September 2022

Alliance RDM Strategy

The RDM Strategy, published in 2022, represents the first phase of the Alliance's Data Management Strategy, and an updated version is being prepared as part of our development of a 2025-30 mandate.

Digital Research Alliance of Canada | **Alliance de recherche numérique** du Canada

Table of Contents

Summary

Alliance RDM's Strategic Plan builds on a long history of RDM activities in Canada. It provides a two-year plan that continues on its strategic course, while scaling up to meet growing need and demand. As part of this strategy planning and development, this process included the adoption and application of the Research Data Framework (RDaF), scoring self-assessment, and evidence gathering process. The results of the RDaF self-assessment provides a clear sense of the current state of RDM in Canada and helps frame the Alliance's focus and desired future state. This Strategy documents the community context and governance that frames this ecosystem, provides a sense of the current state of RDM activities and initiatives, and looks to the future and envisions a path forward.

Purpose of this Document

This document provides a high-level, preliminary view of a revised Strategy for Research Data Management (RDM) under the Digital Research Alliance of Canada (the 'Alliance').
This Strategy includes the application of the Research Data Framework (RDaF), a publication of the US National Institute of Standards and Technology (NIST)[1], to develop an evidence-based review of the national state of RDM currently, and will allow us to envision and leverage future opportunities for action.

The aim of this Strategy is to maintain continuity and momentum for core, ongoing, national RDM Services and Programs, while anticipating and being responsive to DRI ecosystem needs identified through the Alliance's Needs Assessment and other strategic activities. The initiatives described in this Strategy will also seek to foster and forge researcher-centric, pan-DRI synergies with Advanced Research Computing (ARC) and Research Software (RS), and align with the Alliance's Cybersecurity mandate.

This Strategy aligns with the Alliance's Vision and Mission statements:

> *Vision*: To catalyze world-class Canadian research for the benefit of all.

> *Mission*: As a trusted and inclusive partner, the Digital Research Alliance of Canada fosters national and global collaboration to provide researcher-centric, sustainable and integrated digital research infrastructure.[2]

In the next section, we provide an overview of this Strategy, its history, and the RDaF framework it is based on.

Overview

History of RDM Strategy Development

This RDM Strategy reflects a long history of RDM strategic planning and has roots in the LCDRI's Data Management Position Paper (2017), the Alliance's Current State of Research

---

[1] Hanisch. J. R., Kaiser. L. D and Carroll. C. B. NIST Special Publication 1500-18. RdaF. Retrieved from: https://doi.org/10.6028/NIST.SP.1500-18

[2] Digital Research Alliance of Canada. Strategic Plan. https://alliancecan.ca/en/initiatives/strategic-plan. Retrieved August 2022.

Data Management Report (2020),[3] and the Statements of Work submitted to ISED by the Canadian Association of Research Libraries' Portage Network to fund RDM activities nationally for Fiscal Years 2018-19 and 2019-20, prior to the transition of RDM into the Alliance.

Building upon this solid foundation and moving forward, we have also adopted the Research Data Framework (RDaF)[4] into our strategy planning and development. In the first step of applying this industry recognized framework, an RDM Working Group (RDM WG) composed of RDM Subject Matter Experts (SMEs) from the Alliance and the wider, national RDM Community conducted a self-assessment analysis of the Alliance's current maturity related to RDM. The Working Group was community-focused and collaborative, including representation from the Alliance across each of the DRI pillars, the RDM Network of Experts, and community partners (Compute Ontario and the Canadian Association of Research Libraries).

The Research Data Framework (RDaF)

The RDM WG selected a US National Institute of Standards and Technology (NIST)[5] evaluation framework to help select and prioritize Alliance RDM initiatives. NIST is currently leading the development of the Research Data Framework (RDaF)[6] with involvement and input from national and international leaders in the research data stakeholder community. One of the purposes of this framework is for it to act as a tool that can inform and help develop the research data culture in an organization.[7]

The RDaF comprises six Functions, which correspond to stages in the research data lifecycle, with Categories and Subcategories for each Function. These Functions include: Envision, Plan, Generate/Acquire, Process/Analyze, Share/Use/Reuse, and Preserve/Discard. These map directly onto the research data lifecycle. While Figure 1 presents these components as a cyclical continuum, they should be performed concurrently and continuously to address research data management needs.[8] These Functions are further explained in Table 1.

---

[3]Khair, S. et al. (2020). The Current State of Research Data Management in Canada: An Update to the LCDRI Data Management Position Paper. https://zenodo.org/record/6564659

[4] Research Data Framework (RDaF): https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1500-18.pdf

[5] US National Institute of Standards and Technology: https://www.nist.gov/

[6] Hanisch. J. R., Kaiser. L. D and Carroll. C. B. NIST Special Publication 1500-18. RdaF. Retrieved from: Https://doi.org/10.6028/NIST.SP.1500-18

[7] Hanisch. J. R., Kaiser. L. D and Carroll. C. B. NIST Special Publication 1500-18. RdaF. Retrieved from: https://doi.org/10.6028/NIST.SP.1500-18

[8] Hanisch. J. R., Kaiser. L. D and Carroll. C. B. NIST Special Publication 1500-18. RdaF. Retrieved from: https://doi.org/10.6028/NIST.SP.1500-18
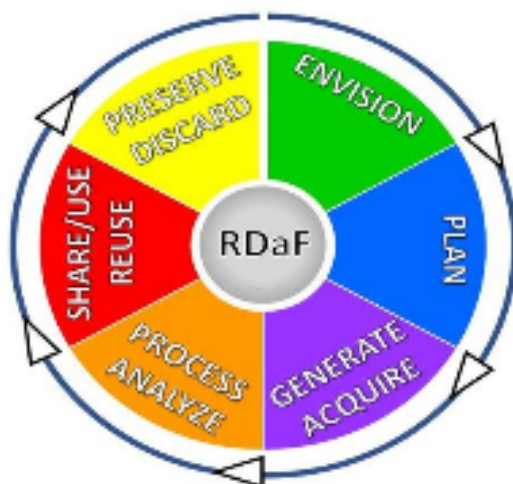
Figure 1: RDaF Functions based on the Research Data Lifecycle

It is important to understand that the RDaF at this initial stage is a reflection of the Current and Target maturity levels for the existing Alliance RDM team and services, as well as the Business Cases that have been prepared for the Multi Year Funding Proposal 2023-2025 (MYFP). In some cases, it is best to consider these Categories and Sub-categories in the broader context of the Alliance/Federation and external partners, but we limited ourselves in most areas to just the Alliance's RDM context, or we considered the element to be either Out of Scope or Future Scope (e.g., part of the next stage of planning or a future iteration of the RDaF exercise).

The Alliance will continue to use and expand application of the evolving RDaF framework, and other frameworks as appropriate, to other areas of Alliance activity. The Alliance is directly engaged with NIST in providing feedback and participating in further development of the RDaF framework.

Table 1: RDaF Functions, Descriptions, and Categories

NIST RDaF Function	Description	Categories Included
Envision	This Function encompasses the review of the overall strategies and drivers of an organization's research data program. The Envision Function is where choices and decisions are made that together chart a high-level	Data Governance Structure, Community Engagement, Data Culture, Reward Structure, Workforce/Career Paths, Data Safety and Security, Strategy, and Data Risk Management.

	course of action to achieve desired organizational goals.	
Plan	This Function encompasses the tactical management positioning in an organization for effective research data management throughout the research data lifecycle.	Chain of Control, Economics and Costs of Planning, Funding Planning, Data Objects, Hardware/Software Infrastructure, Data Management Planning, Scientific Data Standards, and Assessment and Controls.
Generate/Acquire	This Function covers the generation of raw research data, both experimentally and computationally, within an organization, and the collection or acquisition of research data produced outside of an organization.	Sources of Raw Data, Experimental Data Generation, Computational Data Generation, FAIR Principles for Data Generated In-House, External Sources of Data, and Community-Based Standards for Formats.
Process/Analyze	This Function concerns the actions performed on generated or acquired research data to yield processed research data, typically using software, from which observations and conclusions can be made. This Function also concerns the research data stewardship functions performed by an organization	Data Provenance, Data Architecture, Software Tools, Scientific Workflow Processes and Systems, Data Inventory, Data Modeling and Analytics, Data Representation/Models/Structures, Data Curation, and Metadata.
Share/Use/Reuse	This Function outlines how raw and processed research data are disseminated, used, and reused within an organization and any constraints or encouragements to use/reuse. It also includes the dissemination, use, and reuse of raw and processed	Legal and Licenses, Data Publishing, Data Citation, Internal and External Data Access, Levels of Protection, Applications and Analysis, and Data Architecture for Application and Use.

	research data outside of an organization.	
Preserve/Discard	This Function delineates the end-of-use and end-of-life provisions for research data in an organization and includes records management, archiving, and safe disposal.	Criteria, Data Sustainability, Storage and Preservation of Data, Moving Data from One Service to Another across Organizations, and Retention and Disposition Schedules.

The RDaF was selected as an effective tool to provide insight on the RDM-ecosystem, including identifying gaps, challenges, and opportunities for the future. Further, both NIST as an organization and individuals specifically involved in the development of RDaF are well-respected and provide broad and internationally-focused perspectives. Use of this framework involved a scoping exercise to determine what Categories and Subcategories from the RDaF Functions were in scope, followed by a scored self-assessment allowing for focused evaluation of the state of RDM within the Alliance and in the broader Canadian context.

The results of the RDaF self-assessment provide a clear sense of the current state of RDM in Canada and help frame the Alliance's desired future state and focus.

RDaF Process Scoring

While the RDaF provides an effective self-assessment tool to gauge current and future states, in its present form it lacks an objective framework for scoring maturity levels – instead relying upon and recommending use of external, well-accepted frameworks such as the Capability Maturity Model Integration (CMMI).[9] This model was chosen by the RDM WG, and used to score self-assessed maturity levels for each of the RDaF Functions.

