

Making Research Data Publicly Accessible: Estimates of Institutional & Researcher Expenses

February 2024

This work is licensed under a Creative Commons Attribution 4.0 International License.



<https://doi.org/10.29242/report.radsexpense2024>



Authors

Alicia Hofelich Mohr
University of Minnesota
ORCID: [0000-0002-7644-4105](https://orcid.org/0000-0002-7644-4105)

Jennifer Moore
Washington University in St. Louis
ORCID: [0000-0001-6628-6820](https://orcid.org/0000-0001-6628-6820)

Jake Carlson
University at Buffalo
ORCID: [0000-0003-2733-0969](https://orcid.org/0000-0003-2733-0969)

Jonathan Petters
Virginia Tech
ORCID: [0000-0002-0853-5814](https://orcid.org/0000-0002-0853-5814)

Lizhao Ge
George Washington University
ORCID: [0009-0005-7862-6016](https://orcid.org/0009-0005-7862-6016)

Shawna Taylor
Association of Research Libraries
ORCID: [0000-0002-9842-7867](https://orcid.org/0000-0002-9842-7867)

Joel Herndon
Duke University
ORCID: [0000-0001-9995-9040](https://orcid.org/0000-0001-9995-9040)

Cynthia Hudson Vitale
Association of Research Libraries
ORCID: [0000-0001-5581-5678](https://orcid.org/0000-0001-5581-5678)

Wendy Kozlowski
Cornell University Library
ORCID: [0000-0001-6539-3798](https://orcid.org/0000-0001-6539-3798)

This research was funded by the US National Science Foundation under grant [#2135874](https://www.nsf.gov/awardsearch/showAward?AWDNO=2135874).

Suggested citation: Hofelich Mohr, Alicia, Jake Carlson, Lizhao Ge, Joel Herndon, Wendy Kozlowski, Jennifer Moore, Jonathan Petters, Shawna Taylor, and Cynthia Hudson Vitale. *Making Research Data Publicly Accessible: Estimates of Institutional & Researcher Expense*. Washington, DC: Association of Research Libraries, February 2024.

<https://doi.org/10.29242/report.radsexpense2024>

Table of Contents

- Executive Summary.....4**
- Overview 6**
- Methodology 7**
- Institutional Administrative Expenditures 8**
 - Total Annual Expenditures by Institutional Service Area9
 - Anticipated Future Investments by Institutional Service Area.....10
- Researcher Expenses.....12**
 - Average DMS Expenses over the Grant Period 12
 - Researcher Funding Source..... 14
 - Average DMS Expenditures over the Grant Period by Discipline..... 15
 - Average DMS Expenditures over the Grant Period by Funding Agency.....17
- Scaling Institutional Expenses19**
 - Relationship between DMS Expenses and DMS Activities.....20
- Conclusion..... 23**
- Recommendations and Considerations24**
- Limitations and Future Explorations 26**
- Data Availability Statement28**
- Acknowledgements..... 29**

Executive Summary

Public access to research data is critical to advancing science and solving real-world problems. Over the last 15 years, an increasing number of US funding agencies have required the management and broad sharing of research data¹ and other related research outputs to accelerate the impacts of their investments. In response, many academic institutions have built and maintained services and infrastructure to comply with these policies. These resources are often spread across the institution, housed in various administrative units, such as campus IT, the university libraries, and the research office, among others. Given this distributed nature, the total institutional cost of public access to research data has not been well understood. This lack of data on expenses makes strategic planning for research data services and support particularly difficult given the increased scope of research data management and sharing (DMS) services expected as a result of the 2022 Office of Science and Technology Policy (OSTP) memo, “Ensuring Free, Immediate, and Equitable Access to Federally Funded Research.”²

To begin to address this need, research conducted as part of the US National Science Foundation (NSF) grant “Completing the Lifecycle: Developing Evidence-Based Models of Research Data Sharing” (#2135874) collected expense information from six academic institutions. Expense information on data management and sharing activities was collected from institutional units, including the library, IT, research offices, and other institutes and centers for fiscal year 2021–2022. We also retrospectively assessed expenses for federally funded

1 John P. Holdren, “Expanding Public Access to the Results of Federally Funded Research,” The White House, February 22, 2013, <https://obamawhitehouse.archives.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research>.

2 Alondra Nelson, “Ensuring Free, Immediate, and Equitable Access to Federally Funded Research,” Office of Science and Technology Policy, Executive Office of the President, August 25, 2022, <https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf>.

researchers over the life cycle of selected federal awards that required compliance with DMS requirements. Together, these costs represent total institutional expenses for implementing data management and sharing.

Our findings indicate:

- **\$2,500,000**—Average combined annual institutional expense for DMS for researchers and campus service providers (Institutional expenses ranged from approximately \$800,000 to over \$6,000,000.)
- **\$750,000**—Average yearly cost for DMS for all institutional units including the library, IT, research office, and other institutional institutes and centers that have DMS capabilities
- **\$29,800**—Average cost directly incurred by researchers per funded research project for DMS
- **6%**—Average percent of overall grant award that was used by researchers for DMS
- Average DMS expense by researchers per funded project by funding agency (US Department of Energy excluded for small sample size; n=3):
 - **\$36,000**—US National Institutes of Health
 - **\$19,000**—US National Science Foundation

Together, our findings indicate the DMS activities are associated with non-trivial expenses to both researchers and institutions. We include recommendations for institutions, researchers, and considerations for funding agencies to most effectively and efficiently meet DMS policy requirements.

Overview

Academic institutions have made significant investments to support public access to research data requirements, yet little to no data about these services, infrastructure, and costs currently exist or are widely shared. For public access to research data to be optimized, funding agencies, institutions, and organizations must better understand the investments made by institutions and individual researchers toward meeting these requirements.

This mixed-methods study was funded by the US National Science Foundation (grant #2135874). The Association of Research Libraries (ARL) and six research-intensive academic institutions—Cornell University, Duke University, University of Michigan, University of Minnesota, Virginia Tech, and Washington University in St. Louis—used surveys and interviews to provide an initial examination of institutional expenses for public access to research data. Due to the breadth and heterogeneity of research data and funding, we scoped this work to three US federal funding agencies (Department of Energy, National Institutes of Health, and National Science Foundation) and five disciplinary areas (biomedical sciences, environmental science, materials science, physics, and psychology).

