LG-254930-OLS-23, Association of Research Libraries

Institutional Functional and Cost Models for Public Access to Research Data
IMLS NLG Proposal - Association of Research Libraries

Institutional Functional and Cost Models for Public Access to Research Data

The Association of Research Libraries (ARL), along with the Washington University in St. Louis, the University of Minnesota, Duke University, all of whom are members of the Data Curation Network, request $744,650 for a 3-year National Leadership Grant (NLG) in the Applied Research category in alignment with Goal 3, Objective 3.1 of the NLG program for Libraries. Investigators will conduct critical research on the economics of investments in public access to research data. Our research will produce data, information, and models for libraries, institutions, policy makers, and funding agencies to inform strategic decision-making about building and maintaining research data related services, technology, and infrastructure.

Project Justification. Public access to research data is critical to advancing science and solving real world problems. In recent years a number of funding agencies have required the management and broad sharing of research data[1] and other related research outputs to accelerate the impacts of their investments. Many research institutions have developed and launched infrastructure to support faculty in meeting these requirements. These services and infrastructures are often spread across the institution, housed in various administrative units, such as campus IT, the university libraries, and the research office, among others. While many connections and points of collaborations have been established within institutions to support broad access to research data, these collaborations are often informal or ad-hoc. Given the growth in demand, its distributed nature, and functional nuance, the true cost of public access to research data is not well understood. There are many unknowns about the institutional landscape for funded research data sharing, which hampers collaborations and the institutional ability to plan and budget appropriately.

Academic and research libraries have made significant investments to support public access to research data requirements; yet have little comparative data about these services, infrastructure, and costs. For public access to research data to be optimized, funding agencies, institutions, and organizations must understand the local landscape and required investments. This research builds on previous funded research examining the institutional costs and expenses for federally-mandated data sharing.

This research will answer the following questions:
1. What are the different service and cost models institutions put in place to support research data management and sharing policies?
2. What is the direct expense to the institution, and specifically the academic library, to implement federally mandated data sharing policies?
3. What is the cost to the researcher to comply with the funded research data sharing policies?

To answer these questions we will build upon the work plan and mixed-methods developed in the [Realities of Academic Data Sharing](#) (RADS) (NSF# #2135874) research, led by ARL and the Data Curation Network (DCN), that focused on understanding where researchers share their data, the completeness of the metadata supporting that data, and institutional expenses for public access to research data at six academic institutions in five disciplinary areas. By its very nature the RADS project was meant to be exploratory and transformative. Building upon this prior work, the goal of this IMLS research is to **accelerate our knowledge and understanding of expenses and service models for a wider range of disciplines and academic institutions to inform strategic investments for research libraries and institutions.**

The DCN is a membership network[2] composed of seventeen organizations with well-established public access data repositories (15 academic libraries, the Michael J. Fox Foundation, and Dryad), which share functional and subject expertise to facilitate the more robust management and curation of research data. In total, these institutions were awarded

---

[1] Office of Science and Technology Policy, "Expanding Public Access to the Results of Federally Funded Research." February 22, 2013. Accessed: https://obamawhitehouse.archives.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research
[2] Data Curation Network: https://datacurationnetwork.org/

billions in federal and private funding in 2021 alone[3]. These DCN members, ARL, and other research and academic libraries participating in this project have a strategic advantage to: 1) map and articulate public access to research data services across their campuses, 2) leverage existing campus relationships to gather financial information related to federally mandated data sharing, and 3) develop relationships across campus that bolster data sharing support and services within the libraries.

Given the heterogeneity of support and infrastructure for research data services across institutions, a single financial and service model to support public access to research data is not feasible. Depending upon the organizational structure of the institution, the maturity of the research data infrastructure, and other strategic decisions made by university administrators, research data services are different within and across institutions. Yet there are commonalities we can draw across these institutions by employing a common service and cost model.

A clear understanding of possible models and expenses is critical for institutions to appropriately meet revised and forthcoming public access to research data policies as expected from the 2022 Office of Science and Technology Policy (OSTP) memo, Ensuring Free, Immediate, and Equitable Access to Federally Funded Research. The 2022 OSTP memo specifically references the May 2022 report "Desirable Characteristics of Data Repositories for Federally Funded Research"[4], issued by the Subcommittee on Open Science of the National Science and Technology Council, and asks federal agencies to provide guidance to researchers on appropriate digital repositories to use, as digital repositories used by researchers will now have to meet the characteristics requirements referred to in this report. Although research libraries are prepared to meet these requirements[5] and provide researchers and their institutions with the needed infrastructure to support federally mandated data sharing, this support does not come without increased investments from libraries and academic institutions.