Using both of these tools (RDaF and CMMI) members of the RDM WG worked first individually, and then collectively, to score current and future maturity states. 'Current state' reflected where we are now for each of the RDaF functions, given the resources in place today. 'Future state' reflected where we could reasonably anticipate being at the end of the two-year funding period under consideration in this MYFP. The result is a clear picture of the current state of RDM and

---

[9] Capability Maturity Model Integration as presented in the COBIT® 2019 Framework: Governance and Management Objectives

of the desired target state. An overview of the average (mean) scores against each Lifecycle Function are provided in Table 2.

Table 2: Current and Target State Scoring per each Overarching RDaF Lifecycle Function. Target State is what we wish to achieve by the end of Fiscal Year 2024-2025

Lifecycle Function	Current State Self-Assessment using CMMI (0-5) (Mean)	Target State Self-Assessment using CMMI (0-5) (Mean)
Envision	2.1	3.4
Plan	1.3	2.3
Generate/Acquire	0.7	1.9
Process/Analyze	1.4	2.2
Share/Use/Reuse	1.6	2.6
Preserve/Discard	1.5	2.2
AVERAGE	**1.4**	**2.4**

A more granular breakdown of the self-assessment scoring by Category is available in Table 3. Connections with specific RDM Services and activities are listed in Column 5, providing additional context for the approach to achieving the Target Maturity Level. Some of these connections are references to Business Cases in the MYFP – *These Business Cases are included in full in the MYFP document.*

1. Stabilization and Growth of the DMP Assistant Platform and Service
2. Lunaris Expansion
3. National PID Strategy Expansion
4. Distributed Data Curation Network
5. Sensitive Data Repository Project

Table 3: Current and Target State Scoring by RDaF Function and Lifecycle Category

Lifecycle Function	Category	Current State Self-Assessment using CMMI (0-5) (Mean)	Target State Self-Assessment using CMMI (0-5) (Mean)	Connections within Alliance RDM Services and Business Cases to Achieve Target States
Envision Review of the overall strategies and drivers of an organization's research data program	Data Governance Structure ● Identification of Goals and Roles [15, 85, 45, 54, 61, 64, 74, 28, 34, 44, 2] ● Data vision and/or data policy [15, 85, 45, 54, 61, 64, 74, 28, 34, 44, 2] ● Data management value proposition [15, 85, 45, 54, 61, 64, 74, 28, 34, 44, 2] ● Data management organization [15, 85, 45, 54, 61, 64, 74, 28, 34, 44, 2] ● Value of data (quantitative or qualitative) [15, 85, 45, 54, 61, 64, 74, 28, 34, 44, 2] ● Legal and regulatory compliance [15, 85, 45, 54, 61, 64, 74, 28, 34, 44, 2] ● Data quality (including Trust and Certification) [15, 85, 45, 54, 61, 64, 74, 28, 34, 44, 2] ● Data privacy [15, 85, 45, 54, 61, 64, 74, 28, 34, 44, 2] ● Data ethics [15, 85, 45, 54, 61, 64, 74, 28, 34, 44, 2]	2	4	There are a range of policies and governance structures in place or under development within the Alliance, although these need to be harmonized across the DRI pillars and expanded to ensure that RDM stakeholders are appropriately represented. To that end, work is advancing on both an RDM Advisory Committee and a Charter to define and help guide the activities of the RDM Network of Experts. We anticipate considerable progress in these areas over the next two years and the RDM Business Cases included in the Multi Year Funding Proposal will advance this category in several areas. The Sensitive Data Repository Project will harmonize sensitive data policy, advancing sensitive data policy and ethics. The proposed National Curation Network will co-develop curation guidance to promote data quality and data management. This will build on community expertise and best practice, as well as FRDR and Borealis's extensive internal curation policies and workflows.The expansion of the DMP Assistant service and platform will reinforce and support most of the objectives of the overarching data governance structure and provide evidence of the organization's commitment to research data.

Community Engagement ● Stakeholder community(ies)[15] ● Communication with stakeholder community(ies)[15] ● Interactions with other organizations[15] ● Cross-community engagement (across domains and sectors)[15] ● Inclusivity in interactions[15]	2.8	3.8	The Alliance RDM team supports a well-established, community-driven RDM Network of Experts that engages a diverse ecosystem of data and research support professionals. Further work is being done on a Network Charter that aims to define the role of the Network and articulate its formal relationship to the Alliance. Several of the Business Cases will significantly advance this category: Growing the service team for Lunaris will allow us to create a community of practice centered on stakeholder engagement, collaboration, and promotion. The Sensitive Data Repository Project includes the co-development of a Project Road Map, developing an approach for collaboration on harmonizing sensitive data processes and policy, and developing training and other resources for dissemination. The approach being taken for the proposed National Curation Network has been co-developed through consultation with stakeholder communities. In addition, curators will engage in cross-institutional planning and resource sharing. The national PIDs consortia are member-driven and include formalized stakeholder engagement processes via the Canadian Persistent Identifier Advisory Committee, international partnerships with organizations also developing PID strategies, and support for communities of practice The DMP Assistant platform, service, and vision is built upon engagement with multiple stakeholder communities, including funders, researchers, research administrators, ethics offices, and an Alliance-supported Expert Group.
Data Culture ● FAIR data principles[15] ● Value of data[15]	3	4	Alliance RDM has adopted a data-centric, life-cycle approach that embraces the principles of Open Science, CARE, TRUST, OCAP, and OA, and fundamentally, making research data

	• Roles and responsibilities[15]			FAIR. The Business Cases included in this MYFP will continue to promote and advance data culture within the organization and ecosystem-wide. The proposed National Curation Network improves support for FAIR data principles and promotes the value of data at local, regional, and national levels. The Sensitive Data Repository Project enables researchers working with sensitive data to comply with the FAIR principles and optimize the likelihood of data reuse. PIDs support all the FAIR principles and are explicitly cited as a foundational component to support findability. Lunaris promotes the value of research data by making it more widely discoverable. Data Management Plans espouse in practical terms the FAIR principles, and include a section on 'roles and responsibilities,' and ultimately raise researchers' and others' awareness of the value of well-managed data.
	Reward Structure • ~~For data management~~ (N) • Value of data workers[15, 1, 78] • ~~Incentives and institutional credit for data sharing and reuse~~ (N) • ~~Disincentives for data sharing~~ (N) • ~~Human Resources (HR) involvement~~ (N)	2	3	Many of these sub-categories were determined to be out of scope for the Alliance. In general, the Alliance seeks to promote incentives for researchers following Open Science practices within institutions and articulate the value of data workers in support of DRI.
	Workforce/Career Paths • Workforce skills inventory[15] • HR's role in data workforce development[15] • Data management training [15, 7, 78]	1.2	2.2	While much effort has been put into ad hoc training and support for community training efforts, more needs to be done to resource, coordinate, and support DRI/RDM training. The proposed National Curation Network will collect data and

	● Workforce preparedness in new and advancing technologies, e.g., HPC, AI, ML, and computation services [15, 16] ● Promotional paths, continual training, and career development [15]			metrics on curation activities and network capacity supports a workforce skills inventory to inform future funding and network continuation or expansion. The Sensitive Data Repository Project will deliver data management training to university staff & researchers and the DMP Assistant service and platform facilitates training of both data HQP and researchers as well, research support stakeholders .The current Data Champions Pilot Project funds awardees as they develop activities that advance awareness, understanding, development, and adoption of RDM tools, best practices, and resources in Canada.
	Data Safety and Security ● Safety and security assurance[15] ● Data inventory [15]	2	3	The Alliance developed a Cybersecurity Framework that addresses this category. From an RDM perspective, the Sensitive Data Repository Project will provide secure technology and infrastructure for researchers to support deposit and controlled access for sensitive datasets. Lunaris allows for metadata about sensitive datasets to be discoverable, even if the data themselves are restricted, allowing for an inventory of sensitive datasets to exist. The proposed National Curation Network will grow capacity for curation professionals to perform data inventories and other risk management activities in consultation with researchers. The FRDR Service Team performs safety, security, and data inventory tasks within the repository. The DMP Assistant service and platform provides an accessible web-based bilingual platform containing standardized templates and guidance to assist researchers at the earliest stages of their research projects in identifying areas of concern and opportunities for improvement with respect to data safety and security.

	Strategy • Organizational data management[15]	2	3	The Alliance's overall data strategy is an ongoing process that will continue to be developed in the coming years. In the Canadian DRI ecosystem, the Sensitive Data Repository Project will create a collaborative framework for organizational cooperation in strategy development for sensitive research data management nationally.
	Data Risk Management • Risk assessment[15] • Risk mitigation and management[15]	2	3	The Alliance developed a Cybersecurity Framework that addresses this category. From an RDM perspective, the Sensitive Data Repository Project will minimize risks to participant confidentiality and privacy by deploying a repository option which is better equipped to manage sensitive research data, co-developing governance, processes, and policies to help address barriers to sensitive research data deposit, data curation, and data sharing. As a part of the Proposed National Curation Network, curators will perform risk management activities and can suggest mitigation strategies to researchers. The FRDR Service Team engages in risk management activities related to the operation of the repository. Working within the nascent Cybersecurity Framework, more needs to be done to assess and address risk to our research data holdings.
Plan The tactical management positioning in an organization for effective	Chain of Control • ~~Documentation~~ (F) • ~~Communication within organization~~ (F)	1	2	Categories were out of scope for this RDaF exercise, but will be included in a future assessment.
	Economics and Costs of Planning • ~~Decision-making tools for data, including cost-benefit analysis~~ (F) • ~~Cost breakdown, i.e., calculation of~~	0	**1**	Categories were out of scope for this RDaF exercise, but will be included in a future assessment.