Methodology

From October 3, 2022, to December 5, 2022, the research team distributed surveys to two populations at their respective institutions. The surveys included questions on expenses for research data management and sharing:

1. Institutional units that offer DMS services (n = 69 respondents; 58 provided expense data; average response rate of 50% across institutions). These units included offices within research offices, information technology, specialized institutes and centers, and research libraries.
2. Funded researchers who completed a Department of Energy (DOE), National Institutes of Health (NIH), or National Science Foundation (NSF) grant between 2013 and 2022 in one of the following disciplines: biomedical sciences, environmental science, materials science, physics, or psychology (n = 255 respondents; 91 provided expense data; average response rate of 8.4% across institutions).

A full explanation of our methodology and a summary of response rates may be found in [*Realities of Academic Data Sharing \(RADS\) Initiative: Research Methodology 2022–2023 Surveys and Interviews*](#). Links to our survey instruments are in the “Research Instruments” section of the aforementioned report, and are available on the [RADS webpage](#).

Institutional Administrative Expenditures

DMS services require coordination across numerous institutional offices as well as researcher team time and investment to meet the various federal, publisher, and institutional policies.

To assess the expense of providing DMS services within an institution, our research analyzed the yearly staffing and infrastructure expenditures of institution-based offices. Our survey asked about staffing expenses (such as number of staff involved in DMS support, percent effort, and salaries) and annual expenses for infrastructure and technology (such as hardware, software, and licenses)³ to provide services for [DMS Activities](#).⁴ To normalize the variability in how offices and reporting lines are structured within the research universities, units were grouped into four main institutional service areas:

- **IT Offices**—Information technology, including research computing
- **Libraries**—University libraries and archives
- **Institutes and Centers**—Specialized institutes and centers, including collegiate and departmental research offices
- **Research Offices**—Central research offices, including compliance offices, legal offices, and grants management

Average expenses were taken across individual units in each of the four institutional service areas to create the total expense for each institutional service area.

3 These activities were categorized by phase and generally correspond with grant management and data life cycles and include: 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 (hereinafter referred to as “Data Retention”); and Project Closeout and Compliance.

4 The surveys used version 1 of the RADS Public Access DMS Activities (linked in-text). At the time of writing, version 3 of these activities, released in December 2023, was the most up-to-date version. <https://doi.org/10.29242/radsdmsactivities2023>.

Total Annual Expenditures by Institutional Service Area

The average annual expenditures for DMS staffing and infrastructure for administrative areas across our institutions was approximately \$750,000.

When we break this annual expense down by service area, libraries had the greatest average annual cost, followed by IT. These expenditures included both staffing and infrastructure (such as technology) to provide DMS services.

Table 1: Average annual costs to support research DMS activities across the six institutions, by institutional service area (values are rounded to three significant figures).

Service Provider	Total Average Annual Cost	Standard Deviation
Libraries	\$303,000	\$164,000
IT Offices	\$249,000	\$218,000
Institutes & Centers	\$94,900	\$76,500
Research Offices	\$93,800	\$60,400
Total Cost (Sum)	\$740,000	\$107,000

These results align with findings from COGR’s 2023 report, “[The Cost of Complying with the New NIH DMS Policy](#),” which reported that libraries and IT had the greatest financial burden for complying with the NIH data management and policy.

Figure 1 breaks down total annual expenses for each institutional service area by staffing and infrastructure. This figure shows that libraries had higher staffing expenses than other areas, while infrastructure expenses were fairly consistent across all service areas.

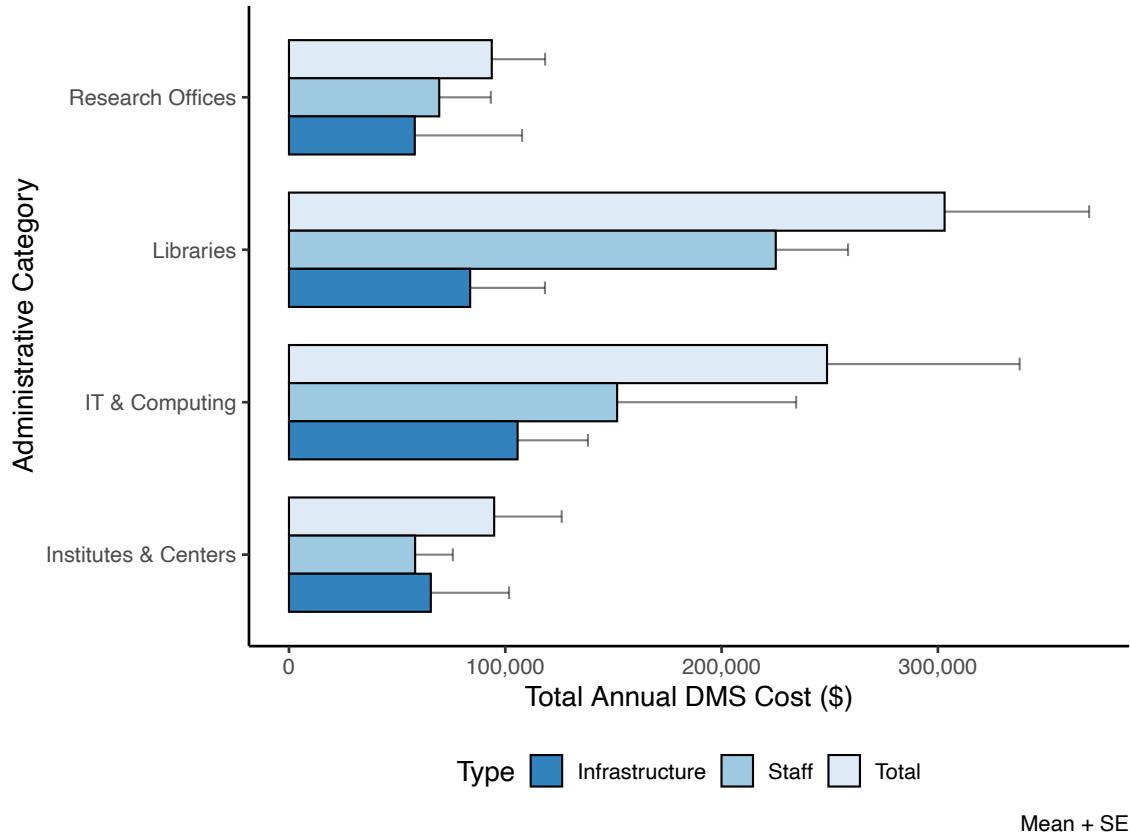


Figure 1: Breakdown of average annual cost for DMS services by staffing and infrastructure for institutional service areas.