Concomitant to the OSTP 2022 public access memo, in January 2023, twelve U.S. federal agencies, including the National Science Foundation (NSF), the National Institutes of Health (NIH), and the Department of Energy (DOE), among others, joined with the White House to affirm their commitment to advancing open and equitable research through launching the Year of Open Science.[6] In doing so, these agencies pledged to spark change and inspire open science engagement to advance the adoption of open, equitable, and secure science. This commitment is a significant signal that research libraries need to increase researcher support in open science practices, especially data sharing.

Existing models for costing public access to research data, such as those developed by the Consortium of European Social Science Data Archives (CESSDA) and the Dutch Data Archive and Network Services (DANS), are scoped to specific segments of the research data sharing workflow such as the data repository or data storage, and do not span across the data sharing lifecycle or the institution. An appropriate economic model for public access to research data in the US would need to take into account not just expense and service information, but also the return on investment (ROI) or value of the item being modeled. The Collaboration to Clarify the Costs of Curation program, partially funded by the European Union, sought to develop a benefits model specifically for digital preservation (see Item 4 in the Appendix). Preservation, as we know, is just one component of appropriate data sharing. Thus, while this model is useful, it would need to be expanded upon to apply for all data sharing activities. The research proposed in this project is necessary to inform future economic modeling. As data citation continues to become the norm for researchers and as research data sharing and management becomes more part of the research culture, future research will be better positioned to develop economic models. However, we can begin the process of assessment to make such economic models possible in future research.

---

[3] Nietzel, M. T. (2022, December 17). Top 25 American universities for R and D spending; Johns Hopkins #1 again. *Forbes*. https://www.forbes.com/sites/michaeltnietzel/2022/12/17/the-top-25-american-universities-based-on-r-and-d-spending-johns-hopkins-again-heads-the-list/

[4] Subcommittee on Open Science, National Science and Technology Council (2022, May). Desirable Characteristics of Data Repositories for Federally Funded Research. Executive Office of the President of the United States. https://doi.org/10.5479/10088/113528

[5] The Data Curation Network. (2022, September 6). DCN is Ready to Support Policies Resulting from OSTP Public Access Memo. https://datacurationnetwork.org/2022/09/06/dcn-is-ready-to-support-policies-resulting-from-ostp-public-access-memo/

[6] Science.gov. Open Science Announcements from Federal Agencies. Accessed 2023, March 15, https://open.science.gov/

Understanding the total institutional costs has been identified as a high priority for higher education institutions and member organizations, including Association of American Universities (AAU), the Association of Public and Land Grant Universities (APLU), and the Council on Governmental Relations (COGR).

Project Work Plan. To answer the research questions above, we propose to conduct research over a three year period and have two aims of research: 1) build upon the research started in the original RADS research to unpack institutional expenses and services for federally mandated data sharing; and 2) expand the available costing and service models for federally mandated data sharing. The full time project manager, along with the institutional co-PIs and PI, will be responsible for overseeing all aspects of this research project. Proper project management is critical for the success of the research, especially when creating the functional models, developing and disseminating surveys, managing and delivering approximately 30+ focus group meetings, collating costing data, and more.

Personnel

Administration:
- Cynthia Hudson Vitale, Director, Scholars and Scholarship, Association of Research Libraries
- Shawna Taylor, Project Manager, Association of Research Libraries
- TBH, Data Curation Network Program Manager, Association of Research Libraries

Research Team, leads from each participating institution:
- Joel Herndon, Director of the Center for Data and Visualization Sciences, Duke University
- Alicia Hofelich Mohr, Research Support Services Coordinator, University of Minnesota
- Jennifer Moore, Head of Data Services, Washington University in St. Louis

We will establish an advisory board composed of research data experts and higher education partners. The advisory board will provide feedback on the general direction of the project, circulate engagement and feedback opportunities with their communities, and provide overall advocacy for the project. These members will include:

- Jake Carlson, Director of Deep Blue Data, University of Michigan Libraries
- Lisa Johnston, Director, Data Governance, University Wisconsin, Madison
- David Kennedy, Vice President and Director of Costing & Financial Compliance, COGR
- Wendy Kozlowski, Library Lead for Research Data Services, Cornell University Library, Cornell University
- Jonathan Petters, Assistant Director, Data Management & Curation Services, Data Services, University Libraries, Virginia Tech
- Kacy Redd, Associate Vice President of Research & STEM Education, APLU
- Tobin Smith, Vice President of Science Policy and International Issues, AAU

Phase 1: Digging Deep on Institutional Expenses and Services (August 2023 - July 2026)
Building upon our initial NSF-funded research we propose to further dig deep into the expenses and investments that academic institutions have made in public access to research data infrastructure and services. Specifically, over the first 18 months of this research we will:

Phase 1, Objective 1 - Review our previous research approach for gathering institutional service and expense information

> Output(s): A common set of data management and sharing activities needed to facilitate federally mandated data sharing, refined and tested by the research data community.