research data management throughout the research data lifecycle.	~~costs by data lifecycle stage~~ (F)			
	Funding Planning ● Models for provisioning resources, i.e., direct, overhead, or mixed[21]	0	1	This work is being performed organization-wide via the development of the New Service Delivery Model and Funding Model. The Alliance RDM team engages with this broader process as appropriate.
	Data Objects ● Quantitative and qualitative data [6, 12, 13, 51, 50, 15] ● ~~Software, models~~ (F) ● ~~Instruments~~ (F) ● Data publications, journal publications [6, 12, 13, 51, 50, 15] ● Presentations [6, 12, 13, 51, 50, 15] ● Other [6, 12, 13, 51, 50, 15]	1	2	Through its National Repository Services and Curation team, the Alliance has a strong understanding around several of the data objects covered in this category, while others, like software models and instruments, will be in scope for future work. The proposed National Curation Network will support expertise and increased capacity around curating quantitative and qualitative data, software, models, and developing shared resources and guidance to support the effective management of this data more widely. Lunaris ensures that research data objects are discoverable, while the national PIDs consortia facilitate linkages among data objects in the research ecosystem.
	Hardware/Software Infrastructure ● Interoperability [57, 56] ● ~~Persistent instrument identifiers~~ (F)	1	2	The FRDR Service Team piloted a national network of Preservation Service Providers (PSPs) to promote software infrastructure interoperability as well as shared preservation storage and access for pilot participants. National PIDs support enables interoperability among disparate hardware and software infrastructures. Expansion for existing programs and the addition of new programs will be instrumental in advancing the Alliance in this category. We are also exploring targeted integrations between existing platforms, such as the DMP Assistant and the National ARC

				and Repository Services.
	Data Management Planning ● Data management plans (DMPs) [15, 72] ● Lifecycle considerations: living documents or static proposals? [15, 72]	3	4	The Alliance supports a national DMP service and platform (DMP Assistant) that is core to meeting the objectives of this category. Increased support for the DMP Assistant service and platform will allow us to advance in this area. The Sensitive Data Repository Project will contain a mandatory DMP component within access requests, creating a pipeline for sensitive RDM to include DMPs.
	Scientific Metadata Standards ● Sources of standards [66, 23, 24, 53, 83, 92, 65, 15] ● General, domain-specific [66, 23, 24, 53, 83, 92, 65, 15, 59]	2	3	The FRDR Service team applies relevant metadata standards within the repository context and preservation workflows. The Sensitive Data Repository Project will help create guidance on metadata schemas and elements related to sensitive data and access-restriction on data. The proposed National Curation Network will work with researchers to apply relevant metadata standards and enhance existing metadata, often as part of a data deposit process, and create guidance materials for general and domain-specific use. Data Management Plans speak directly to the need for choosing and using appropriate metadata standards to support data reuse and discovery. Lunaris maps metadata to Lunaris' schema for discovery, enhancing and standardising metadata for exporting.
	Assessment and Controls ● Goals/definition of success [52] ● Metrics or metrics structure, tracking use and impact measures [52]	2	3	The Alliance has developed a variety of Key Performance Indicators (KPIs) in accordance with our funder requirements that help us to track usage and impact across our RDM platforms. Each of the Business Cases referenced in this strategy also include KPIs to track and measure their individual success.

Generate/Acquire The generation of raw research data and/or the acquisition of research data by an organization.	Sources of Raw Data ● ~~Generated In-house experimentally or computationally~~ (F) ● ~~Collected from external sources~~ (F)	0	1	Categories were out of scope for this RDaF exercise, but will be included in a future assessment. The creation of the Alliance, with a merged DRI ecosystem, will see greater coordination of the assignment of metadata and other RDM practices with raw data. We expect modest improvements in these areas, but this is a gap to be addressed in a future MYFP.
	Experimental Data Generation ● ~~Specification and recording of instruments and associated metadata~~ (F) ● ~~Description and recording of measurement protocols~~ (F) ● ~~Methods for data and metadata capture and recording~~ (F)	0	1	
	Computational Data Generation ● ~~Commercial and/or custom software~~ (F) ● ~~Methods for computational variables (metadata) capture and recording~~ (F)	0	1	
	FAIR Principles for Data Generated In-House ● Data born FAIR ● Data made FAIR	0.5	1.5	National Repository Services support data being generated on ARC systems being made FAIR. Improved integration between RDM and ARC pillars needed to ensure that generated data is born FAIR. Continued use and adoption of the DMP Assistant will support more data being born FAIR within the Canadian research environment. The proposed National Curation Network will increase capacity locally and nationally for data curation expertise, which leads to the creation of FAIR data both in the active research and data publication stages of the research lifecycle. Improved access to PIDs services will ensure that all datasets are assigned a digital object identifier. The Sensitive Data Repository Project is designed to enhance findability,

				accessibility, and re-usability of sensitive research data.
	External Sources of Data Data acquired FAIR ● Identification, collection, and recording [29, 39, 40, 42, 41, 43, 82, 55, 66] ● Metadata harvesting [76, 86]	1.7	3.3	The Lunaris service harvests metadata from Canadian repositories and makes it available in a federated search. Lunaris' index of metadata is also harvested by other discovery services. Expansion of the service team will greatly improve the Alliance's performance in these areas. National Repository and Curation services ensure that data from external sources made FAIR prior to acquisition.
	Community-Based Standards for Formats ● Standards development organizations/sources [66, 23, 24, 53, 83, 92, 65, 85, 15] ● General, domain-specific [66, 23, 24, 53, 83, 92, 65, 15, 59]	2	3	The Alliance promotes the use of metadata best practices and recommended standards across disciplines. The FRDR Service Team applies community standards and best practices for the use of metadata and file formats within the repository and preservation contexts. Continued training and the development of educational resources in this area is needed.
Process/Ana lyze The actions performed on generated or acquired research data to yield processed research data, and the research data stewardship functions	Data Provenance Original authoritative copy ● Version identification ● Provenance of data derived from other data ● Provenance of scientific records across all the individual outputs ● Timestamping	1	2	Provenance information for datasets deposited in National Repository Services (FRDR) are captured, described, and maintained through automated and manual processes as part of repository and preservation workflows. Versioning procedures in FRDR are preliminary and will need to align with international best practices and standards (e.g., W3C-PROV). The proposed National Curation Network will support data provenance and versioning both for active and published data by capturing relevant metadata and creating documentation such as READMEs and curation logs. Metadata harvesting through Lunaris provides a pointer back to the dataset source and all of this work is enabled by the national PIDs services.

performed by an organization.				More work is needed to expand data provenance tracking for active datasets stored on the National ARC systems.
	Data Architecture ● Design ● Security ● Configuration management ● Hosting and storage ● Use of cloud	0.2	1	While the Alliance developed a Cybersecurity Framework to articulate the security functions, Data Architecture is in a nascent state. More work is needed in this area.
	Software Tools ● ~~Data lifecycle~~ (F) ● ~~Management and analysis~~ (F) ● ~~Commercial and/or custom tools~~ (F) ● ~~System resilience and adaptability~~ (F) ● ~~Maintenance~~ (F)	1	1	Categories were out of scope for this RDaF exercise, but will be included in a future assessment.
	Scientific Workflow Processes and Systems ● ~~Workflow tools~~ (F) ● ~~Laboratory notebooks, i.e., electronic, paper~~ (F)	1	1	Categories were out of scope for this RDaF exercise, but will be included in a future assessment.
	Data Inventory ● Formats and standards [37] ● Catalogues [37] ● ~~Interoperability (across instrument~~ (F) ~~manufacturer file formats)~~ (F)	1.7	2.7	The primary function of Lunaris is to provide a comprehensive catalogue of published research data in Canada. Its harvester normalizes metadata from disparate sources into a consistent format for indexing. Expansion of this service will allow for more Canadian repositories and datasets to be added to the national data inventory. A national catalogue of research data would not be possible without the broad application of PIDs to digital research objects. Growth of the PIDs programs and software integrations between publishing platforms will greatly increase

				our ability to harvest metadata from repositories. The National Repository Services support the application of standards and appropriate formatting of datasets for publication purposes. The proposed National Curation Network will further support and promote the use of formats and standards through the data review process. While some tools do exist to support data indexing in the active phase of research(e.g., RADIAM), more work is needed to create a data inventory within the National ARC systems. DMPs support the creation of data inventories by informing data format decisions and facilitating application of metadata and other standards during research.
	Data Modeling and Analytics ● ~~Processes~~ (F) ● ~~Tools~~ (F)	2	2	Categories were out of scope for this RDaF exercise, but will be included in a future assessment. The current state reflects known work happening within the ARC sector.
	Data Representation/ Models/Structures ● ~~Dynamic data~~ (F) ● General, domain-specific	1	2	This category is well-understood within the data publishing context with our National Repository Services adhering to consistent standards for data representation. More work needs to be done to integrate RDM practices with data in the active research phase being stored on the National ARC systems.
	Data Curation ● Policies and processes [28, 40, 31, 30, 33, 29, 75, 17, 18, 19, 20] ● Human Resource Capacity [28, 40, 31, 30, 33, 29, 75, 17, 18, 19, 20]	3	4	From an organizational perspective, the Alliance's approach to and understanding of data curation is relatively mature. The proposed National Curation Network will support the development of shared resources and increase human resource capacity for curation activities nationally for the broader ecosystem.

				The Sensitive Data Repository Project will be working with institutions and their staff on cultivating sensitive research data curation capacity. Improved support for sound Data Management Planning and the DMP Assistant will facilitate successful curation by encouraging researchers to consider in advance how they will manage their data into repository or preservation storage. Applying DMPs during the data collection, processing, and analysis stages of the research life cycle facilitates data curation by helping to ensure data are prepared for deposit, preservation, discovery, and reuse.
	Metadata ● Types of metadata [86] ● Responsible parties [86] ● Specification of metadata standards [86] ● Linked data structure ● Persistent identification (PID)	2.6	4.4	Metadata is a core area of focus for the Alliance RDM team. The FRDR Service Team performs metadata activities supporting repository and preservation functions. Lunaris maintains a metadata profile composed of commonly used metadata standards with crosswalks to other, often more specialized, schemas. Growing the Lunaris service team will greatly expand our capacity to support this work. The PIDs programs, and in particular the DataCite Canada Consortium, ensure that Canadian research data include digital object identifiers. The development of a National PIDs Strategy and continued growth of support for these programs will expand PID access for Canadian repositories and institutions. The Sensitive Data Repository Project will advance sensitive data metadata schemas and elements.