Anticipated Future Investments by Institutional Service Area

Libraries, IT, and Research units reported anticipating substantial future investments in staff or infrastructure to support DMS services in the next five years.

In addition to assessing the current annual expenses to support DMS activities, we also asked about areas (by phase) in which institutional offices anticipate making future investments. Administrators were asked to anticipate future expenses for major categories of [DMS Activities](#). Figure 2 shows the additional level of investment (staff and infrastructure) respondents anticipated making within the next five years, relative to their current expenses.

Nearly half (44%) of the respondents from libraries anticipated substantial future investments (over 100% of their current investments)

in Data Retention, with a smaller proportion reporting similar investments in Data Collection, Storage, and Management and in Making Data Broadly Available.

Approximately 13% of IT offices reported anticipating substantial investments (over 100% of their current investments) in Data Retention, and a smaller proportion (7%) reported similar investments in Data Collection and Management. A small number of Research Offices anticipated substantial investments in Planning, in Making Data Broadly Available, and in Project Closeout phases.

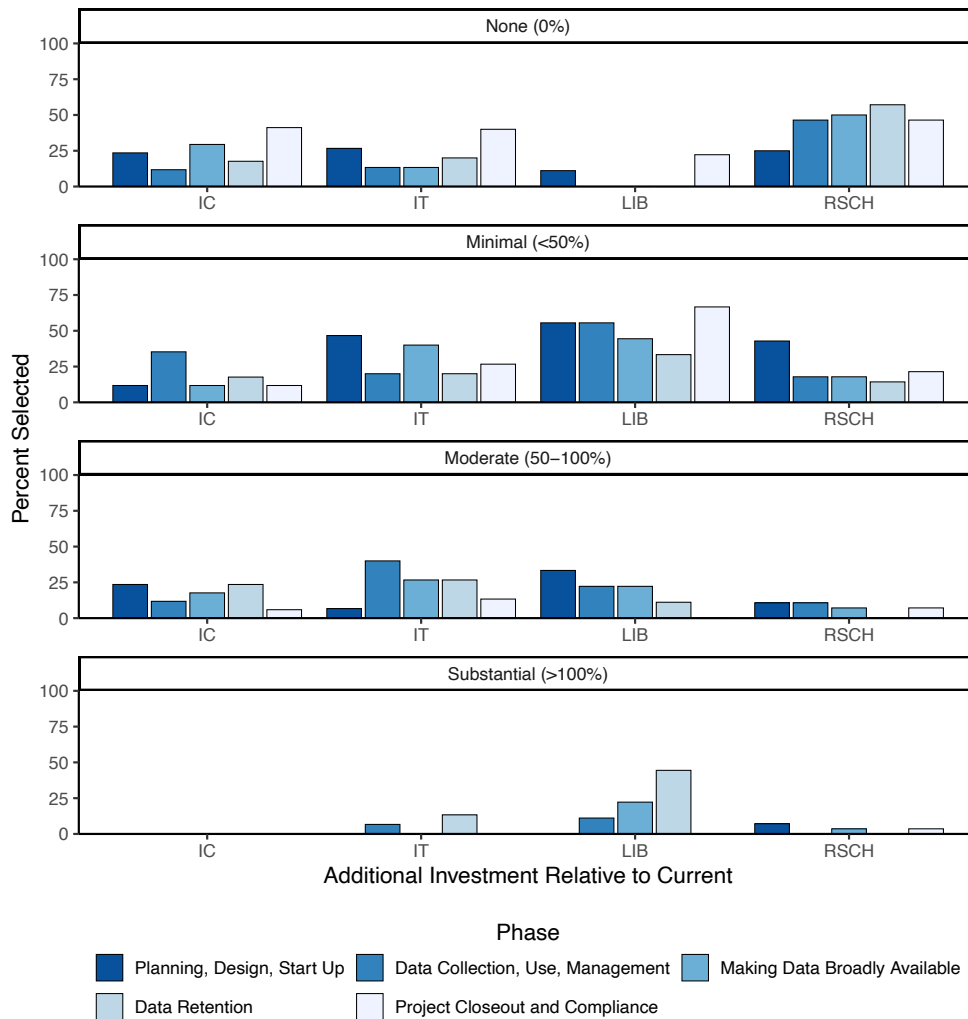


Figure 2. Anticipated level of future investment (staff or infrastructure) within the next five years relative to current expenses by DMS category. Percentage of respondents who selected each response are grouped by their institutional service area (IC = Institutes and Centers; IT = Information Technology; LIB = Libraries; RSCH = Research Offices).

Researcher Expenses

Compliance for DMS policies is often fulfilled by a combination of institutional units and researchers who receive grant awards. Therefore, in addition to assessing the institutional expenses for infrastructure, staffing, and services to support DMS, we also assessed the expenses incurred by researchers in meeting federal DMS policies.

To assess researcher expenses, we asked principal investigators (PIs) on completed grants about the expenses they incurred for DMS activities, including technical infrastructure (such as hardware, storage, and software) and staffing (such as time and pay for students and data managers). We also asked whether these expenses were funded solely through the grant, or were supplemented by lab or departmental funding. Finally, we assessed the relationship between expenses and the DMS activities reported by researchers. Our analysis looked retrospectively at grants that were completed after the issue of the 2013 Office of Science and Technology Policy (OSTP) memorandum, [“Increasing Access to the Results of Federally Funded Scientific Research,”](#) a milestone mandate for the creation of US federal agency policies for the management of data resulting from federally funded research.

Average DMS Expenses over the Grant Period

The average expenditure for researchers for DMS services and infrastructure throughout the grant period was \$29,800 or 5.83% of their award amount.

Average expenditures across all respondents for DMS during the entire grant period was \$29,800, which equated to nearly 6% of the total award amount. To capture differences across grant sizes, we divided our sample by percentile of overall grant award amount. Research grant awards in the upper 25% of total grant amount had higher expenditures for DMS than those in the middle 50% or the lower 25% (Table 2).