As a first step in this new research, we will systematically review our previous data collection instruments, interviewer guides, and data management and sharing activities for clarity and purpose. One of the key takeaways, or lessons learned, from the original RADS research was the importance of collecting expense and research data services information in ways

that maximize the context for this information, as well as being more comprehensive and specific in who we collect the expense data from across administrative units. While surveys were beneficial for getting an overview of expenditures and activities to make research data publicly available, the follow up interviews with a small number of researchers and administrators allowed us to further refine the expense and investment data, test the terms and activities validity with researchers and administrators, and allowed us to gain an understanding of researcher and administrator data management and sharing practices.

Review data collection instruments. As part of the original research project, we developed four data collection instruments, two survey instruments and two focus group questionnaires, one of each for campus administrators and funded researchers. Our process for reviewing these instruments will be to: 1) more closely align our survey questions with previous cost or expense study analysis, in collaboration with higher education associations such as COGR; 2) analyze survey and interview responses to identify item-level non-response bias, inconsistent estimations in survey versus interview formats, and refine those questions with the goal of creating more robust methods capturing a comprehensive picture of services and expenses cross each institution.

Review and refine campus administrator and researcher population. Another lesson learned from the RADS project was the importance of interviewing and surveying comparable administrative offices across each institution. This is a particularly difficult task given the variety of institutional organizational structures. While each academic institution may have IT support, this support may be centralized, distributed, or outsourced at different academic institutions. This complicates comparing expenses and services across institutions. Additionally, our original research did not return a comprehensive view of infrastructure and services across any one institution. In this next phase of the research, we will account for this by leveraging partnerships and colleagues in higher education associations to make connections with their members. For example, we will work closely with EDUCAUSE, COGR, and AAU to identify and encourage the appropriate institutional representative to complete the data collection process. This combined, on the ground and higher education coalition approach, will support more comprehensive data collection. To review and refine the researcher population we will build on what we know about the primary research conducted at the original RADS institutions and leverage our compiled researcher contact lists. We will identify comparable (based on discipline or methodology) research projects across the three institutions to complete Phase 1: Objective 3. Our goal is to develop a framework that can be used across various institutions to systematically identify relevant administrative units and comparable researchers and research.

Review and refine DMS activities through community engagement. The understandability of the data management and sharing (DMS) activities developed during the original RADS research will also be refined and tested. The original 28 and 27 administrative and researcher-focused activities were used to describe common public access to research data activities. These terms were developed with colleagues at the COGR and the validity of them were lightly tested during the RADS data collection protocols. These categories are generally based on the research data lifecycle; a sample of them is below and the full list is available on the ARL [website]. To further align and test the understandability and validity of these activities, we will gather structured feedback from campus administrators and funded researchers while consulting established data taxonomies, such as the NIST RDaF[7] and Lifecycle Decisions for Biomedical Data[8]. Refining these DMS activities will result in a common set of data management and sharing activities required for federally mandated data sharing. This will not only illustrate to other academic libraries what data management and sharing activities are common at our participating institutions, but will also serve as a guide for new researchers[9], and demonstrate to funders the complexity of data sharing within an institution.

---

[7] https://www.nist.gov/programs-projects/research-data-framework-rdaf
[8] https://nap.nationalacademies.org/catalog/25639/life-cycle-decisions-for-biomedical-data-the-challenge-of-forecasting
[9] Note: The Office of Sponsored Projects at Dartmouth College, neither a participating RADS institution nor a DCN member, has listed the DMS activities identified in the RADS study as a resource for researchers in the pre-award phase of the grant lifecycle. OSP. (2023, March 10). NIH Data Management and Sharing Policy. Dartmouth College.
https://www.dartmouth.edu/osp/pre-award-development/writing_a_proposal/nih_data_management_and_sharing_policy.php

DMS Phase	DMS Activity
Planning, Design, and Start Up of Projects	Preparing a Data Management Plan (DMP) or Data Management and Sharing (DMS) Plan
	Preparing IRB protocols and informed consent for data sharing
	Determining storage solutions for active research data
	Identifying an appropriate repository (or repositories) for making research data broadly available

Figure 1: A sample of the public access data management sharing activities developed in the NSF RADS study and included in researcher and campus-administrator surveys and interviews.

Phase 1, Objective 2 - *Build upon and refine our existing survey, focus groups and interview protocols*

Output(s): A revised set of questions and methodological approach to acquire data on federally-funded data sharing expenses and activities at academic institutions

Based on the reviews conducted in P1: Objective 1, we will develop a revised set of questions and refined methodological approach for campus administrators and researchers. These revised approach and questions will focus on collecting information and data on the:

- Additional burden that federally mandated research data sharing services place on campus infrastructure providers
- Financing and other cost-recovery approaches for paying for public access to research data activities within libraries
- Additional services and infrastructure put in place to meet upcoming public access policies
- Identify information to contextualize spending on federally mandated data sharing in the scope of their overall budget

The refinement of the methods will focus on finding efficient and streamlined approaches to expand the depth of what is known at existing institutions, facilitate more effective information gathering for institutions included in the expansion (Phase 2), and ensure the data collected produces comparable results across institutions of varying sizes and structures.