				The proposed National Curation Network will advance the quality of metadata in a variety of ways. Curators apply metadata standards, enhance existing metadata, and support linked data structures and the use of PIDs (i.e., DOI, ORCiD, ROR). Increased adoption of DMPs will ensure that researchers are guided appropriately in the selection of a metadata standard at their project's outset, leading to better metadata application for data throughout the research lifecycle.
Share/Use/Reuse How research data are disseminated, used, and reused within and outside an organization.	Legal and Licenses ● Ownership of data [15, 85, 45, 54, 61, 64, 74, 28, 34, 44, 2] ● Constraints and encouragement for data use [15, 85, 45, 54, 61, 64, 74, 28, 34, 44, 2] ● Intellectual property rights/restrictions [15, 85, 45, 54, 61, 64, 74, 28, 34, 44, 2] ● Usage agreements/terms/licenses and required permissions [15, 85, 45, 54, 61, 64, 74, 28, 34, 44, 2] ● Terms of service [15, 85, 45, 54, 61, 64, 74, 28, 34, 44, 2] ● Data sharing agreements and licensing [15, 85, 45, 54, 61, 64, 74, 28, 34, 44, 2] ● Data citation [15, 85, 45, 54, 61, 64, 74, 28, 34, 44, 2]	3	4	The National Repository Services require the use of licenses that indicate how a dataset may be reused and provide clear guidance on data citation. The Services themselves have established Terms of Service and other policies. Curators review data for potential copyright and license issues and Terms of Service violations, work with researchers to find solutions, and support data linking and citation by reviewing datasets for presence or use of third-party data. The proposed National Curation Network will advance our capabilities in this category nation-wide. The Sensitive Data Repository project will establish governance mechanisms supported by a harmonized data access agreement template and the negotiated terms and permissions of a co-developed restricted access model. The DMP Assistant provides essential guidance to researchers on legal requirements and considerations for their data.

		3	4	
	Data Publishing ● Repositories [15, 85, 45, 54, 61, 64, 74, 28, 34, 44, 2] ● Referencing data/digital objects from journal articles [15, 85, 45, 54, 61, 64, 74, 28, 34, 44, 2] ● Supplementary material [15, 85, 45, 54, 61, 64, 74, 28, 34, 44, 2] ● Data linking [15, 85, 45, 54, 61, 64, 74, 28, 34, 44, 2]	3	4	The Alliance directly supports two complementary, generalist National Repository Services (FRDR and Borealis) that facilitate data publishing for Canadian researchers. Future work will focus on creating support mechanisms for Canadian domain-specific repositories. The current National Curation Team is an instrumental component of data publishing services. The proposed National Curation Network will support data deposit and publishing within generalist and domain-specific repositories, data linking through the review process, and recommending references for supplementary material, other digital objects or publications as appropriate. Data Management Plans directly increase the likelihood of successful publishing of data into repository storage for discovery and access by asking researchers to consider plans for data deposit, discovery and reuse at the outset of a project. Developing new integrations to support PIDs, and different PID types, in publishing platforms will enhance our repository systems and provide increased opportunities for creating data linkages and linking datasets to journal articles.
	Data Citation ● Citation metrics [52] ● Citation impact [52]	2	3	The National Repository Services have adopted internationally recognized best practices for data citation and usage metrics gathering. Growing a National Curation Network will support sound data citation practice by reviewing data, recommending, and applying appropriate identifiers and other metadata that allow users to reliably cite data.

Internal and External Data Access - Access internally, e.g., the data generator - Access externally [9, 39, 40, 42, 41, 43, 82, 55, 66, 79] - Programmatic access, aka Smart API [36] - Data access vs. data visiting [38]	1	2		The National Repository Services support programmatic access to data through APIs, as well as a variety of traditional downloading methods. Future work should explore the concept of data visiting within our repository systems, especially for very large datasets. The National Curation Network will support external data access through data publication and deposit, improving FAIRness of data. DMPs help researchers map out how and where they will share their research data and to address the steps that will allow this to occur.
Levels of Protection - Unclassified but sensitive information, e.g., de-identification, enclaves [14, 73] - Security classification [14, 73] - Protecting limited data/secure platforms/enclaves [14, 46, 47] - Data anonymization [14, 46, 47]	0	1		The Sensitive Data Repository Project will establish and implement risk frameworks with associated access restriction capacities. The FRDR Curation Team reviews data deposited into FRDR for sensitive information and facilitates the use of embargos when requested by researchers. The National Curation Network will increase capacity nationally for reviewing data for sensitive information and working with researchers to resolve issues, i.e., removing sensitive data or suggesting de-identification methods. Lunaris records indicate whether the dataset is open or restricted. The DMP Assistant guides researchers as they consider the sensitivity of their data and helps them identify appropriate steps to take during and after the research life cycle.

	Applications and Analysis ● ~~Technologies for use and analytics, e.g., AI, ML~~ (F)	0	1	Category was out of scope for this RDaF exercise, but will be included in a future assessment.
	Data Architectures for Application and Use ● ~~Extensibility across communities, including machine-based interactions~~ (F) ● ~~Capturing insights from ML and use of these to improve datasets for future AI applications~~ (F) ● ~~Capturing data performance characteristics~~ (F) ● ~~Location of data (e.g., relative to instruments, in the cloud, transient copies)~~ (F)	0	1	Categories were out of scope for this RDaF exercise, but will be included in a future assessment.
Preserve/Discard The end-of-use and end-of-life provisions for research data in an organization, including records management, archiving, and safe disposal.	Criteria ● Use and impact [39, 40, 43, 48]	2	3	Curators supported through the National Curation Network may be involved in data appraisal, which is a process that assesses use and impact. Curators may work with dedicated archivists or data preservationists to support preservation activities.The FRDR Preservation Coordinator makes appraisal decisions based on established criteria and policies for data deposited in FRDR with support from the Curation Team. PIDs help to support an understanding of data use and impact by facilitating the creation of a networked PIDs graph that shows the interlinkages among elements of the research ecosystem, e.g., researchers, data objects, journal publications, grants, etc.
	Data Sustainability ● ~~Data longevity and support~~ (F)	0	1	Category was out of scope for this RDaF exercise, but will be

	● Orphan datasets (F)			included in a future assessment.
	Storage and Preservation of Data ● Media to store and preserve data [39, 40, 43, 48] ● Data back-up [39, 40, 43, 48] ● Data repositories [39, 40, 43, 48]	2	3	Nascent preservation platforms and processes are connected to the National Repository Services, but more work is needed to formalize and structure this work. DMPs help ensure that data are not only backed up, but also ready for deposit into repository storage, and as a consequence, have the potential to be processed into preservation storage for longer term archiving.
	Moving Data from One Service to Another across Organizations ● Media to store and preserve data ● Roles and responsibilities [22, 30, 39, 40, 42, 33, 32, 35, 43, 61, 74] ● Moving data from one agency to another, e.g., from a funded research agency to an agency with a permanent repository [27, 68] ● Registration of repositories: roles and responsibilities [82] ● Disciplinary archives [27, 68]	1.8	1.8	Work is being done on a pilot basis to support the movement of data among services and organizations as a part of a distributed Preservation Service Provider Network, but more work is needed to formalize and structure this work. Efforts to register Canadian repositories in international registries are being performed on an ad hoc basis through the RDM Network of Experts.
	Retention and Disposition Schedules ● Data archiving, i.e., what is kept and not kept [48, 49, 62] ● Decision processes [40, 48, 49, 62] ● End-of-life issues (F) Example: Responsible party for keeping raw data feeds; Example: Store (or not) raw data, given the large amount of storage needed	1.5	2.5	Both National Repository Services have implemented preservation activities and workflows, although these are still at a relatively immature stage. The Preservation Expert Group within the national RDM Network of Experts develops community resources and guidance materials to support preservation activities, including data retention and disposition activities. Through the National Curation Network, curators may be

	● Deaccessioning/End-of-life [40, 43] ● Recognition of removed data (gravestone) [40]			involved in data appraisal – that is, the process of assessing use and impact to make decisions about what is kept and what is not. Curators may also work with dedicated archivists or data preservationists to support preservation activities. Data Management Plans prompt researchers to begin considering key questions related to data retention and destruction at the early stages of a project. Such planning helps to inform data stewardship decision making, including by curators or archivists.
AVERAGE		1.4	2.4	

REFERENCES providing curated evidence of internal and/or external standards, best practices, and other material that support the analysis of the RDaF, noted above:

1. Advisory Council on Economic Growth. (2017). *Learning Nation: Equipping Canada's Workforce with Skills for the Future*. Government of Canada. https://budget.gc.ca/aceg-ccce/pdf/learning-nation-eng.pdf
2. Alliance Research Data Management (RDM) Advisory Committee. (2022). *Advisory Committee: Terms of Reference*. Digital Research Alliance of Canada.
3. Artefactual Systems Inc. (n.d.-a). *External tools | Documentation (Archivematica 1.13.2)*. Archivematica. Retrieved August 23, 2022, from https://www.archivematica.org/en/docs/archivematica-1.13/getting-started/overview/external-tools/#jhove
4. Artefactual Systems Inc. (n.d.-b). *Ingest | Documentation (Archivematica 1.13.2)*. Archivematica. Retrieved August 23, 2022, from https://www.archivematica.org/en/docs/archivematica-1.13/user-manual/ingest/ingest/#normalize
5. Artefactual Systems Inc. (n.d.-c). *What is Archivematica? | Documentation (Archivematica 1.13.2)*. Archivematica. Retrieved August 23, 2022, from https://www.archivematica.org/en/docs/archivematica-1.13/getting-started/overview/intro/
6. Aspler, J., & Brown, J. (2022). *National PID Strategy*. National Data Services Framework Summit. https://www.youtube.com/watch?v=OM9yBWdgVwQ&t=5162s
7. Bakker, K., Davis, R., Merschrod, E., Bitsuamlak, G., Després, P., Meurs, M.-J., Blais, B., Dickie, E., Rosolowsky, E., Bourque, G., Dupont, B., Pillai-Riddell, R., Brown, S., Estill, L., Sinervo, P., Côté-Lussier, C., Fung, B., Sobie, R., Crompton, C., … D'Arcy, C. (2021). *Meeting the Digital Research Infrastructure Needs of the Canadian Research Community*. Digital Research Alliance of Canada. https://alliancecan.ca/sites/default/files/2022-04/Researcher%20Council%20Priorities%20EN_0.pdf
8. Borealis: The Canadian Dataverse Repository. (2022a, June 23). *About*. Borealis. Retrieved August 24, 2022, from https://borealisdata.ca/about/
9. Borealis: The Canadian Dataverse Repository. (2022b, June 23). *Borealis Preservation Plan*. Borealis. Retrieved August 24, 2022, from https://borealisdata.ca/preservationplan/

