Realities of Academic Data Sharing (RADS) Initiative

Public Access Data Management and Sharing Activities for Academic Administration and Researchers

November 21, 2022

Shawna Taylor, MA, MLIS, Association of Research Libraries

with:

Jake Carlson, MA, MLIS, University of Michigan

Joel Herndon, PhD, Duke University

Alicia Hofelich Mohr, PhD, University of Minnesota

Wendy Kozlowski, MS, Cornell University Library

Jennifer Moore, MLIS, Washington University in St. Louis

Jonathan Petters, PhD, Virginia Tech

Cynthia Hudson Vitale, MA, MALIS, Association of Research Libraries

Note: All secondary authors have contributed equally; the secondary author list is alphabetical.

ASSOCIATION OF RESEARCH LIBRARIES

Table of Contents

Introduction

Federally mandated policies requiring public access to funded research data have impacted how funded researchers and institutions support data throughout the research life cycle. The 2013 White House Office of Science and Technology Policy (OSTP) memo, "Increasing Access to the Results of Federally Funded Scientific Research" (also referred to as the Holdren memo), formalized requirements for making funded research data publicly accessible. Many other federal agencies, as well as academic institutions internally, followed suit by implementing their own data management policies. This trend is continuing as the 2023 NIH Data Management and Sharing Policy and the revised federal agency policies that will result from the 2022 OSTP memo, "Ensuring Free, Immediate, and Equitable Access to Federally Funded Research" (also referred to as the Nelson memo), will transform how institutions and researchers manage their research data.

In response to the growing number of federal requirements to share publicly funded research data, many academic institutions have developed and launched a variety of support services to reduce faculty burden in meeting these requirements. These services are often spread across the institution, such as campus IT, the university libraries, and research offices, among others. A multitude of platforms exist for sharing research data and researchers store their data in institutional repositories, generalist repositories, discipline-specific repositories, or in other locations (such as personal websites) to comply with these mandates. The extent of where and how funded research data are shared, as well as the costs to support this sharing, are not fully understood. The Realities of Academic Data Sharing (RADS) Initiative, funded by the US National Science Foundation (NSF #2135874), considers these questions and examines the activities required to enable research data sharing and associated costs at six academic institutions: Cornell University, Duke University, University of Michigan, University of Minnesota, Virginia Tech, and Washington University in St. Louis.

To better understand which services, infrastructure, and staffing across these six campuses are needed to make research data publicly accessible, the activities required to enable these services, infrastructure, and staffing had to be unpacked first. By identifying specific public-access data management and sharing (DMS) activities (listed below), the RADS project team established a foundation for a common language and life cycle that could capture the actions of both funded researchers and campus administrators at these six universities regarding public access DMS activities.

Public Access to Research Data Life-Cycle Phases

Defining public access DMS activities required organizing them in a manner consistent with the research and grant life cycles. To do this, activities identified by RADS to support public access to research data were categorized into life-cycle phases. These phases are:

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

The phases and their associated activities were developed in collaboration with the Council on Governmental Relations (COGR), and focus on data management and sharing activities in a typical research life cycle. For the RADS project team, collaboration with COGR on developing shared definitions for the life-cycle phases of public access to research data helped forge a common understanding of what is needed to make wide-scale data management and sharing a reality. A shared language around public access phases and DMS activities, as well as costing, will engage funders, administrators, librarians, and researchers. While the final COGR and RADS phases and activities differ slightly to reflect different association member

audiences,[1] the core definitions reflect a shared language that will help institutions as they plan for new and revised federal policies.

RADS Data Management and Sharing Activities

The RADS team defined activities for data management and sharing, which are listed in the tables that follow. The activities are grouped in two categories—one for researchers and one for institutional service providers (such as libraries, IT, research office employees, etc.). While quite similar, institutional and researcher activities differ primarily in that institutional activities support the researcher in meeting award requirements. Due to these differences, the authors felt it necessary to modify the language to reflect the perspectives of the two groups within the institution.

Some activities, such as "Reviewing the data management plan (DMP)" or "Assessing data security needs," may occur during multiple phases in a project but are not necessarily listed in each phase. Project phases, and their associated DMS activities, are not considered linear and may be repeated at various stages during the research life cycle. Additionally, the authors acknowledge that not all activities pertinent to data management and sharing are listed below, as these activities can vary according to discipline and project need. Similarly, not all activities below may be applicable for every project or researcher making funded research data publicly available, and some activities are good data management practices, even for data that may not immediately be made publicly available. Therefore, while these activities are necessary for data sharing, they are certainly not exclusive to research life cycles that end with public data access. Several activities include tasks the authors recognize as important in the research and grant life cycles; however, tasks have mostly been omitted in order to focus primarily on the activities required or necessary in making public access to research data a reality.

Endnotes

1 COGR's interest in articulating their own set of DMS activities was motivated by providing guidance for their membership around the 2023 NIH Data Management and Sharing Policy. Chapter 3 Part II: Implementation Roles & Responsibilities—Roles & Responsibilities Matrix in *COGR's NIH Data Management and Sharing Policy Readiness Guide* includes many of the RADS DMS activities.