However, while absolute expenses were higher for those with a larger overall grant award, these expenses were a smaller percentage of the grant total. Expenses for the smaller grants in our sample (which had an average award amount of \$149,000) had average DMS expenses of a little over 15% of their total award amount, with two-thirds of this number going to staff time for DMS activities. This finding further suggests that DMS activities have non-trivial associated expenses, even for small grant awards. Consequently, this may have implications for differences in the perceived burden of DMS requirements for PIs with smaller versus larger grant awards.

Table 2: Average DMS expenditures, by award amount percentile.

Percentile	Average Total Grant Award	Average DMS Expenditures	Percent of Total Grant Award Used for DMS		
			Total	Infrastructure	Staff
Lower 25%	\$149,000	\$17,600	15.27%	5.11%	10.38%
Middle 50%	\$639,000	\$29,200	4.25%	1.34%	2.69%
Upper 25%	\$6,560,000	\$41,000	1.05%	0.33%	0.73%
All Combined	\$201,000	\$29,800	5.83%	2.01%	3.89%

The striking differences in overall expense versus award proportion can be seen in Figure 3. This figure shows the total average DMS expenditures and percent of total award used for DMS by the percentile of average total grant. While total DMS expenses were higher for larger awards, these expenses were a considerably smaller proportion of the overall grant compared to projects of smaller award sizes.

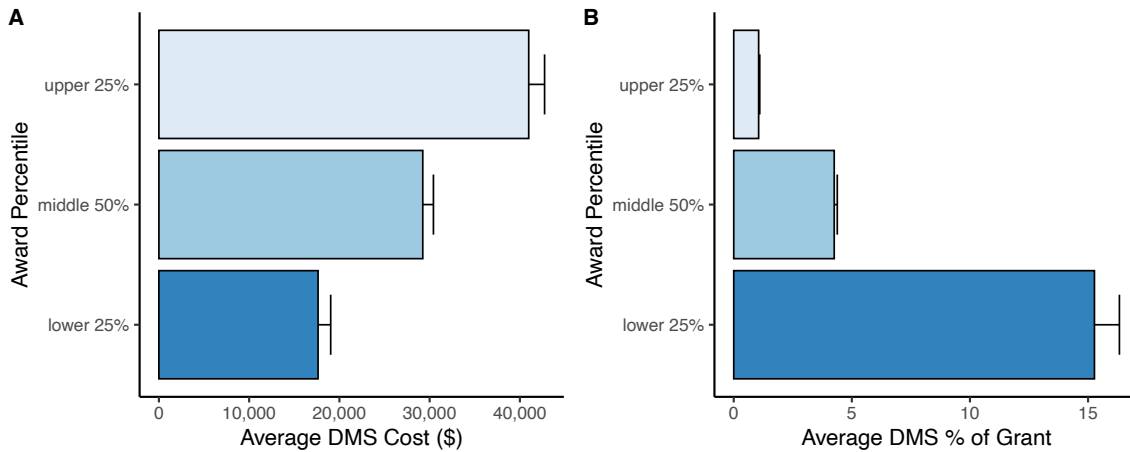


Figure 3: Side-by-side comparison of average DMS cost and average DMS expenses as a percentage of the grant when bucketed into overall award percentiles.

Researcher Funding Source

In addition to collecting information on expenses, we also asked researchers what the budgetary sources were for their DMS expenses. As shown in Figure 4, the majority of the budget came from the grant award, but percentages also came from department, college, lab, or other funds. This figure also shows that the source of funding seems independent of the total grant award amount. In other words, researchers who received smaller and larger awards both used funds outside of the grant to cover some of their DMS expenses.

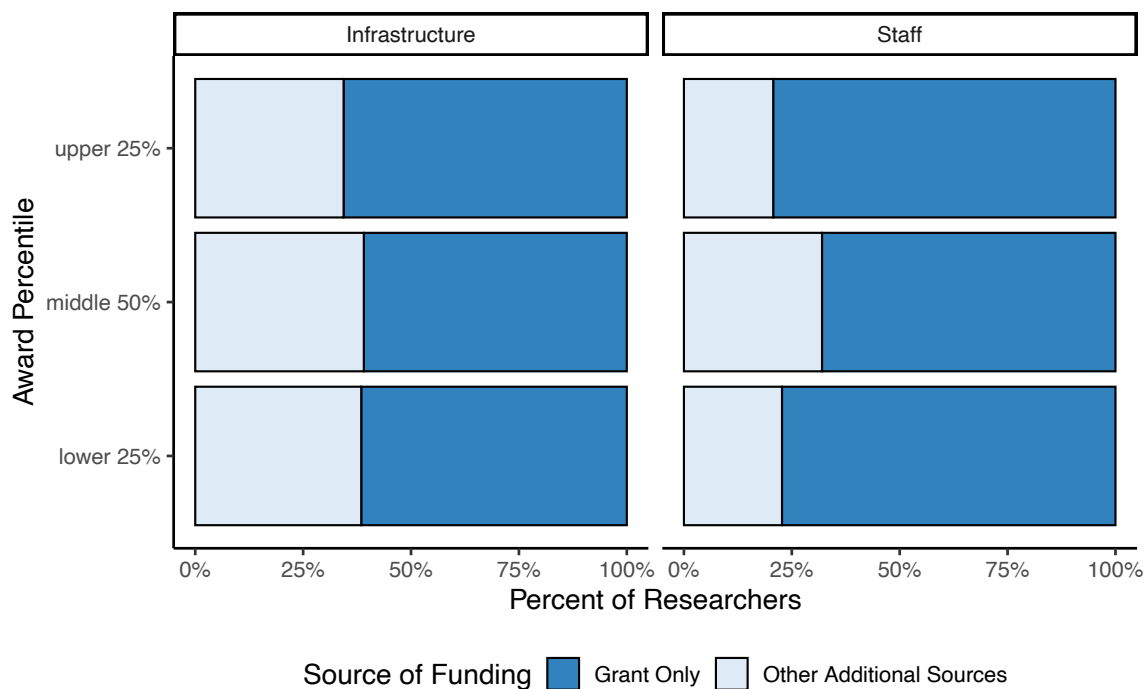


Figure 4: Comparison of where researchers received funding for data management and sharing activities by award percentile, as reported for projects completed from 2013 to 2022.

Average DMS Expenditures over the Grant Period by Discipline

Our research collected expense information retrospectively (2013–2022) across grants within five disciplinary areas: biomedical sciences, environmental science, materials science, physics, and psychology.

