ARL/CARL Joint Task Force on Research Data Services: Final Report

July 16, 2021
Association of Research Libraries / Canadian Association of Research Libraries
Joint Task Force on Research Data Services

Task Force Members

Martha Whitehead, Chair, Harvard University
Dale Askey, University of Alberta
Donna Bourne-Tyson, Dalhousie University
Karen Estlund, Colorado State University
Susan Haigh, Canadian Association of Research Libraries
Claire Stewart, University of Nebraska–Lincoln
Kornelia Tancheva, University of Pittsburgh
Tyler Walters, Virginia Tech

Working Group Members

Ibraheem Ali, UCLA
Thea Atwood, University of Massachusetts Amherst
Jonathan Cain, University of Oregon
Jake Carlson, University of Michigan
Wind Cowles, Princeton University
Renata Curty, UC Santa Barbara
Marcel Fortin, University of Toronto
Jimmy Gaphery, Virginia Commonwealth University
Lisa Johnston, University of Minnesota
Amy Koshoffer, University of Cincinnati
Wendy Kozlowski, Cornell University
Sherry Lake, University of Virginia
Tim McGeary, Duke University
Andi Ogier, Virginia Tech
Plato Smith, University of Florida
John Watts, Texas A&M University

ARL Staff Leads

Jennifer Muilenburg, University of Washington, ARL visiting program officer
Judy Ruttenberg, ARL senior director of Scholarship and Policy
Table of Contents

About the Task Force 4

Objective 1: Develop a shared understanding among ARL and CARL members of the roles of research libraries in the research data ecosystem 5

Objective 2: Develop a roadmap with recommendations for the roles of research libraries with regard to research data principles, policies, and approaches to managing research data in the context of the Open Science by Design framework and recommendations 7

Recommendation 1: Conduct a cross-campus mapping of existing campus resources and researcher needs for RDS 8

Recommendation 2: Define a library portfolio and strategy for RDS 8

Recommendation 3: Articulate library and institutional research data services and partnerships 8

Recommendation 4: Formalize partnerships through development of a service catalog 9

Recommendation 5: Document services by elements of data management requirements 10

Recommendation 6: Evaluate the program on a spectrum of maturity 12

Recommendation 7: Define an institutional strategy for RDS 12

Objective 3: Develop strategies for discipline-specific research data approaches, priorities for automation of processes, economic models to scale and sustain shared resources, prioritization of research data to steward, and decision-making rubrics 13

Discipline-specific research data 13

Priorities for automation of processes 14

Economic models for shared resources 14

Prioritization of data to steward 15

Next Steps 15

Endnotes 16
About the Task Force

The Association of Research Libraries (ARL) Action Plan 2019–2021 advances an objective under the Scholars and Scholarship priority to position ARL members to lead within their institutions on “open science by design”—a reference to a 2018 consensus report by that title published by the US National Academies of Sciences, Engineering, and Medicine. In 2019, the ARL Scholars and Scholarship Committee charged a task force composed of both ARL member directors and data librarians to work with ARL staff (including visiting program officers) to develop resources members could use to advance this objective with respect to research data services (RDS). The committee recommended partnering with the Canadian Association of Research Libraries (CARL), based on CARL’s leadership in developing Portage, an initiative that has built a national community of practice supporting research data management in Canadian research institutions, and has worked collaboratively to develop tools, services, and best practices for research data stewardship in Canada.

In charging the task force, the Scholars and Scholarship Committee wanted to ensure it would build on prior work (citing in particular the OCLC Research Realities of Research Data Management series and the National Academies’ Open Science by Design report) and connect to emerging initiatives internally and among partners.

The purpose of the task force was twofold: (1) to demonstrate and commit to the roles research libraries have in stewarding research data and as part of institution-wide research support services and (2) to guide the development of resources for the ARL and CARL memberships in advancing their organizations as collaborative partners with respect to research data services in the context of FAIR principles and the Open Science by Design framework. In keeping with the ARL Action Plan, research libraries will be successful in meeting these objectives if they act collectively and are deeply engaged with disciplinary communities.

