

## Institutional Functional and Economic Models for Public Access to Research Data

The Association of Research Libraries (ARL) requests \$742,957 for a 3-year National Leadership Grant 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 institutional 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<sup>1</sup> and other related research outputs to accelerate the impacts of their investments. Many academic institutions have developed and launched infrastructure to support faculty in 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 public access to research data, these collaborations are often not based on evidence and are informal, at best. Given this distributed nature, the true cost of public access to research data is not well understood. In truth, there are many unknowns about the institutional landscape for public access to research data.

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 understand the local landscape and investments required of academic institutions. This research will catalyze this effort by building on previous funded research to understand the costs and expenses, across the institution, for public access to research data.

This research will answer the following questions:

1. What are the different models institutions put in place to support research data management and sharing policies?
2. What is the cost to the institution to implement public access to research data policies?
3. What is the cost to the researcher to comply with the public access to research data policies?

To answer these questions we will build upon the work plan and methods developed in the [Realities of Academic Data Sharing](#) (RADS) (NSF# #2135874) project that focused on understanding where researchers share their data, what is the completeness of that metadata, and how much were institutions spending at six institutions in five disciplinary areas. By its very nature the RADS project was meant to be exploratory and transformative. **Accelerating our knowledge and understanding of expenses for a wider range of disciplines and organizational models is the goal of this research.** Given the heterogeneity of support and infrastructure for research data services across institutions, a single financial and sustainability 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 look different within and across institutions. Yet there are commonalities we can draw across these institutions, and newer cost modeling frameworks we can put into practice and pilot. 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 memo, [Ensuring Free, Immediate, and Equitable Access to Federally Funded Research](#).

**Project Work Plan.** This research will take place over a three year period and is broken into two concurrent phases. Each year we will recruit 7-8 new institutions to join the research project - where they will take part in Phase 1 and Phase 2 of the research - for a total of 24 institutions. Recruitment will be prioritized for organizations that contribute to diversity in size, student demographic, and federal funding levels (including non-research 1 Carnegie Classification institutions). ARL has established, and currently operates, many programs and projects that extend beyond the ARL membership, including the [Kaleidoscope program](#).

**Phase 1: Conduct Campus Infrastructure Mapping.** The recently released NSF-funded, Association of American Universities and Association of Public and Land Grant Universities Guide for Accelerating Public Access to Research

---

<sup>1</sup> 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>

Data<sup>2</sup> recommends, as one of the initial steps for institutions, the formation of campus-wide coordinating teams to identify and catalog campus infrastructure for public access to research data. 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. This phase of the project will leverage the tools and approach developed in the exploratory RADS research project - a survey and focus groups to campus administrators at six institutions to determine the extent of their 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 public access to research data, the service mappings 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<sup>3</sup>. The project team will leverage the NIST Function Modeling Method<sup>4</sup> to design the institutional functional models for public access to research data. These models will show the varying roles, responsibilities, and services of different types of academic campuses, as well as the flow of activities.

**Phase 2: Gather Expense Data Related to Public Access to Research Data.** Leveraging funding agency API's, institutional affiliated and completed research projects will be identified for each institutional participant. Once identified, institutional teams will look retrospectively at the finances of completed federally funded research leveraging our piloted methods for collecting this information, through surveys and focus groups and then categorizing costs into the data management and sharing (DMS) activities the research team has developed with colleagues at the Council on Government Relations (COGR). 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](#).

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
Data Collection, Storage, and Management	Managing active data (e.g., storage, security, backup, lab notebooks)
	Creating quality control mechanisms or procedures
	Developing documentation of data (e.g., data dictionary, protocols)

Once collected, this data will be de-identified as much as possible, sensitive information removed, and the financial information shared as case studies mapped to the functional models.

### Project Results

- Data and information about expenses and costs for public access to research data
- Creation of functional models for institutional support for public access to research data
- Development of disciplinary and institutional case studies

**Budget.** The Association of Research Libraries is requesting \$742,957 for the three year project. Personal costs will total \$344,440 over three years and includes \$258,977 for salaries and \$85,462 for fringe benefits at a 33% rate. We are requesting \$37,500 for PI and project manager travel to share results of the research. \$3,600 is requested for survey software. \$45,000 is requested for statistical consultant support. To be accountable for institutional time (communications, administrative management, etc.) we are requesting \$60,000 in honoraria. Kickoff meetings will be held each year with the new institutional research teams, we estimate these costs to be \$84,000 over three years. We have \$574,540 in direct costs and \$168,418 in indirect costs based on the agreed upon 37.75% rate.

<sup>2</sup> Available at:

<https://www.aau.edu/sites/default/files/AAU-Files/Key-Issues/Public%20Access/AAU%20APLU%20Guide%20to%20Accelerate%20Public%20Access%20to%20Research%20Data.pdf>

<sup>3</sup> 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/](https://www.idef.com/idefo-function_modeling_method/)

<sup>4</sup> <http://www.idef.com/?download=250>