We intentionally limited our search to these disciplines to capture the breadth of research conducted at each institution. However, as seen in Table 3 and Figure 5, there were fewer differences in DMS costs across disciplines than expected, with the exception of psychology research (Table 3). As can be seen by the standard deviation, there was considerable variability in DMS costs within a discipline. However, it is important to note that the sample sizes when broken out by discipline are quite small, so caution must be used when drawing inferences about disciplinary differences based on this data.

Table 3: Average data management and sharing expenditures, by average disciplinary grant award amounts. Note: Materials Science had fewer than 5 respondents, so their averages are not shown.

Discipline	Average Total Grant Award	Average DMS Expenditures	Standard Deviation	N
Materials Science	--	--	--	<5
Physics	\$304,000	\$11,900	\$20,100	6
Environmental Science	\$500,000	\$14,000	\$9,380	14
Biomedical Sciences	\$3,130,000	\$30,600	\$35,100	44
Multidiscipline	\$952,000	\$35,400	\$75,800	18
Psychology	\$163,000	\$71,500	\$79,100	6

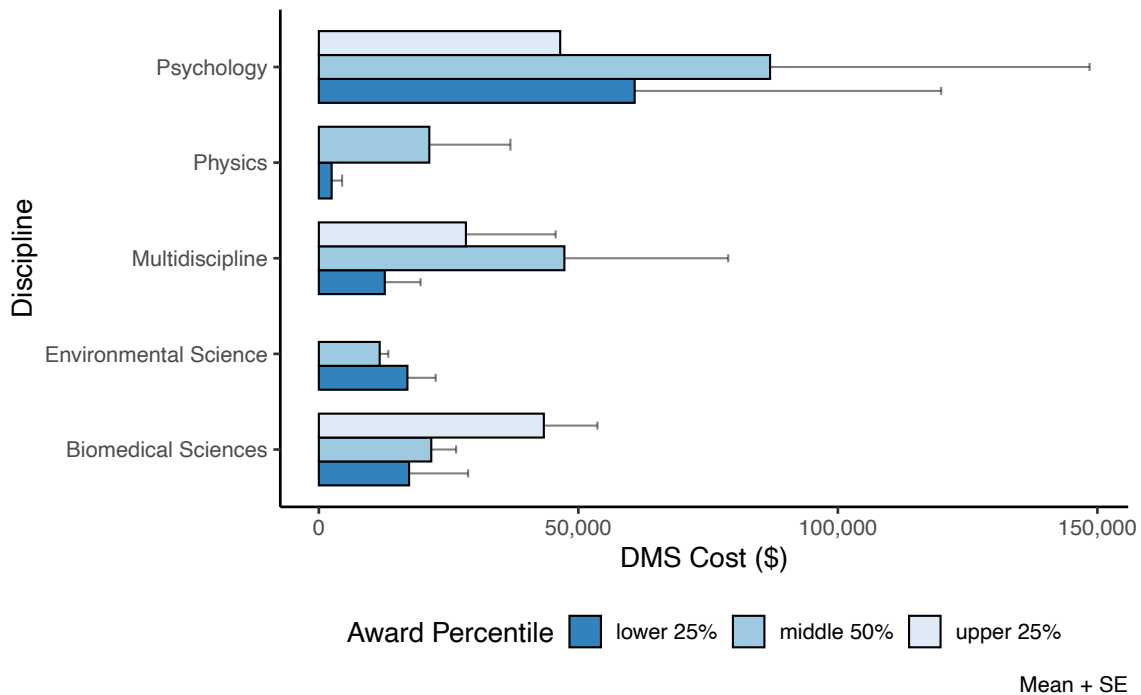


Figure 5: Average DMS expenditures by discipline and percentile of total awarded grant amount. As award percentile was calculated across the entire sample, not all disciplines had awards in each percentile group (for example, there were no awards in Physics or Environmental Science in the upper 25% of our sample).

Average DMS Expenditures over the Grant Period by Funding Agency

- **NIH-funded researchers reported, on average, a total DMS expense of \$36,000 across the lifetime of the project.**
- **NSF-funded researchers reported, on average, a total DMS expense of \$19,000 across the lifetime of the project.**
- **Due to sample sizes <5, expenses specific to DOE-funded researchers were not analyzed.**

We analyzed DMS expense data by funding agency and total grant award. While these expenses were not necessarily directly budgeted in the initial grant proposals, nor even completely directly covered by grant funds, this analysis is provided to aid institutions and funders in better anticipating future DMS costs in agency-specific grant proposals⁵.

When we analyzed the results according to percentiles of funding, we found that the expenses for DMS were fairly consistent between the lower 25% and middle 50% of total average grant awards within each funder. For NIH, these expenses were around \$28,000 and for NSF these expenses were around \$12,000 (see Table 4 and Figure 6). However, it is worth noting that these expenses correspond to dramatically different percentages of the grant award for these two groups, with smaller awards facing a much higher burden.

For grant awards in the top 25% of the sample, the costs for DMS activities were nearly double, and on average represented a much smaller percentage of the grant.

⁵ Please note that some of the DMS activities (such as those included in the Planning phase) researchers provided expense information for may not be considered as direct costs in grant budgets, but were included when calculating the average percent of grant award used for DMS.

Table 4: Average award percent dedicated to data management and sharing expenses by funder.

Funder	Percentile	Average Total Grant Award \$	Average DMS Expenditures	Average Percent of Grant Award Used for DMS	N
NIH	Lower 25%	\$422,000	\$28,400	11.30%	14
NIH	Middle 50%	\$850,000	\$28,700	3.14%	27
NIH	Upper 25%	\$1,900,000	\$57,100	1.26%	15
NSF	Lower 25%	\$150,000	\$12,400	32.30%	5
NSF	Middle 50%	\$300,000	\$10,700	3.42%	16
NSF	Upper 25%	\$470,000	\$32,400	4.02%	11

Institutions and data management professionals may want to consider offering DMS budgeting advice in alignment with these expense percentages by funding agency and average grant award amount.