For the applicable survey portions of the research we will leverage our previous experience and build the revised survey on the Alchemer platform. Raw identifiable data will be stored with ARL on secure servers and in Alchemer during the course of the project. Assets with personally identifiable information, specifically the survey responses, will be destroyed at the end of the project. For the interview portion of the research, interviews will be recorded via an ARL Zoom account. For more details on how the recordings and survey data will be managed, please refer to the Data Management Plan.

Phase 1, Objective 3 - *Conduct deep dive follow up surveys and interviews with campus-based administrators and researchers*

Output(s): Refined expense and activity data from funded researcher and campus administrators, with analysis of library investments

We will again take a mixed-methods approach using our refined methodology for the data collection and deep dives analyses for institutional expenses for federally mandated data sharing.

Researcher Deep Dive Analysis
Our methods for researcher inclusion in this component of the study will shift slightly from our original workflows - including both retrospective and prospective analyses on services and expenses for federally mandated data sharing requirements.

For the retrospective study we will leverage the results of our original research and conduct in-depth interviews with previously funded researchers to do deep dives analyses of their use of campus-based services and expenses to support their data sharing. In our original study we reached out to over 3000 researchers across six institutions to collect information about services and expenses through a survey format. Our response rate across all institutions was approximately 7%. While we were able to conduct follow up interviews with a small number of researchers in the original research, this research will increase the number of follow up interviews using the refined protocol meant to reduce noise in the data which will further allow us to make comparisons in expenses and service use across disciplinary and methodologically-similar research. We are requesting $75 for each researcher as an incentive for completing the interview.

In addition to the retrospective study, we will also look prospectively at researcher service use on campus and their expenses for data sharing. To do this we will recruit at least one researcher from each institution to allow us to follow along and track expenses and service utilization throughout a 2 year grant period. Potential researchers will be identified through existing collaborations with institutional research offices. Our inclusion criteria may include: 1) awarded a grant from IMLS, NIH, NSF, or DOE within the calendar year or the year before; 2) grant period of 2-3 years; 3) data management and sharing responsibility as a condition of grant funding. To monitor services and expenses over the course of the grant we will ask prospective researchers to complete a series of interviews and/or questionnaires on a quarterly basis about the activities they recently completed, campus-based services they used, and expenses they incurred for managing and sharing their research data. Please see Item 5 in the Appendix for a sample set of questions for this prospective research. We are requesting $500 as incentive for each research team to participate in this prospective study over the 2-3 year period.

Campus Administrator Deep Dive Analysis

As previously mentioned, our original RADS research took a broad and extensive approach to gathering service and expense information from campus-based administrators. In our previous research our goal was to identify as many offices on campuses as possible to understand the extent of services. Given the heterogeneity of institutions, comparable offices were difficult to ascertain. For this reason, we are scoping this phase of the research to only conduct follow up interviews with specific campus-based offices that are consistent across the three institutions.

Institution	Number of responsives offices	Number of offices identified for RADS study inclusion	Response Rate
Cornell	13	17	76.47%
Duke University	9	15	60.00%
University of Michigan	15	23	65.22%
University of Minnesota	18	33	54.55%
Virginia Tech	8	17	47.06%
Washington University in St. Louis	10	34	29.41%
Grand Total	**73**	**139**	52.52%

Figure 2: The table above illustrates the heterogeneity of the offices, identified by the project PIs in the original RADS study, which provides support to federally-funded data sharing activities.

During the RADS research project, the project team categorized and grouped the array of offices that support federally funded research data into four main categories: Libraries, Information Technology offices, Research offices (such as the Office of the Vice Chancellor for Research or Research Development), and Institutes or Research Centers. Grouping offices in these four categories was the methodology in which we were able to compare diverse services across all six institutions. Please see Item 2 in the Appendix for a sample campus map that visualizes institutional activities for the original research across all six original institutions.

This model is helpful to compare a wide range of services, but does not fully represent the extent to which offices support these activities. In this research project, further investigation, via interviews or focus groups, with only specific offices that are consistent across the three institutions, will provide more detail to better understand and represent services across-campuses to support federally-funded data sharing.

Phase 2: Expand the Expense and Service Models for Public Access to Research Data (March 2025 - July 2026)

The original RADS research on expense models and services for data sharing engaged a subset of institutions that are members of the Data Curation Network. While this was helpful in that each institution had robust research data services, it did not necessarily reflect the varying levels of resources and organizational strategies for supporting data focused research in a wide-range of institutions. To address this, we propose recruiting a more diverse set of institutions to conduct campus mappings and gather service and expense information from researchers and campus services providers.