10. Borealis: The Canadian Dataverse Repository. (2022c, June 23). *Borealis Technology Infrastructure and Security Information*. Borealis. Retrieved August 24, 2022, from https://borealisdata.ca/technology/

11. Borealis: The Canadian Dataverse Repository. (2022d, June 23). *Borealis Terms of Use*. Borealis. Retrieved August 24, 2022, from https://borealisdata.ca/termsofuse/

12. Brown, J. (2019). *Developing a Persistent Identifier Roadmap for Open Access to UK Research*. Jisc. https://repository.jisc.ac.uk/7840/2/PID_roadmap_for_open_access_to_UK_research.pdf

13. Brown, Josh, Jones, Phill, Meadows, Alice, & Murphy, Fiona. (2022). *The case for investment in a UK persistent identifier strategy: Resilience, insight, and leadership in global research and innovation* (Version v2). Zenodo. https://doi.org/10.5281/ZENODO.6012367

14. Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, & Social Sciences and Humanities Research Council of Canada. (2018). *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans*. Government of Canada. https://ethics.gc.ca/eng/policy-politique_tcps2-eptc2_2018.html

15. Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, & Social Sciences and Humanities Research Council of Canada. (2021). *Tri-Agency Research Data Management Policy*. Government of Canada. https://science.gc.ca/eic/site/063.nsf/eng/h_97610.html

16. CIFAR, AMII, Mila, & Vector Institute. (2017). *The Pan-Canadian AI Strategy*. CIFAR. https://cifar.ca/ai/

17. Clary, E. (n.d.). *FRDR Administrative Tasks*. Internal FRDR Documentation.

18. Clary, E. (2020). *FRDR Curation Onboarding*. Internal FRDR Documentation.

19. Clary, E. (2022). *FRDR Curation Workflow*. Internal FRDR Documentation.

20. Cloudera. (2022). *Cloudera*. Cloudera. https://www.cloudera.com/

21. Committee on Forecasting Costs for Preserving and Promoting Access to Biomedical Data, Board on Mathematical Sciences and Analytics, Committee on Applied and Theoretical Statistics, Computer Science and Telecommunications Board, Board on Life Sciences, Board on Research Data and Information, Division on Engineering and Physical Sciences, Division on Earth and Life Studies, Policy and Global Affairs, & National Academies of Sciences, Engineering, and Medicine. (2020). *Life-Cycle Decisions for Biomedical Data: The Challenge of Forecasting Costs* (p. 25639). National Academies Press. https://doi.org/10.17226/25639

22. Cooper, A., Steeleworthy, M., Paquette-Bigras, È., Clary, E., MacPherson, E., Gillis, L., Wilson, L., & Brodeur, J. (2021). *Dataverse Curation Guide (Version 01)*. Zenodo. https://doi.org/10.5281/zenodo.5579820

23. CoreTrustSeal. (n.d.). *About – CoreTrustSeal*. CoreTrustSeal. Retrieved August 23, 2022, from https://www.coretrustseal.org/about/

24. CoreTrustSeal Board. (2022). *CoreTrustSeal*. CoreTrustSeal. https://www.coretrustseal.org/

25. CoreTrustSeal Standards and Certification Board. (2019). *CoreTrustSeal Trustworthy Data Repositories Requirements 2020–2022* (v02.00-2020-2022). Zenodo. https://doi.org/10.5281/zenodo.3638211

26. Databricks. (n.d.). *Databricks*. Databricks. Retrieved August 22, 2022, from https://www.databricks.com/

27. DeMichelis, N., Poloney, K., Rancourt, Y. G., & Stathis, K. (2021, October 21). *Research data repository migration: process and lessons learned*. [Conference presentation]. Access 2021 Conference, Virtual.

28. Digital Research Alliance of Canada. (n.d.-a). *About DMP Assistant*. DMP Assistant. Retrieved August 29, 2022, from https://assistant.portagenetwork.ca/about_us

29. Digital Research Alliance of Canada. (n.d.-b). *About - FRDR*. Federated Research Data Repository. Retrieved August 23, 2022, from https://www.frdr-dfdr.ca/docs/en/about/

30. Digital Research Alliance of Canada. (n.d.-c). *After Depositing*. Federated Research Data Repository. Retrieved August 23, 2022, from https://www.frdr-dfdr.ca/docs/en/after_depositing/
31. Digital Research Alliance of Canada. (n.d.-d). *Data Champions Pilot Project Call*. Digital Research Alliance of Canada. Retrieved August 23, 2022, from https://alliancecan.ca/en/funding-opportunities/data-champions-pilot-project-call
32. Digital Research Alliance of Canada. (n.d.-e). *Describing Your Data*. Federated Research Data Repository. Retrieved August 23, 2022, from https://www.frdr-dfdr.ca/docs/en/describing_your_data/
33. Digital Research Alliance of Canada. (n.d.-f). *Depositing Data*. Federated Research Data Repository. Retrieved August 23, 2022, from https://www.frdr-dfdr.ca/docs/en/depositing_data/
34. Digital Research Alliance of Canada. (n.d.-g). *Help*. DMP Assistant. Retrieved August 29, 2022, from https://assistant.portagenetwork.ca/help
35. Digital Research Alliance of Canada. (n.d.-h). *Preparing Your Data*. Federated Research Data Repository. Retrieved August 23, 2022, from https://www.frdr-dfdr.ca/docs/en/preparing_your_data/
36. Digital Research Alliance of Canada. (n.d.-i). *Search API*. Federated Research Data Repository. Retrieved August 23, 2022, from https://www.frdr-dfdr.ca/docs/en/advanced/#search-api
37. Digital Research Alliance of Canada. (n.d.-j). *Site Metrics*. Federated Research Data Repository. Retrieved August 23, 2022, from https://www.frdr-dfdr.ca/stats/en/
38. Digital Research Alliance of Canada. (2020a, December 11). *Access and Reuse Policy*. Federated Research Data Repository. Retrieved August 23, 2022, from https://www.frdr-dfdr.ca/policies/en/access_reuse/
39. Digital Research Alliance of Canada. (2020b, December 11). *Collections Development Policy*. Federated Research Data Repository. Retrieved August 23, 2022, from https://www.frdr-dfdr.ca/policies/en/collections_development/
40. Digital Research Alliance of Canada. (2020c, December 11). *Data Retention and Deaccession Policy*. Federated Research Data Repository. Retrieved August 23, 2022, from https://www.frdr-dfdr.ca/policies/en/data_retention/
41. Digital Research Alliance of Canada. (2020d, December 11). *Metadata Harvesting Policy*. Federated Research Data Repository. Retrieved August 23, 2022, from https://www.frdr-dfdr.ca/policies/en/metadata_harvesting/
42. Digital Research Alliance of Canada. (2021a, May 18). *Data Submission Policy*. Federated Research Data Repository. Retrieved August 23, 2022, from https://www.frdr-dfdr.ca/policies/en/data_submission/
43. Digital Research Alliance of Canada. (2021b, November 26). *Preservation Policy*. Federated Research Data Repository. Retrieved August 23, 2022, from https://www.frdr-dfdr.ca/policies/en/preservation/
44. Digital Research Alliance of Canada. (2022a). *Digital Research Alliance of Canada Research Data Management (RDM) Network of Experts Charter DRAFT*. Digital Research Alliance of Canada.
45. Digital Research Alliance of Canada. (2022b). *FRDR CoreTrustSeal TDR Application DRAFT*. Digital Research Alliance of Canada.
46. Digital Research Alliance of Canada, & The Federation. (2022c). *Cybersecurity Strategy & Roadmap*. Digital Research Alliance of Canada.
47. Digital Research Alliance of Canada. (2022d). *The Alliance Data Architecture Strategy Draft*. Digital Research Alliance of Canada.
48. Dorey, J., Hurley, G., & Knazook, B. (2022). *Appraisal Guidance for the Preservation of Research Data*. Zenodo. https://doi.org/10.5281/zenodo.5942236
49. Dorey, J., Hurley, G., Rancourt, Y. G., (2022, September 9). *Workshop on preservation appraisal*. [Workshop]. Curation Expert Group & Preservation Expert Group. https://www.eventbrite.ca/e/workshop-on-preservation-appraisalatelier-sur-la-preservation-a-long-terme-tickets-380468931817