The task force formed three working groups of data practitioners,
representing a wealth of expertise, to research the institutional landscape and policy environment in both the US and Canada, setting three core objectives for the work:

1. Develop a shared understanding among ARL and CARL members of the roles of research libraries in the research data ecosystem

2. Develop a roadmap with recommendations for the roles of research libraries with regard to research data principles, policies, and approaches to managing research data in the context of the *Open Science by Design* framework and recommendations

3. Develop guidance for research libraries and for representing research libraries’ work with policy makers, including strategies for discipline-specific research data approaches, priorities for automation of processes, economic models to scale and sustain shared resources, prioritization of research data to steward, and decision-making rubrics

**Objective 1: Develop a shared understanding among ARL and CARL members of the roles of research libraries in the research data ecosystem**

ARL and CARL are engaged in their respective national and international policy discussions around research data—through, for example, the Board on Research Data and Information (BRDI) of the US National Academies of Sciences, Engineering, and Medicine; Canada’s New Digital Research Infrastructure Organization (NDRIO); and the International Science Council’s Committee on Data (CODATA). While broadly informed by recent national and international developments in research data management, the ARL/CARL joint task force working groups concentrated on the role research libraries play within their institutions, in collaboration with campus partners, researchers, and each other.
As educators and stewards of the scholarly and scientific record, research libraries have a significant interest in accelerating open research and scholarship on their campuses. The broad adoption of open research principles and strategies benefits the individual researcher through increased citations and scholarly impact, spurs scientific advancements through the rapid sharing of data, and provides more equitable access to research. Research universities are promoting open science practices and principles as they relate to funder and publishing requirements, reflecting a growing impatience with a system of incentives and rewards that many perceive to be out of alignment with scientific values. Academic research library leaders have a unique position on campus, supporting every discipline with services, expertise, collections, and infrastructure.

For more than a decade, as key research funding and policy making agencies have steadily increased their requirements of institutions and investigators to manage, preserve, share, and describe research data, libraries have been in the forefront of institutional efforts to meet those mandates. Data librarians have worked alongside researchers and tool builders to create and commit to FAIR—findable, accessible, interoperable, and reusable—data principles. And libraries have launched collaborative, multi-institution networks of expertise and/or infrastructure, such as the Data Curation Network in the US and the Portage Network in Canada.

The specific policy environment and the extent of coordination of national infrastructure differs between Canada and the United States, but core elements of research data management as required by major funding agencies, and instantiated in institutional services, are similar enough to collaborate on a shared understanding of library roles. These roles include:

- Providing services for faculty on the most commonly required elements for data management by funding agencies in Canada and the United States: assisting with data management planning, assisting with data description (including metadata), consulting
on data ethics and privacy, data sharing through deposit or consultation, and retention and preservation

- Partnering on grants to ensure these practices are embedded into projects from the start
- Providing education and training that has driven researcher interest and influenced the growth of research data services within the institution
- Leading the development, advocacy, and adoption of persistent identifiers (PIDs)
- Influencing and consulting on copyright, licensing, and disciplinary expertise
- Shaping and socializing open science norms and standards, including FAIR data principles

Objective 2: Develop a roadmap with recommendations for the roles of research libraries with regard to research data principles, policies, and approaches to managing research data in the context of the Open Science by Design framework and recommendations

What follows is a set of recommendations based on proven practices among ARL and CARL libraries. While most ARL and CARL libraries provide research data services, the extent of their service offerings, level of staff, and integration with related services within their institutions vary. These recommendations may be best used as a checklist or pathway for developing and maturing research data services. A library that is still developing an RDS program might want to begin by conducting a campus-mapping of existing research data service points across the institution. Another library may have an existing RDS program but lack formal partnerships and defined roles and responsibilities with other infrastructures and services across the institution. A next step in this case may be the creation of a formal service catalog.
In Canada, the Tri-Agency Research Data Management Policy requires institutional grantees to develop and publish a research data management strategy. In the United States, there is no such requirement, but recommendations from the Association of American Universities/Association of Public and Land-grant Universities Accelerating Public Access to Research Data (APARD) initiative include creating or updating institutional data policies. Successful institutional strategies and policies will both address the elements required by key funding agencies for sharing and managing data, and include provisions for both sensitive and open data.

**Recommendation 1: Conduct a cross-campus mapping of existing campus resources and researcher needs for RDS**

- Example: [University of Michigan Data Services—Mapping Campus Landscape](#)

**Recommendation 2: Define a library portfolio and strategy for RDS**

- Leverage the campus-mapping conducted in step one; and complete a strengths, weaknesses, opportunities, and threats (SWOT) analysis for potential library RDS services. (See, for example, the [UC Merced RDS SWOT analysis](#))
- Create a library RDS strategic plan. (See, for example, “[Strategic Planning for Research Data Services](#)”)

**Recommendation 3: Articulate library and institutional research data services and partnerships**

Compile an institution-wide list of research data service points.