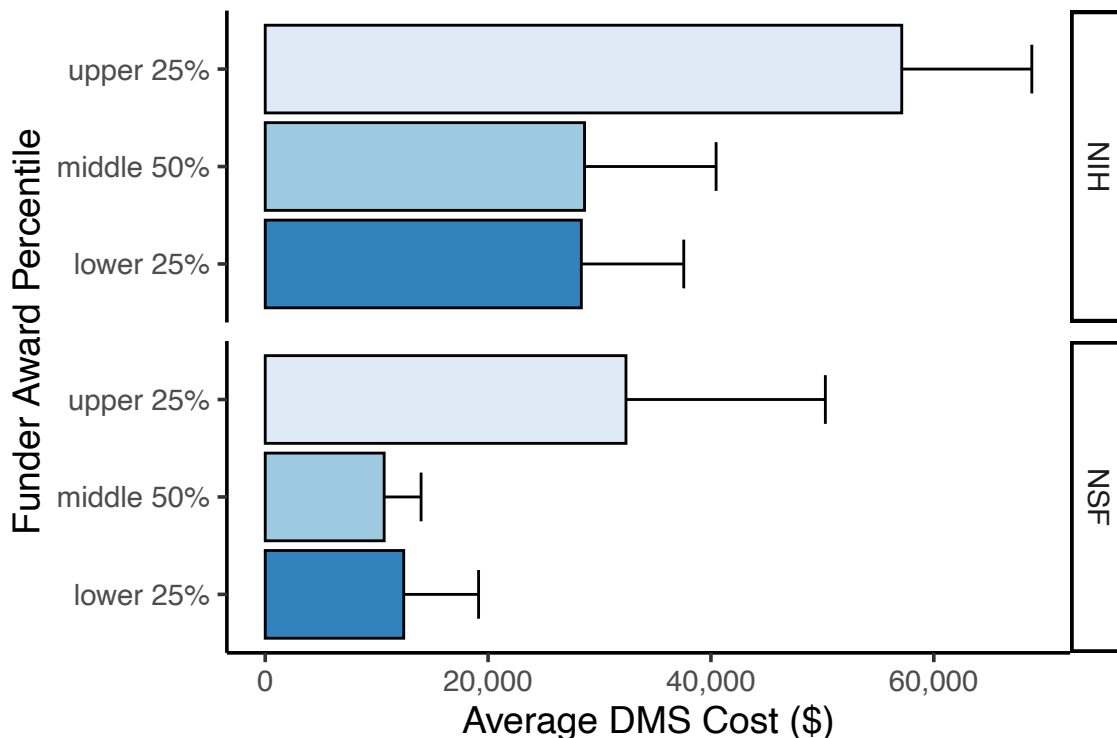


Figure 6: Average DMS expenditures by funder and percentile of total awarded grant amount (percentiles here are calculated within funder, rather than across all awards). DOE is excluded due to small sample size (n=3).

Scaling Institutional Expenses

The average annual expenditure for institutions (combining administrative costs with researcher costs) for DMS services and infrastructure was approximately \$2,500,000, with a range between \$800,000 and \$6,000,000, among the institutions included in this research.

Institutes of higher education have vastly different organizational structures, infrastructure, and services. Our estimated expenses for individual administrative units and funded researchers will scale within a given institution based on the number of offices and funded projects. To estimate how these expenses may scale into a total annual expenditure for our individual institutions, we combined the administrative expenses (summed across the four institutional service areas) with the average per-project cost for a research award multiplied by the average number of funded NIH, NSF, and DOE projects per year at that institution over our time period of interest. The combined yearly institutional expenses for DMS ranged between \$800,000 and \$6,000,000 at the six institutions surveyed.

Table 5: Total annual institutional data management and sharing expenses, by researcher and administrative costs. Note: Participating institutions have been de-identified.

Institution	Total Federal Sponsored Research Expenditures ⁶	Total Researcher Cost	Total Administrative Cost	Total Cost
A	\$636,000,000	\$283,000	\$525,000	\$808,000
B	\$240,000,000	\$217,000	\$680,000	\$897,000
C	\$681,000,000	\$1,160,000	\$398,000	\$1,558,000
D	\$902,000,000	\$2,060,000	\$510,000	\$2,570,000
E	\$652,000,000	\$2,120,000	\$958,000	\$3,078,000
F	\$971,000,000	\$5,040,000	\$1,030,000	\$6,070,000

⁶ Total federal sponsored research expenditures data were collected from the Higher Education Research and Development (HERD) Survey 2022, National Center for Science and Engineering Statistics, US National Science Foundation, accessed January 17, 2024, <https://nces.nsf.gov/surveys/higher-education-research-development/2022>. These include amounts from separate medical schools.

The variances between these expenses are likely due to a myriad of factors, including:

- Total researcher costs are a driver in the overall institutional cost. Consequently, an institution that receives a greater number of grant awards will have a higher institutional cost.
- The schools that make up the institution also have an impact on these expenses. Institutions C, D, E, and F in Table 5 all have medical schools that were included in this analysis. The other institutions (A and B) have separate medical schools that are independently awarded grants, and these awards were not included in our population.

Relationship between DMS Expenses and DMS Activities

Our analysis found that researchers who completed DMS activities independently had an increased DMS expense of approximately \$2,700 per activity.

As shown in Figure 7, the number of [DMS activities](#) completed by the researcher on their own (without support from institutional service units or external support, such as a disciplinary repository) was significantly associated with greater total DMS costs.⁷ Similar regression analyses found no significant association between DMS expenses and DMS activities done with institutional support, but observed a negative estimated relationship, as seen in Figure 7. We found no statistically significant relationship between DMS activities completed with external support.

⁷ $b = 2735.5$, $t(88) = 2.96$, $p = .004$; adjusted $R^2 = 0.08$; all other $p > .28$