As an incentive for participation in this research, we will offer each institution a two year membership to the Data Curation Network. This no-cost membership will include access to the DCN expertise and knowledge around research data management, repository infrastructure, and CURATE(D) workshop attendance. The project team is requesting a full time position, the DCN Program Manager, to work within the DCN to provide research data support to new members starting in year 2 of the research project. The DCN Program Manager will help build local capacity and assess research data service needs within the new library cohort institutions, as well as develop or enhance networks across institutional services providers. For each library cohort member, the DCN Program Manager may collaborate with the cohort library to offer such services as data management plan reviews, education and training to researchers or students, or data curation services. This individual will also work directly with the new institutions in carrying out the work of Phase 2, Objectives 2 and 3. A draft job description is available in the Appendix - Item 3.

Phase 2, Objective 1 *- Recruit a group of academic libraries from diverse institutions for participation*
> Expected output(s): The successful recruitment of five non-research 1 academic libraries, which represent a diverse student population, federal funding level, into this study and the DCN.

Recruitment will be prioritized for organizations that represent diversity in size, student demographic, and federal funding levels (including non-research 1 Carnegie Classification institutions) from the current membership of both the DCN and ARL. ARL has established, and currently operates, many programs and projects that extend beyond the ARL membership, including the Kaleidoscope program. We would like to recruit up to 5 additional institutions to participate in this research.

To ensure a diverse representation of institutions (including non R1 academic libraries) and individuals, the project team will leverage the wide communication networks of ARL and the DCN to promote participation. Interested library participants will be asked to complete a brief questionnaire as part of the recruitment process to ensure the program is recruiting institutions that need additional support in building and understanding their campus-based services and expenses, as well as ensuring we are creating a diverse and inclusive cohort. A draft version of this questionnaire is available in the Appendix - Item 1.

Phase 2, Objective 2 *- Draft Campus Infrastructure Maps with the New Library Cohort*
> Expected output(s): Campus infrastructure maps for each of the new five libraries participating in this study

The recently released NSF-funded, Association of American Universities and Association of Public and Land Grant Universities Guide for Accelerating Public Access to Research Data[10] recommends, as one of the initial steps for institutions, the formation of campus-wide coordinating teams to identify and catalog campus infrastructure for public

---

[10] Association of American Universities and Association of Public and Land-grant Universities (2021).
Guide to Accelerate Public Access to Research Data. Washington, DC. DOI: https://doi.org/10.31219/osf.io/tjybn

access to research data - or create a campus map of services (see Appendix for an example of a service map developed during the RADS research phase).

While seemingly straightforward, the process of conducting a campus mapping may be complicated by the size and complexity of the University, as well as by the research data needs of the research project. At Pennsylvania State University, for example, a researcher may have to interact with 22 different institutional service or infrastructure providers to address their research data needs[11] (see Item 6 in the Appendix).

Without a clear map or web portal of these services, inefficiencies, inequities, and gaps in services and infrastructure develop. Yet, few institutions have completed a rigorous and comprehensive review to move forward strategically with institutional research data services. To identify and understand the extent of services within institutions, we will collect data from research, IT, library and research institute administrators on expected and actual roles and responsibilities, services, and infrastructure support to enable federally funded research data sharing. Gathering data about existing institutional policies and/or strategies for public access to research data for all participating institutions will also inform these campus maps. This step is also critical for explaining and accounting for any differentiations that may be seen in the financial information gathering phase of the project.

Data from the RADS administrator surveys shows the distribution of data sharing activities across each institution (Figure 2). As mentioned above, campus offices and departments were categorized into one of four service areas: Institutes/Center, IT, Libraries, and Research Offices. Data from the survey results show that libraries support core data management and sharing activities at each institution, illustrating libraries' central role on campus to support federally funded research. The new library cohorts that participate in the campus mapping process, and whose research data management services and infrastructure will vary in maturity, will have the advantage of better situating their services and role within their institutions.
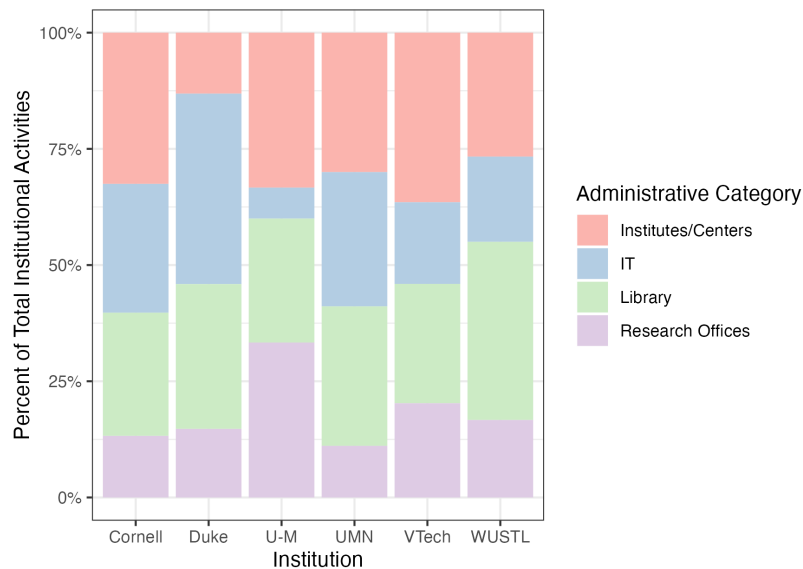
Figure 3: Distribution of overall data institutional sharing activities among four high level administrative categories. Bars represent the total sum of Data Sharing and Management (DMS) activities completed by units within each administrative category (out of 28 activities listed on the RADS survey).