Realities of Academic Data Sharing (RADS)
Public Access Data Management and Sharing (DMS) Activities v1

Researcher Activities

DMS Phase	DMS Activity	Task
Planning, Design, and Start Up of Projects	Preparing data management plans (DMPs) or data management and sharing (DMS) plans	• Reviewing data sharing expectations (for example, scientific data to share, relevant standards, repository selection, timelines) that apply and should be reflected in a plan • Ensuring all policy requirements are met
	Identifying data management and sharing costs to be included in grant budgets	
	Preparing institutional review board (IRB) protocols and informed consent for data sharing	
	Determining storage solutions for active research data	
	Selecting an appropriate repository (or repositories) for making research data broadly available	
	Evaluating data security needs	
	Determining intellectual property and copyright considerations	• Reviewing data use agreement (DUA) for any use of secondary data
	Developing materials transfer agreements and/or data use agreements (DUAs)	

DMS Phase	DMS Activity	Task
	Reviewing disciplinary or institutional standards and/or best practices for handling, collecting, and documenting data	
Data Collection, Storage, and Management	Developing documentation of data (for example, data dictionary, protocols)	
	Creating quality-control mechanisms or procedures	
	Evaluating data-analysis tools and processes to support sharing and reproducibility	
	Managing active data (for example, storage, security, backup, lab notebooks)	• Monitoring data management plans; reviewing and updating during the course of the award/support period to reflect any changes in the management and sharing of scientific data (for example, new scientific direction, new repository option, timeline revision) • Storing data across the project life cycle
Making Data Broadly Available	Making decisions about what data to share or host	
	Preparing data for sharing (for example, de-identification, check privacy/personally identifiable information (PII)/protected health information (PHI), selection, curation, data cleaning, validation, and quality control)	

DMS Phase	DMS Activity	Task
	Submitting data into a data sharing platform (for example, institutional repository, generalist repository, disciplinary repository)	
	Creating documentation for research data (for example, structured metadata, README files)	
	Selecting or applying licenses to data	
	Migrating data file formats to be more open or accessible	
	Creating persistent identifiers (PIDs; for example, digital object identifiers (DOIs))	
	Checking for compliance with any existing data use agreement (DUA)	
Data Retention, Including Preservation, Archive, and Long-Term Access	Migrating files to new formats or systems as needed	
	Monitoring integrity of preserved data	
	Making decisions about de-accessioning and removal of research data	
	Ensuring data security when appropriate (for example, PHI/Health Insurance Portability and Accountability Act (HIPAA), export controls, Federal Information Security Management Act (FISMA), student data, and intellectual property)	
Project Closeout and Compliance	Ensuring funding agency requirements for data sharing have been met	
	Providing compliance support around research project reports	

Realities of Academic Data Sharing (RADS)
Public Access Data Management and Sharing (DMS) Activities v1

Institutional Activities

DMS Phase	DMS Activity	Task
Planning, Design, and Start Up of Projects	Reviewing or preparing data management plans (DMPs) or data management and sharing (DMS) plans	• Reviewing data sharing expectations (for example, scientific data to share, relevant standards, repository selection, timelines) that apply and should be reflected in a plan • Ensuring all policy requirements are met
	Reviewing data management and sharing costs and expenses to be included in grant budgets	
	Reviewing of IRB protocols and informed consent for data sharing	
	Developing, building, or recommending storage solutions for active research data	
	Supporting an appropriate repository (or repositories) for making research data broadly available	
	Assessing data security needs and recommending solutions	

DMS Phase	DMS Activity	Task
	Supporting intellectual property and copyright considerations	• Reviewing data use agreement (DUA) for any use of secondary data
	Developing or reviewing materials transfer agreements and/or data use agreements (DUAs)	
	Referring to disciplinary or institutional standards and/or best practices for handling, collecting, and documenting data	
Data Collection, Storage, and Management	Developing or reviewing documentation of data (for example, data dictionary, protocols)	
	Creating quality-control mechanisms or procedures	
	Evaluating or recommending data-analysis tools and processes to support sharing and reproducibility	
	Managing active data (for example, storage, security, backup, lab notebooks)	• Monitoring data management plans; reviewing and updating during the course of the award/support period to reflect any changes in the management and sharing of scientific data (for example, new scientific direction, new repository option, timeline revision) • Storing data across the project life cycle

DMS Phase	DMS Activity	Task
Making Data Broadly Available	Consulting on decisions about what data to share or host	
	Providing or hosting repositories for making data available	
	Preparing or consulting on preparing data for sharing (for example, de-identification, check privacy/personally identifiable information (PII)/ protected health information (PHI), selection, curation, data cleaning, validation, and quality control)	
	Submitting data into a data sharing platform (for example, institutional repository, generalist repository, disciplinary repository)	
	Creating or reviewing documentation for research data (for example, structured metadata, README files)	
	Consulting, selecting, or applying licenses to data	
	Recommending or migrating data file formats to be open or more accessible	
	Creating or recommending persistent identifiers (PIDs; for example, DOIs)	
	Checking for compliance with existing data use agreements (DUAs)	

DMS Phase	DMS Activity	Task
Data Retention, Including Preservation, Archive, and Long-Term Access	Consulting on or migrating files to new formats or systems as needed	
	Monitoring integrity of preserved data	
	Making decisions about de-accessioning and removal of research data	
	Ensuring data security when appropriate (for example, PHI/Health Insurance Portability and Accountability Act (HIPAA), export controls, Federal Information Security Management Act (FISMA), student data, and intellectual property)	
Project Closeout and Compliance	Ensuring funding agency requirements for data sharing have been met	
	Providing compliance support around research project reports	

Association of Research Libraries
21 Dupont Circle, NW
Suite 800
Washington, DC 20036
T 202.296.2296
F 202.872.0884

ARL.org
pubs@arl.org