50. FAIR Digital Object Fabric Interest Group. (n.d.). *FAIR DO: RDA Recommendations & Outputs*. Research Data Alliance. Retrieved August 24, 2022, from https://www.rd-alliance.org/group/data-fabric-ig/wiki/fair-do-rda-recommendations-outputs
51. FAIR Digital Objects Forum. (n.d.). *Our Working Groups*. FAIR Digital Objects Forum. Retrieved August 24, 2022, from https://fairdo.org/wg/
52. FAIR Metrics Group. (2017). *FAIRMetrics*. Github. http://fairmetrics.org
53. First Nations Information Governance Centre. (2022). *The First Nations Principles of OCAP*. First Nations Information Governance Centre. https://fnigc.ca/ocap-training/
54. FRDR-DFDR. (n.d.). *Terms of Use*. FRDR-DFDR. Retrieved August 22, 2022, from https://www.frdr-dfdr.ca/policies/en/terms_of_use/
55. Global Biodata Coalition. (2022). *Global Core Biodata Resource*. Global Biodata Coalition. https://globalbiodata.org/scientific-activities/gcbr-selection/
56. Global Open Research Commons (GORC) International Model Working Group. (2020). *Global Open Research Commons (GORC) International Model Working Group*. Research Data Alliance. https://www.rd-alliance.org/groups/gorc-international-model-wg
57. Global Open Science Cloud Initiative. (2021). *Invitation to Collaborate on the Global Open Science Cloud Initiative*. CODATA. https://codata.org/initiatives/decadal-programme2/global-open-science-cloud/
58. GO FAIR Implementation Network. (n.d.). *FAIR Principles*. GO FAIR. Retrieved August 22, 2022, from https://www.go-fair.org/fair-principles/
59. GO FAIR Implementation Network. (n.d.). *Implementation Networks*. GO FAIR. Retrieved August 30, 2022, from https://www.go-fair.org/implementation-networks/
60. IBM. (n.d.). *IBM Watson is AI for smarter business*. IBM. Retrieved August 22, 2022, from https://www.ibm.com/watson?utm_content=SRCWW&p1=Search&p4=43700050290113193&p5=e&gclid=Cj0KCQjw_viWBhD8ARIsAH1mCd6rkyg_el3Z5n3CnEKG0dWDXBnVR573eT_1hsSyoRDTqQIUxnnmj08aAgxHEALw_wcB&gclsrc=aw.ds collaboration
61. Kawula, B., & Newson, K. (2022, June 23). *Dataverse Home*. Scholars Portal Dataverse Spot Docs. Retrieved August 23, 2022, from https://spotdocs.scholarsportal.info/display/DAT/Dataverse+Home
62. Knazook, B. (n.d.). *Long-term Preservation Appraisal Checklist. Internal FRDR Documentation.* Retrieved August 24, 2022, from https://docs.google.com/document/d/1jKaH5Yu2EHd7sYjOY3OBR-V4EahRa_lengmyKQc-yVU/edit?usp=sharing
63. Lander, E. S. (2021). *Office of Science and Technology Policy Data Management Policy*. Office of Science and Technology Policy. https://www.whitehouse.gov/wp-content/uploads/2022/02/2021-Public-Access-Congressional-Report_OSTP.pdf
64. Leahey, A., & Newson, K. (2021, January 5). *Current Projects*. Scholars Portal Dataverse Spot Docs. Retrieved August 23, 2022, from https://spotdocs.scholarsportal.info/display/DAT/Current+Projects
65. Lin, D., Crabtree, J., Dillo, I., Downs, R. R., Edmunds, R., Giaretta, D., De Giusti, M., L'Hours, H., Hugo, W., Jenkyns, R., Khodiyar, V., Martone, M. E., Mokrane, M., Navale, V., Petters, J., Sierman, B., Sokolova, D. V., Stockhause, M., & Westbrook, J. (2020). The TRUST Principles for digital repositories. *Scientific Data*, *7*(1), 144. https://doi.org/10.1038/s41597-020-0486-7
66. Lister, A., Thurston, M., Sansone, S.-A., Rocca-Serra, P., Granell, R., Batista, D., & Dauga, D. (n.d.). *FAIRsharing* [2022/08/22]. FAIRsharing. https://fairsharing.org/
67. MoreBrains Cooperative. (2022). *Towards a national PID strategy for Canada: CPIDAC PID consultancy final report*. MoreBrains Cooperative. https://docs.google.com/document/d/1LkvXg64tUofZ_X87bpLQ0CRRac1XK9my/edit#
68. Mukunda, K. (2021, September 1). *Transitioning from Radar to FRDR and Summit | SFU Library.* SFU Library. Retrieved August 24, 2022, from https://www.lib.sfu.ca/help/publish/research-data-management/radar-frdr

69. National Research Council Canada. (n.d.). *Canadian Astronomy Data Centre research facility*. Government of Canada. Retrieved August 22, 2022, from https://nrc.canada.ca/en/research-development/nrc-facilities/canadian-astronomy-data-centre-research-facility

70. National Research Council Canada. (2021). *Artificial Intelligence for Design Challenge program*. Government of Canada. https://nrc.canada.ca/en/research-development/research-collaboration/programs/artificial-intelligence-design-challenge-program

71. Natural Resources Canada. (n.d.). *The Federal Geospatial Platform*. Government of Canada. Retrieved August 22, 2022, from https://www.nrcan.gc.ca/earth-sciences/geomatics/canadas-spatial-data-infrastructure/geospatial-communities-and-canadian-geosecretariat/federal-geospatial-platform/11031

72. Office of Science and Technology Policy. (n.d.). *Office of Science and Technology Policy*. The White House. Retrieved August 22, 2022, from https://www.whitehouse.gov/ostp/

73. Office of Vice-Principal Research and Innovation. (n.d.). *McGill University Research Data Management Strategy Draft v1.0*. McGill University. Retrieved August 24, 2022, from https://www.mcgill.ca/drs/files/drs/mcgill_rdm_strategy_-_draft_v1.0.pdf

74. Ontario Council of University Libraries. (n.d.-a). *Scholars Portal Dataverse*. Scholars Portal. Retrieved August 23, 2022, from https://learn.scholarsportal.info/all-guides/dataverse/

75. Ontario Council of University Libraries. (n.d.-b). *Your Research Data*. Scholars Portal Dataverse. Retrieved August 24, 2022, from https://learn.scholarsportal.info/all-guides/dataverse/datasets/#Datasets-and-Restricted-File-Permissions

76. OpenAIRE. (n.d.). *OpenAIRE Guidelines*. OpenAIRE. Retrieved August 30, 2022, from https://guidelines.openaire.eu/en/latest/

77. Pecoskie, J. (n.d.). *Interim Report for the COVID-19 Data Curation Funding Initiative*. Internal Alliance Documentation. Retrieved August 23, 2022, from https://docs.google.com/document/d/11t0txgzUaHL8QmZQ3PnyNtJozDoz5LBhfiE2S-HkiTs/edit?usp=sharing

78. Pérez-Jvostov, F., Iron, K., Khair, S., Sahrakorpi, S., & Zhang, Q. (2021). *Researcher Needs Assessment: Summary of What We Heard*. Digital Research Alliance of Canada. https://alliancecan.ca/sites/default/files/2022-03/needsassessment_alliance_20220126.pdf

79. ProQuest. (2022). *ProQuest*. ProQuest. https://www.proquest.com/

80. Rauber, A., Asmi, A., Uytvanck, D. V., & Proell, S. (2016). *Data Citation of Evolving Data: Recommendations of the RDA Working Group on Data Citation (WGDC)*. https://doi.org/10.15497/RDA00016

81. re3data. (n.d.-a). *Canadian repositories*. Re3data. Retrieved August 24, 2022, from https://www.re3data.org/search?query=canada

82. re3data. (n.d.-b). *Re3data*. Re3data. Retrieved August 22, 2022, from https://www.re3data.org/

83. Russo Carroll, S., Hudson, M., Chapman, J., Figueroa-Rodriguez, O. L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., & Walker, J. (2018). *The CARE Principles for Indigenous Data Governance*. Global Indigenous Data Alliance. https://www.gida-global.org/care

84. Simons, Natasha, & Burton, Adrian. (2021). *National PID Strategies: Australia*. https://doi.org/10.5281/ZENODO.5024148

85. Standards Council of Canada. (2020). *Canadian Data Governance Standardization Collaborative*. Standards Council of Canada // Le Conseil Canadien Des Normes. https://www.scc.ca/en/flagships/data-governance

86. Stathis, K. (n.d.). *FRDR Metadata Application Profile (MAP)*. Internal FRDR Documentation. Retrieved August 23, 2022, from https://docs.google.com/spreadsheets/d/1WbMY4eNDeF0FvKX5CYjgc8HvnduuQBXK3yNotHXX3B4/edit?usp=sharing

87. The Portage Preservation Expert Group, & The Canadian Association of Research Libraires Digital Preservation Working Group. (2020). *Digital Preservation and NDRIO: A White paper*. The Canadian Association of Research Libraires Digital Preservation Working Group. https://alliancecan.ca/sites/default/files/2022-03/ndrio-preservation-white-paper.pdf

88. The President & Fellows of Harvard College. (2022, January 31). *Dataset + File Management*. Dataverse Project. Retrieved August 24, 2022, from https://borealisdata.ca/guides/en/latest/user/dataset-management.html#permissions

89. The President & Fellows of Harvard College. (2022, April 6). *Finding and Using Data*. Dataverse Project. Retrieved August 24, 2022, from https://borealisdata.ca/guides/en/latest/user/dataset-management.html#permissions

90. Vouk, I., Burcher, R. S., Johnston, C. M., Jenkinson, R. W., Saad, D. A., Gaiot, J. S., Benoy, G. A., Robertson, D. M., & Laitta, M. (2018). *Binational dataset used to develop nutrient SPARROW models for the Midcontinental region of Canada and the United States* (p. 1 zip file containing readme.txt, data1_midcont.txt, and geospatial vector data files midcont_reaches.zip and midcont_catchments.zip, 1 fichier zip contenant readme.txt, data1_midcont.txt, et fichiers de données géospatiales vectorielles midcont_reaches.zip et midcont_catchments.zip) [Zip]. National Research Council of Canada. https://doi.org/10.4224/300.0001

91. Western University, & Fraunhofer Institute of Chemical Technology. (n.d.). *The Fraunhofer Innovation Platform for Composites Research*. The Fraunhofer Innovation Platform for Composites Research. https://www.eng.uwo.ca/fraunhofer/

92. Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18

*Strikethrough ~~"Strikethrough"~~ Note for Table 3: The Subcategory elements that are formatted with strikethrough font are considered Not in Scope for this initial stage of the strategic development. Not in Scope elements are further described by an N (No, Not considered relevant for the Alliance, or needs to be considered in the broader Alliance context), or an F (Future, To be considered in the next stage of RDaF review).