This phase of the project will leverage the tools and approach developed in the original RADS research project and refinements made in Objective 1 to gather this information and determine the extent of local services and expenses. To enable macro-level understanding of how different institutions provide infrastructure and support (inclusive of research data policies and/or strategies) for data sharing, the campus maps will be generalized into a series of functional models. A functional model is a structured representation of the activities, actions, processes, operations within the modeled system or subject area.[12] These models will show the varying roles, responsibilities, and services of different types of academic campuses, as well as engagement with activities across the data management and sharing phases.

Phase 2, Objective 3 *- Develop Expense Models Related to Public Access to Research Data with Expansion Cohort*

> Expected output(s): Expense models of federally-funded research expenditures for both institutions and funded researchers, representing expenses at each of the five new library cohort institutions

During the RADS study the research team developed expense models across six academic institutions in five disciplinary areas. Through Phase 1 of this research we will refine and deepen the analysis of expenses across three of the original six institutions. The research team will leverage the refined methodologies and data collection instruments developed in Phase 1 of this research for application with the new institutions with both campus administrators/infrastructure providers and researchers.

Campus-Based Administrator Expenses Data Collection. Using the campus infrastructure models developed in Phase 2, Objective 2, administrators from each unit will be identified and interviewed or surveyed based on the refined methodologies developed in Phase 1. Expenses in terms of staffing and infrastructure for federally mandated data sharing will be assessed, as well as anticipated growth. To enhance comparability across units and institutions, these numbers will be gathered along with overall operating budgets and total number of staff to better understand the proportion of budget and labor needed to support data sharing. A full version of the campus-based questions used in the original RADS research are available on our website.

Researcher-Based Expenses Data Collection. Leveraging the IMLS, National Science Foundation (NSF), National Institute of Health (NIH), and Department of Energy (DOE) public grants databases and APIs, institutional affiliated and completed research projects will be identified for each new library involved in this research. Grants-based metadata will be downloaded for each found research project. Principal investigator names and emails will be verified against employee directories to confirm they are still institutionally affiliated with the University. Once identified, institutional teams will be invited to participate in this research and report retrospectively at the finances of completed federally funded research. This component of the research will leverage our piloted and refined methods for collecting this information through surveys and focus groups and then categorize expenses into the data management and sharing (DMS) activities referenced above and developed through the original RADS research. A sample of the researcher-based questions used in the original RADS research are presented below and a full version of the survey instrument is available on our website.

- Of the staff members hired or responsible for making research outputs broadly available, please consider the five positions most responsible for this work. If there were less than five positions responsible for this work, please fill out the information for the number of positions accordingly. What percentage of time did each of these five positions dedicate to data sharing and what was their approximate salary?
- For services, infrastructure, or staffing costs related to data sharing across the research lifecycle, what was your approximate budget (e.g., software, contracts, fees) for 2021-2022?
- What additional level of investment (staff, infrastructure, or services) does your unit anticipate making within the next 5 years to support data sharing?

