org