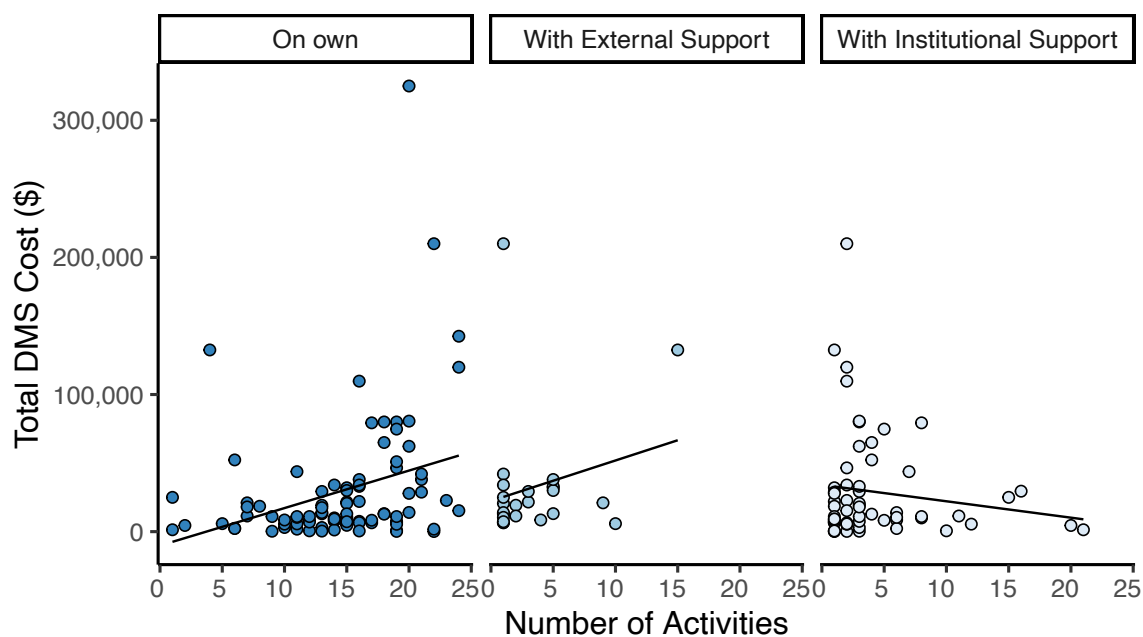


Figure 7: Relationship between DMS expenses and who is performing the DMS activity.

Average DMS expenses by data-sharing location (institutional, generalized, or specialized repositories, and personal websites)

On average, researchers spent approximately \$7,200 for DMS when they shared their data in an institutional repository or institutional data repository versus \$35,000 for researchers who used other platforms, such as generalized and specialized repositories and personal websites.

As shown in Figure 8, the average DMS expenses of principal investigators (PIs) who use an institutional repository were considerably lower than those of researchers who instead used other platforms and applications for sharing their data. The DMS expense for a researcher who shared some or all of their data using an institutional repository or institutional data repository⁸ averaged \$7,200 compared with \$35,000 for those researchers who used methods other than an

⁸ Most participants who used an institutional repository (IR) did so as their only sharing mechanism (n=10); others who used an IR also shared their data by request, in supplementary materials, or in a personal website (n=4). Two participants who used an IR also used a disciplinary repository.

institutional repository. While our data do not point to why this difference might be so stark, future research is needed to explore the roles of institutional repositories and how researchers make decisions about data-sharing locations.

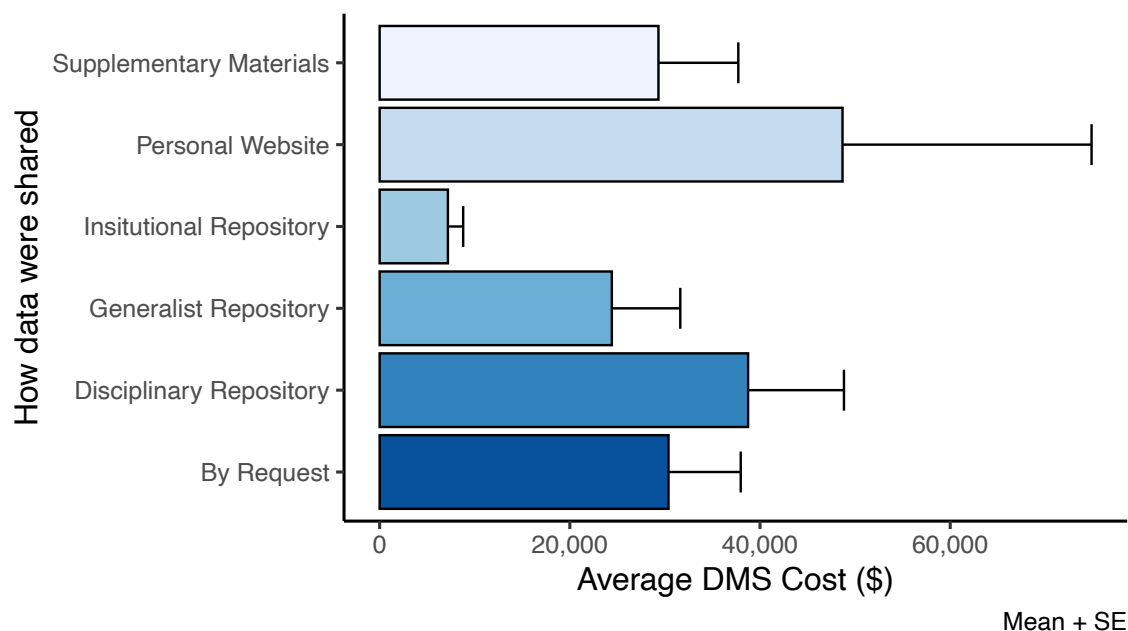


Figure 8: Average DMS expenses per funded research project by how research data were shared. In cases where multiple options were selected, the researcher’s DMS expenses were counted in each category.

Overall, we found the average total expenses for DMS costs were higher for researchers who shared their data using any of the available options other than an institutional repository.

It is important to note that these data represent total DMS expenses across the project, and are not restricted to repository expenses. Our results may suggest more fiscal efficiency when researchers use institutional services for data sharing, such as an institutional repository. However, in these cases it is likely that institutions are absorbing the expenses for DMS through indirects or are seeking funding elsewhere. As previously shown, library expenses for DMS (which, in all six of the institutions, include support for the institutional repositories) vary based on organization, but certainly have a cost.

Conclusion

Institutions of higher education have made significant investments in services and infrastructure to support federal policies for research data management and sharing. Our research represents a first and exploratory step in estimating the costs of DMS to both institutional offices and federally funded researchers at large US research universities. Our findings affirm that making research data publicly accessible does have a cost, which is incurred across many types of institutional service areas but is most heavily shouldered by IT units and libraries. These two areas also anticipate making substantial additional investments in these services over the next five years.

Funded researchers within these institutions also face individual project costs, which are an especially large proportion of costs for smaller grant awards. Our results show that while DMS expenses increased with higher grant awards, the proportion of the award spent on DMS decreased dramatically in larger award amounts. The burden placed on researchers in covering these expenses is significant, especially for smaller awards that already have more limited funds available for science and scholarship. Our results suggest that this burden has the potential to be lessened with further engagement with institutional services; these services also have a cost, but one that is borne by institutional service areas, rather than the researcher. Future work is needed to determine how to most efficiently and effectively balance institutional and direct cost coverage of DMS expenses.