---

[12] Defined according to the NIST Federal Information Processing Standards Publication 183: Integration Definition for Function Modeling (IDEF0). https://www.idef.com/idefo-function_modeling_method/
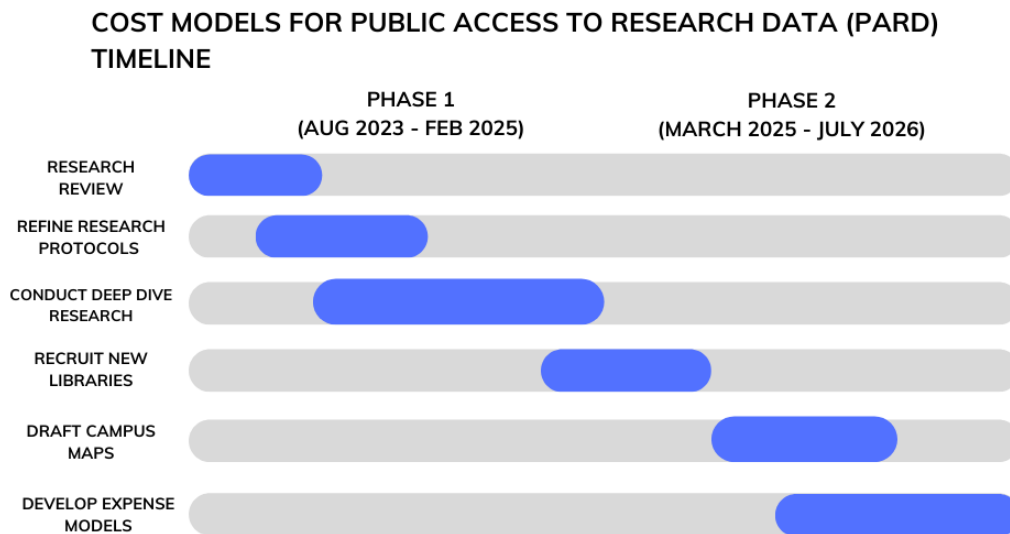
The original RADS research focused on five disciplinary areas to scope and pilot out the methods. Disciplinary scope for this research will be based on the frequency by which the related institutions received grants from IMLS, NSF, NIH, and DOE in the last ten years by subject/disciplinary area. This will be completed by utilizing the metadata gathered from the grants databases mentioned above.

Expense Data Analysis
Expense models across both the new and original RADS libraries will be compared to determine commonalities in the institutional and researcher expenses for federally funded data sharing, as well as assessing if financial burden falls disproportionately to libraries, certain units, or researchers across various institutional structures. This will be assessed with descriptive statistics as well as inferential proportional equality tests to assess independence of financial burden across institutional types and units within institutions. In addition to overall expense to institutions, proportions of budget required to support federally mandated data sharing will also be assessed. This is critical for understanding the extent to which smaller units and institutions are bearing an outsized burden to support federal sharing mandates.

Project Results
Library investments in infrastructure and services play a crucial role in enabling researchers to secure funding, publish research, and openly share data and publications related to their research. As current RADS results demonstrate (see Figure 3 above), university libraries support the majority of the activities required for federally mandated data sharing, and at three of our six institutions, are the largest university-wide contributor in these spaces. The RADS project empowers libraries to make meaningful investments in data management and sharing that reflect institutional partnerships and capacity while complying with federal policies. By providing data, models, and case studies that reflect both the novel and systemic ways that academic institutions organize to support researchers, we seek to empower libraries that wish to expand resources for research data management and sharing support while simultaneously informing US federal policy on data sharing. ARL's engagement with stakeholders in higher education (e.g., COGR, APLU, AAU) demonstrates to academic libraries that increased collaboration with other campus service providers is essential for optimal data management and sharing capacity building. This research will catalyze the body of information libraries have about these services and expenses, which will then allow for strategic choices to be made. The below timeline (Figure 4) indicates the phases and research objectives of the project.

Figure 4: A timeline of the proposed project, by phase and research objective.

Narrative Service and Cost Models for PARD - Schedule of Completion

Major Activities	Aug - Oct 2023	Nov '23 - Jan 2024	Feb - April 2024	May - July 2024	Aug - Oct 2024	Nov '24 - Jan 2025	Feb - April 2025	May - July 2025	Aug - Oct 2025	Nov '25 - Jan 2026	Feb - April 2026	May - July 2026
Year →	Year 1				Year 2				Year 3			
Phase →	Phase 1						Phase 2					
Phase 1												
Systematically review previous data collection instruments	■											
Review and refine funded researcher population		■										
Review and refine campus administrator population		■	■									
Review and refine data management & sharing activities			■									
Build upon and refine existing data collection protocols			■	■								
Develop a revised survey instrument				■								
Develop a revised interview methodology				■								
Collect data from funded-researchers retospectively				■	■							
Collect data from funded-researchers prospectively		■	■	■	■	■						
Analyze administrator data						■						
Analyze researcher data							■					
Phase 2												
Develop an outreach strategy to recruit non-ARL, non-DCN academic libraries						■						
Select and formally recruit five new libraries/institutions							■					
Collect data for campus mapping								■				
Draft campus infrastructure maps for the five new institutions									■			
Collect data for public access data sharing expense models - researchers										■		
Collect data for public access data sharing expense models - administrators										■		
Expense data analysis & modeling - researchers											■	■
Expense data analysis & modeling - administrators											■	■

Types

The primary digital products that will be generated during this research will be data about research data services and expenses at several institutions.

Additionally, we will have documentation, survey instruments and protocols for all of the data collection activities completed through this research.