Our current state averages 1.4, which on the CMMI scale is defined as "Initial." This acknowledges some activities are underway, but many are incomplete and in need of more work. The target state score of 2.4 envisions us moving RDM to "Managed" on the CMMI scale, a state where activities are described as basic, yet complete. Given the scope of RDM, this self-assessment scoring is appropriate and shows the areas of targeted growth that are necessary and possible for us to build upon.

What stands out from this process is the fact that every RDaF Function in both the Current- and Target-state self-assessments is addressed. This illustrates that the focus for RDM spans the whole Research Lifecycle and all its Functions, and as noted in the RDaF, "the Functions should be performed concurrently and continuously to create a dynamic operational culture that addresses the research data management needs."[10] It is therefore essential to focus on the whole of the research data lifecycle to ensure that appropriate consideration is provided to the activities that fall under each Function. Clear gap areas, which were determined to be largely out-of-scope for this iteration of the RDaF exercise, are Categories/Subcategories that intersect with activities that are currently under the primary purview of ARC. These Categories need to be further explored in a future RDaF exercise and we anticipate that more work will be needed to ensure that RDM and ARC activities are considered holistically from a DRI-wide perspective. This is most evident in Figure 4, which visualizes findings from the "Generate-Acquire" Lifecycle Function.

Current and Future states for RDM are further explained in a series of visualizations:

---

[10] Hanisch. J. R., Kaiser. L. D and Carroll. C. B. NIST Special Publication 1500-18. RdaF. Retrieved from: Https://doi.org/10.6028/NIST.SP.1500-18
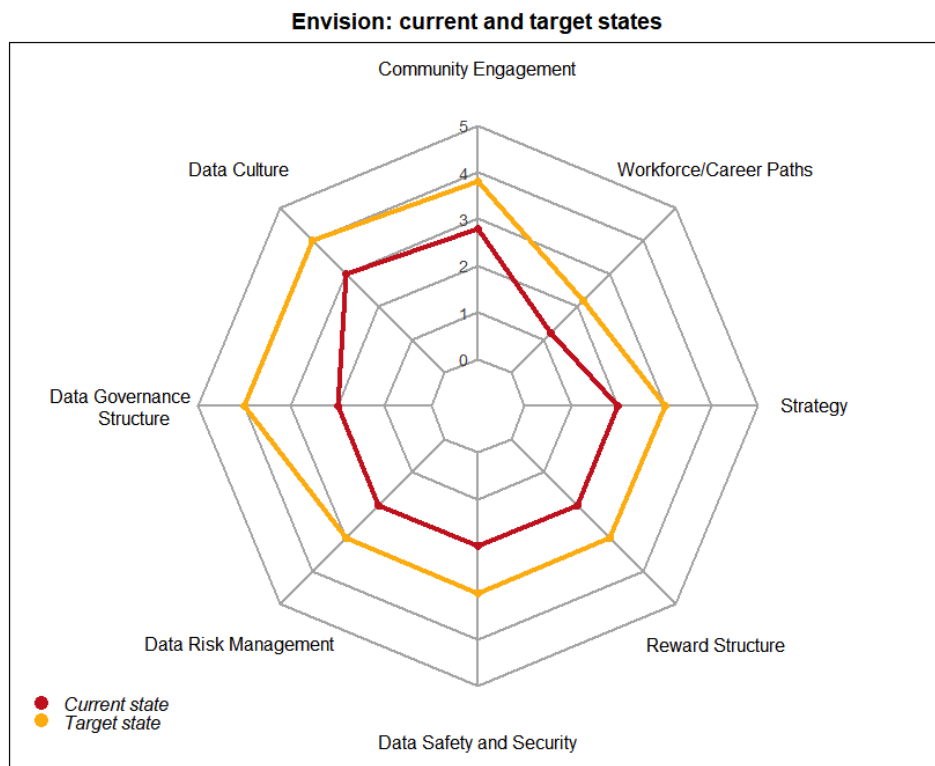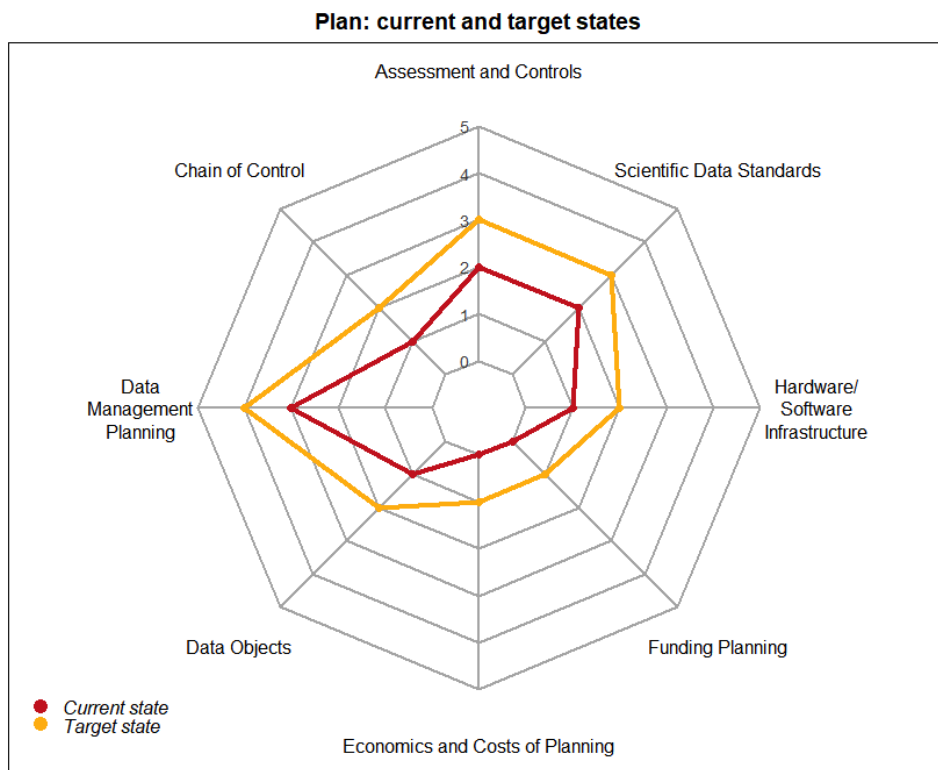
Envision: current and target states

Figure 2: Envision

Plan: current and target states

Figure 3: Plan
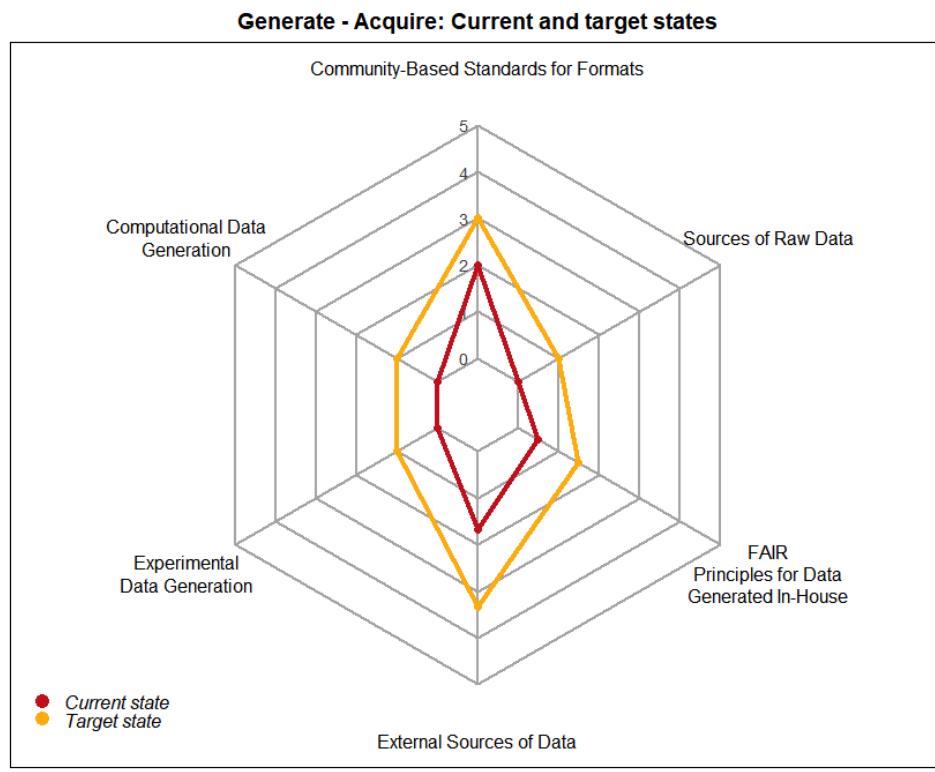
Figure 4: Generate - Acquire
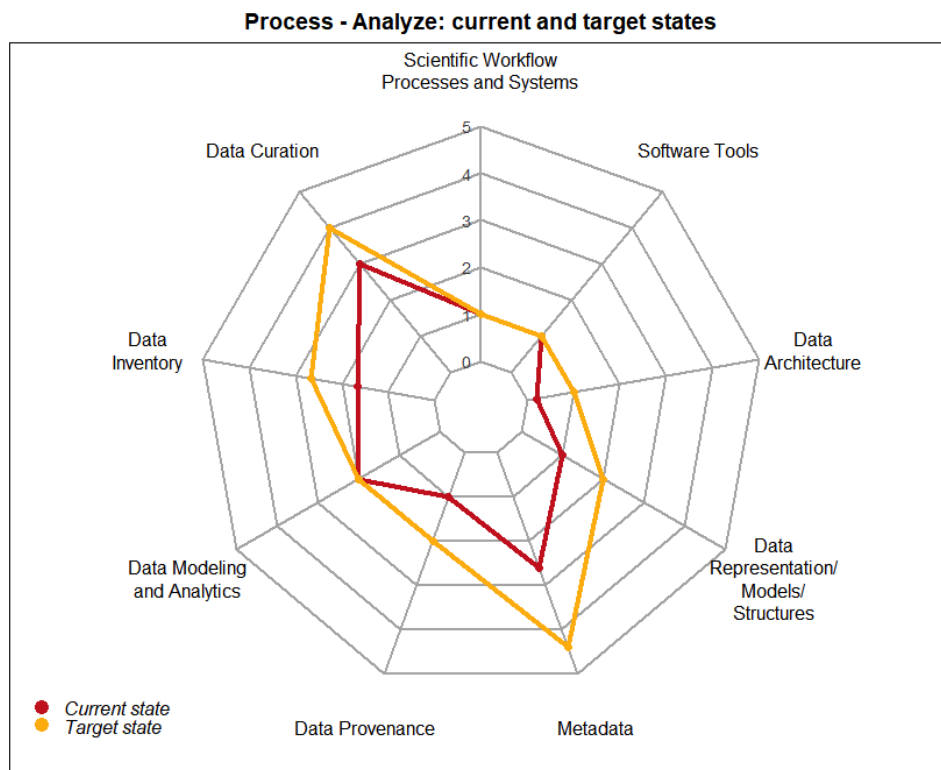
Figure 5: Process-Analyze