Our results are limited to a small sample of well-resourced research-focused universities. Undoubtedly, these costs and models will be different for other types of institutions. Future research using these methods can help refine these institutional cost estimates to better reflect the realities of data sharing across more diverse institution types, such as liberal arts colleges or less research-intensive institutions.

Recommendations and Considerations

Based on our results, we have laid out the following recommendations and considerations for researchers, institutions, and funding agencies to most effectively budget for and meet data management and sharing requirements:

Researchers:

- Seek out and make use of institutional DMS services and infrastructure to decrease overall expenses.
- Recognize that research data management and sharing has a baseline cost in terms of time and infrastructure. Ensure these costs are sufficiently captured in grant budgets.

Institutions:

- Invest in research data management and sharing services and infrastructure, such as institutional repositories.
- Engage in broader outreach to their researchers and departments to advertise library and IT resources currently available.
- Reward researchers for doing DMS as credit towards career growth, such as promotion and tenure.
- Leverage expense models to conduct local DMS expense analyses in order to understand local expenses and make strategic decisions about DMS infrastructure investments, especially if coordination across units is possible.
- Develop formal partnerships between key DMS administrative units on campus to increase costs efficiencies and administrative coordination for DMS support, such as adequately capturing DMS costs in grant budgets.
- Invest in comprehensive communication plans and strategies when sharing information about institutional DMS services and infrastructure.
- Conduct cost-efficiency studies to evaluate offering centralized DMS services through locally hosted or vended solutions.

Funding Agencies:

- Invest in research data management and sharing services and tools, such as the [DMPTool](#).
- Recognize institutional repositories, especially those that meet the “[Desirable Characteristics of Data Repositories for Federally Funded Research](#),” as a legitimate method for sharing research findings and data.
- Recognize that the cost of research data management and sharing is shared among researchers, institutions, and federal agencies. An equitable distribution of these expenses across stakeholders is important to advance scientific research and ensure the United States continues to be a leader in research. This is especially important for small awards, as the costs of DMS encompass a much higher proportion of these budgets.
- Acknowledge that time and effort for research data management and sharing are large components of researcher expenses. Consider models that prioritize the professionalization of research data managers to improve both the cost efficiency and quality of research data that is shared.

Limitations and Future Explorations

We recognize and strongly acknowledge that the data and results of this research have under-reported the actual expenses for research data management and sharing for the following reasons:

Limitations of Survey Instrument

- For research projects, staff cost was reported based on annual effort on DMS activities and salary. Because we did not capture the number of grant years the specific staff were funded, the salary and percent effort was taken for one year, which likely underestimates the true expense.
- Our methods asked researchers which institution-based services they used, not whether the PIs themselves, graduate students, or project staff completed each activity. Therefore, we do not know the distribution of work and associated cost across the DMS activities and involved personnel.

Limitations of Time Frame and Sample

- We identified researchers who had completed grants within a 10-year period (2013–2022) to align with DMS requirements and identify projects in which data have been shared. However, we found it was difficult for researchers to recollect exact grant expenditures over this time period, which likely contributed to lower cost estimates than were actually incurred. Additionally, researchers with multiple grants during this time period often reported having trouble determining expenses specifically for the identified grant.
- Our research was limited to research funded by DOE, NIH, and NSF. While this comprises a significant portion of the overall funded research happening on an academic campus, it does not capture **all** funded research. Given the variability we found between NIH and NSF data management and sharing expenses, we were cautious to make generalizations about funding agency average expenses. Additional research should apply similar protocols to research projects funded by other federal agencies.

- On the administrative side, we estimated annual costs across categories of institutional offices. Many universities have multiple offices in each of these categories that have independent budgets and costs. Our results do not account for the number of offices within each institution. If this was accounted for, it is likely institutional costs would be higher than what is reported here.

General Challenges

- The distinction between research data management and sharing activities and “good scientific practices” was difficult for researchers and administrators to parse out. While our research produced a set of activities that comprised DMS, budgets are not structured in this manner and, in subsequent interviews, researchers indicated a significant overlap between DMS practices and “good science” practices.
- Responses provided in follow-up interviews demonstrated the difficulty for both researchers and administrators in estimating their costs in terms of time. Some researchers and administrators did not include their own time in their estimates of cost, meaning that PI effort on DMS activities may not be reflected in our expense estimates. Additionally we saw high nonresponse rates for the financial questions for researcher respondents, with some volunteering that they did not know how to answer those questions. The small sample size also prevented much interpretation of disciplinary differences, as we had counts less than 20 in every discipline except biomedical sciences.

Despite these limitations, our research fills a critical gap in knowledge and represents a first step in understanding the costs of data management and sharing to both institutional offices and federally funded researchers in large US research universities. Our findings can be used to help institutions make decisions about strategic investments and services for DMS, as well as to help researchers develop initial budgets for new and forthcoming DMS requirements. Our methodology can also be used as a starting point for other institutions to evaluate their own costs and service models for supporting DMS activities and policy requirements.

Data Availability Statement

De-identified response data and data dictionaries for both the Institutional Infrastructure and Researcher surveys are located in the Washington University in St. Louis WashU Research Data (WURD) repository, at <https://doi.org/10.7936/6RXS-103654>.

Acknowledgements

Funding for this research and the Realities of Academic Data Sharing (RADS) initiative was provided by the US National Science Foundation (NSF), [award #2135874](#), EAGER grant: “Completing the Lifecycle: Developing Evidence-Based Models of Research Data Sharing.” We would especially like to thank Martin Halbert, science advisor for public access at NSF, for his considerations regarding interoperable research infrastructure, data, and metadata. We would also like to thank the members of the Data Curation Network (DCN) for engagement and discussion around these topics. Finally, we’d like to thank reviewers of the draft report for their feedback, including Anurupa Dev, Association of American Medical Colleges; Jason Gerson, Patient-Centered Outcomes Research Institute; David Kennedy, COGR; Christopher Marcum, US Office of Management and Budget; Kacy Redd, Association of Public and Land-grant Universities; Krystal Toups, COGR; Tyler Walters, Virginia Tech; and Martha Whitehead, Harvard University.