Availability

All data collected through this research will be deidentified and made openly accessible online and disseminated alongside the other data collection documents and protocols. As appropriate, materials will also be deposited into the University of Minnesota's Digital Conservancy repository for long-term preservation and access including the assignment of Dublin Core metadata assigned for discovery.

Access

All materials and data will be published and made accessible under a standard Creative Commons license (CC-BY or CC-BY-NC) to allow for broad reuse by the community. A standard citation of the content producers will be provided to facilitate proper acknowledgement by secondary users.

Sustainability

Digital assets archiving and preservation is supported through an approach that starts with adequate documentation of data using metadata and other formats appropriate for long-term preservation. The digital assets from this project will be retained and curated for a minimum of 10 years. This archiving will protect against the fragile nature of commercial and other open-source projects, thus providing dedicated backup services for the project assets.

Data Management Plan

Types of Data:
This research will produce a number of text documents, models, use cases, and data about costing information. This includes functional models of how researchers leverage institutional services for meeting public access to research data requirements and use cases based on discipline and institution.

Data & Metadata Standards:
Whenever possible, open data formats or formats that do not use closed proprietary specifications will be adopted as asset accessibility and archiving standards for the project. For example, all text/data will be encoded using Unicode to prevent data loss. Uncompressed TIFF (or comparable) will be used for all images. Archival copies and originals of the data will be maintained according to archiving policies outlined below.

Metadata will be created and saved throughout the lifecycle of the research project and will be in line with the commonly accepted scholarly standards. All collections will have a well-structured metadata record created. This record includes elements such as author, license, abstract, publication date, funder, and others. For continued preservation of the data and materials, metadata elements consistent with the PREMIS data dictionary and data model will be implemented.

A "read-me" file will also be created that includes an asset inventory, general rights for reuse, contact information, a recommended citation, and a synopsis of the project.

Policies for Access and Sharing:
To improve dissemination of the research outputs a digital object identifier (DOI) will be assigned to the collection or at necessary granularity of the project. The dissemination information package (DIP) will include access copies of the workflows and models, the DataCite metadata record(s), and the "read-me" file.

Policies for re-use, re-distribution, derivatives:
The Principal Investigator will be responsible for ensuring the adherence of this data management plan within the project period. The project team will steward the data and digital assets. All assets will be made available for re-use through a research team member's institutional repository, Zenodo, and also made accessible through ARL's report and data platform. All assets will be licensed as Creative Commons-Attribution (CC-BY 4.0) to allow for the broad sharing and adaptation of the materials.

Plans for archiving and preservation:

Digital assets archiving and preservation is supported through an approach that starts with adequate documentation of data using metadata and other formats appropriate for long-term preservation. The digital assets from this project will be retained and curated for a minimum of 10 years. This archiving will protect against the fragile nature of commercial and other open-source projects, thus providing dedicated backup services for the project assets.

Organizational Profile: Association of Research Libraries (ARL)

The Association of Research Libraries (ARL) was established in 1932 in recognition of the need for coordinated action and a forum to address common problems among research libraries and archives. Since that time, the organization has grown from the founding 42 research libraries to 127 of the largest academic and public research libraries in the US and Canada. ARL's vision is that research libraries will be even more intimately engaged in supporting the full life cycle and activity range of knowledge discovery, use, and preservation, as well as the curation and sharing of knowledge in diverse contexts of the university's mission and of society more broadly. ARL achieves the vision by fostering the open exchange of ideas and expertise; advancing diversity, equity, and inclusion; and pursuing advocacy and public policy efforts that reflect the values of the library, scholarly, and higher education communities.

ARL seeks to advance an equitable and enduring research information environment that meets the needs of scholars (researchers, students, and independent scholars) now and in the future. With its members and in collaboration with partners around the world, ARL explores, pilots, enhances, and scales initiatives to ensure scholars can do their best work, by ensuring research libraries can do their best work.

In order to achieve our shared objective, the Association engages directly with and influences research libraries and their parent organizations, research and learning communities, and public policy makers. ARL does this by:
- Advocating for public policies in support of the mission and shared objective
- Catalyzing collective efforts to achieve enduring and barrier-free information
- Creating diverse, equitable, inclusive, and accessible services, collections, and work environments
- Providing data and analytics on research library practices, effectiveness, and impact

ARL focuses our efforts with leaders in our member institutions and other research library communities to:
- Shape and inform leadership practice throughout research libraries
- Facilitate a culture of innovation within research libraries

This is accomplished while engendering a member experience in which all members feel welcome and can participate in and benefit from the Association.

Given the volatility of the contemporary research and learning landscape, ARL continues to use its collective strength in convening, supporting, and endorsing others' work when it advances our shared objective, and, where appropriate, reprioritizes our resources as critical issues emerge.