*Resources and examples*

- [Research Data Services Checklist](#)
- [Taxonomy of research data services](#)
Recommendation 4: Formalize partnerships through development of a service catalog

For the past decade or more, ARL and CARL members have cultivated key partnerships with senior research officers, chief information officers, high-performance computing units, and other faculty-facing units. These partnerships can be vulnerable in their dependencies on personal relationships, rather than codified into official relationships between campus units. Service catalogs are a common practice in information technology management for managing collaborations. A service catalog establishes a compact between users and service providers, and encourages a continual assessment of current areas of emphasis and potential avenues for investment in the future.

The following framework is a tool for assessing RDS partnerships through six facets:

1. **Research Data Service**: Does the partnership have a focus on a specific service area (for example, education, consultation, technology, publishing, stewardship)?

2. **Research Data Life Cycle**: What stages of the research data life cycle does the partnership advance?

3. **Best Practices**: What RDS best practices are represented in the partnership? (FAIR; CARE; ethics; diversity, equity, and inclusion; reproducibility and replicability; compliance; institutional mission; open science/research)

4. **Affiliation of Partner**: Who is the partner?

5. **Audiences**: Who are the intended audiences of the partnership?

6. **Partnership Life Cycle**: What is the current maturity of the partnership?
Tools for creating a service catalog

- Research Data Curation: A Framework for an Institution-Wide Services Approach
  - EDUCAUSE Data Curation Roles Planning Matrix
- RDS Organizational Service Layers and Infrastructure checklist
- RDS partnership framework for a catalog
  - Example: RDS Partnership Catalog

Recommendation 5: Document services by elements of data management requirements

Government funding requirements in Canada and the US share basic elements of data management. These elements map to functional service areas of data description, ethics and privacy, intellectual property rights, storage and security, data sharing, deposit, and preservation.

Table of RDS Funder Requirements and Associated Tools and Checklists

<table>
<thead>
<tr>
<th>Data Description</th>
<th>Data Curation Network CURATED checklists</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access and Sharing</td>
<td>Data Repository Feature and Function Evaluation Checklist</td>
</tr>
<tr>
<td><strong>Institutional examples:</strong></td>
<td></td>
</tr>
<tr>
<td>• Virginia Tech Repository Evaluation Matrix</td>
<td></td>
</tr>
<tr>
<td>• Penn State University ScholarSphere policies on content &amp; deposit, access, preservation, and curation</td>
<td></td>
</tr>
<tr>
<td>Metadata</td>
<td>Research Data Alliance (RDA) Metadata Directory</td>
</tr>
<tr>
<td></td>
<td>Implementing Effective Data Practices: Stakeholder Recommendations for Collaborative Research Support</td>
</tr>
<tr>
<td><strong>Intellectual Property Rights</strong></td>
<td>Cornell University Research Data Management Service Group, <em>Introduction to Intellectual Property Rights in Data Management</em>&lt;sup&gt;27&lt;/sup&gt;</td>
</tr>
<tr>
<td>---------------------------------</td>
<td>--------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Ethics and Privacy</strong></td>
<td>CARE Principles for Indigenous Data Governance&lt;sup&gt;28&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>Format</strong></td>
<td>University of Washington Libraries data format best practices&lt;sup&gt;29&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>Archiving and Preservation</strong></td>
<td>Canadian Federated Research Data Repository <em>(FRDR)</em>&lt;sup&gt;30&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>NIH Workshop on the Role of Generalist Repositories to Enhance Data Discoverability and Reuse&lt;sup&gt;31&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Scholars Portal Dataverse&lt;sup&gt;32&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>Storage and Backup</strong></td>
<td>University of Toronto Libraries data storage and backup best practices&lt;sup&gt;33&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>Data Management Planning</strong></td>
<td>DMPTool&lt;sup&gt;34&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>ezDMP&lt;sup&gt;35&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>DMP Assistant&lt;sup&gt;36&lt;/sup&gt;</td>
</tr>
</tbody>
</table>
Recommendation 6: Evaluate the program on a spectrum of maturity

Assess institutional services according to maturity and capabilities models.

Examples

- Research data services maturity model
- Capabilities Model

Recommendation 7: Define an institutional strategy for RDS

Absent the creation of an institutional policy or strategy, external mandates can elicit a diffuse response across campus, whereby disparate units create redundant and siloed services. Lack of coordination also poses a risk to the institution that key needs will go unmet. Like data management planning itself, policies protect institutions against risk related to anything from breaches of sensitive data to being out of compliance. A well-articulated policy can be part of supporting responsible conduct of research. Since the AAU/APLU APARD work began in 2017, AAU, APLU, and ARL have pushed to make data sharing part of institutional policies, mirroring the Tri-Agency policy evolution.

Examples

- Institutional Research Data Management Strategy Template
- Dalhousie University Institutional Research Data Management Strategy

Institutional data policies

In US institutions, institutional data policies are more common. Key parts of an institutional policy include: ownership, security, storage, retention, transfer, access/sharing, unit responsibilities, PI responsibilities, policy webpage.
Data policy examples

- Utah State University research data policy\textsuperscript{41}
- Iowa State University research data policy\textsuperscript{42}

Data policy resources

- Starting the Conversation: University-Wide Research Data Management Policy\textsuperscript{43}
- Guidance for Developing a Research Data Management (RDM) Policy\textsuperscript{44}

**Objective 3: Develop strategies for discipline-specific research data approaches, priorities for automation of processes, economic models to scale and sustain shared resources, prioritization of research data to steward, and decision-making rubrics**

Much of the work associated with this objective is outside of individual institutions and involves professional societies, national funding agencies, and interdisciplinary research communities. What follows are pathways for ARL and CARL members to engage in collaborative work that helps position research libraries in this broader context.

**Discipline-specific research data**

Key strategies for libraries to develop discipline-specific RDS capacity include: participating in inter-institutional collaborations (such as the DCN and Portage), partnering with scholarly society or disciplinary repositories (such as FRDR), establishing a faculty champions program, and facilitating the creation and development of data communities.

**Examples**

- Portage Curation Expert Group\textsuperscript{45}
- Data Curation Network\textsuperscript{46}
- Federated Research Data Repository\textsuperscript{47}
• Re3data.org Registry of Research Data Repositories
• Contribute library expertise to emerging data communities
• Data Curation Network and Ithaka S+R collaboration on data communities
• A Tool for Assessing Alignment of Biomedical Data Repositories with Open, FAIR, Citation and Trustworthy Principles

Priorities for automation of processes

ARL staff held six focus group sessions with the sixteen working group members and additional data practitioners from ARL and CARL institutions. While there are developments in large, well-funded data science environments, participants described process automation as largely aspirational for libraries and identified the following priorities:

• Metadata creation, including assigning PIDs
• Preservation systems integration (e.g., Archivematica)
• Supporting research graph initiatives through OpenAire and DataCite

Resources

• BRDI committee on automating scientific research workflows
• Implementing FAIR Data for People and Machines: Impacts and Implications—Results of a Research Data Community Workshop

Economic models for shared resources

In 2020, the US National Academies published a roundtable report on life-cycle decisions for biomedical data and cost forecasting. The framework can be applied to non-biomedical data and there is great interest in its application.

• Life-Cycle Decisions for Biomedical Data: The Challenge of Forecasting Costs
• Research Data Preservation in Canada
• Recommendations for a National Canadian Dataverse

Members of the ARL/CARL RDS Task Force are participating in an
OCLC Research project on making strategic choices about library collaboration in this area.\(^{59}\)

**Prioritization of data to steward**

Strategies and priorities for data selected for long-term stewardship are still developing within institutions. ARL and CARL can influence these conversations through partnerships with disciplinary societies and repositories.

**Next Steps**

1. Continued engagement between ARL and CARL on the role of research libraries in RDS

2. In partnership with AAU and APLU, convene ARL members to gather feedback on the [National Institute of Standards and Technology Research Data Framework]^{60}

3. Build upon the AAU/APLU APARD work to develop institutional functional models for public access to research data

4. Investigate the open by design approach with regard to Indigenous data sovereignty, community expectations, and ethical, legal, and commercial obligations of researchers

5. Examine costs related to public deposit of NSF-funded research

6. Work with disciplinary societies and repositories on coordinating resources and services

7. Hold a series of CARL/ARL member discussions on emerging areas of interest for research libraries, including big data, sensitive data, AI, data repository certification, and security
Endnotes


9. Cynthia Hudson-Vitale and Judy Ruttenberg, “Persistent Identifiers Connect a Scholarly Record with Many Versions,” ARL


22. ARL Research Data Services Stakeholders and Partners Working Group, “RDS Partnership Catalog Example” (unpublished document, last modified April 24, 2021), https://docs.google.com/document/d/1x6_2BTJOb4mHjZN8Rrs6nqZXsAaN2SipP4CtDb4I/edit?usp=sharing.


42. “Research Data,” Iowa State University Policy Library, accessed