Figure 6: Share-Use-Reuse

Preserve - Discard: current and target states
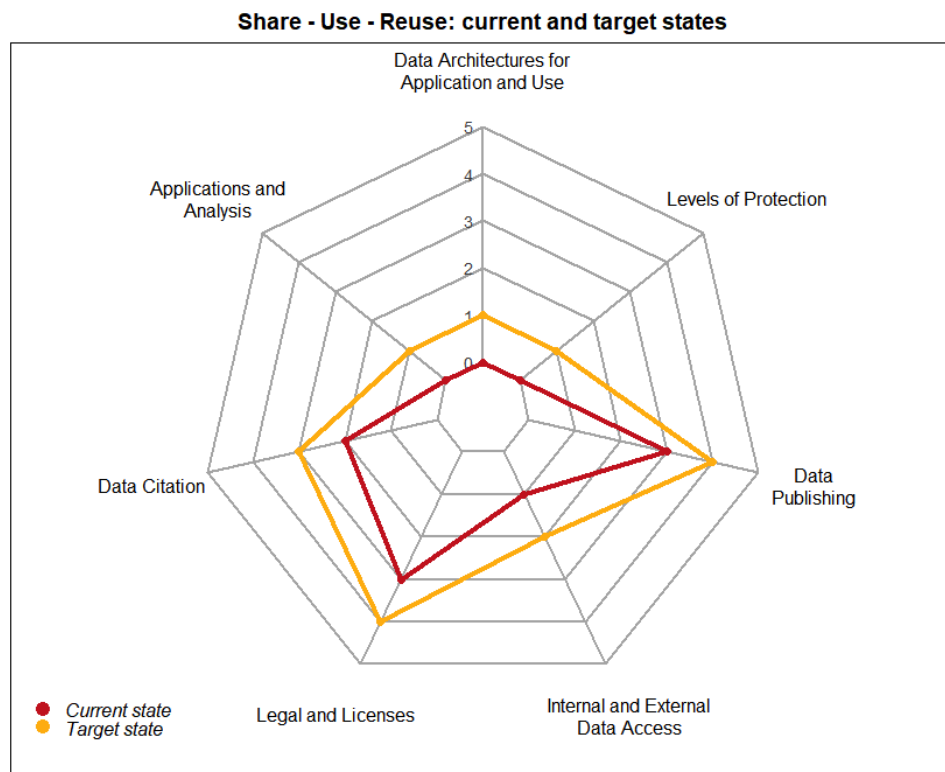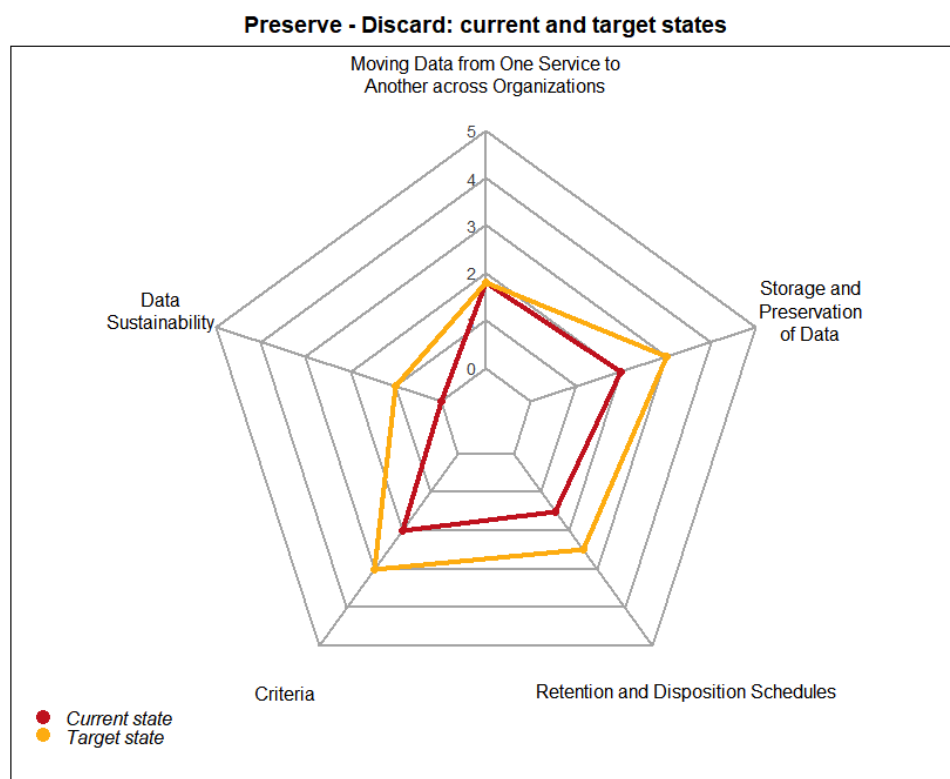
Figure 7: Preserve - Discard

In the next section, we provide an overview of the important community connections and governance structures necessary to undertake this data stewardship work.

RDM Community Context and Governance Structures

In this section, we focus on the significant pan-Canadian community connections and governance structures that are developed as part of this research data stewardship work, including those that will be emphasized through current and upcoming initiatives.

National RDM Support Team

The national Alliance RDM Team was constituted to help researchers meet their RDM needs by facilitating and coordinating the development and deployment of a diverse array of RDM platforms, supports, services, and training, also closely aligned with key stages of the research data life cycle. The work of the RDM Expert Groups is coordinated and supported nationally by this Team, whose role is essential to the Network's continued productivity and success. The national support provided to the pan-Canadian grassroots Network underpins the success of the Alliance by earning strong community buy-in and effectively meeting researchers' and institutions' RDM needs.

National RDM Network of Experts

The work of the Alliance RDM Team is community-driven and foundational, drawing upon a well-established and diverse ecosystem of research support professionals, including, but not limited to, librarians, data management professionals, institutional research officers, research ethics professionals and others who work to support and advance RDM activities and services. This Network contributes to a broad vision for RDM through their generous voluntary engagement in a number of RDM Expert Groups reflective of the stages of the research data life cycle (Figure 8). This pan-Canadian Network of Experts is a key driver of progress in the Canadian RDM landscape. The investment and commitment of local institutions in sharing this nation-wide pool of expertise, *pro bono*, is significant and has had a strong multiplier effect in the development and delivery of national RDM support to researchers and institutions in Canada. Currently, the RDM Network of Experts consists of over 160 individuals from over 70 institutions/organizations. Work is underway on a Charter document outlining the roles, responsibilities, and relationships between the Alliance and the RDM Network of Experts. This is anticipated to be completed in Fall 2022.
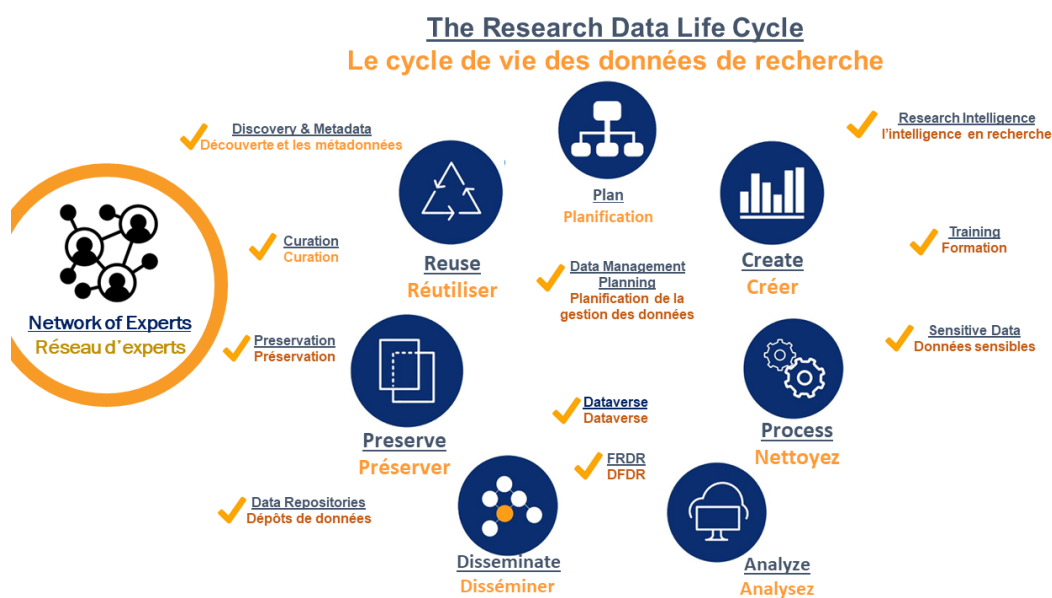
The Research Data Life Cycle
Le cycle de vie des données de recherche

Discovery & Metadata
Découverte et les métadonnées

Research Intelligence
l'intelligence en recherche

Curation
Curation

Plan
Planification

Create
Créer

Training
Formation

Reuse
Réutiliser

Data Management Planning
Planification de la gestion des données

Network of Experts
Réseau d'experts

Preservation
Préservation

Sensitive Data
Données sensibles

Dataverse
Dataverse

Preserve
Préserver

Process
Nettoyez

Data Repositories
Dépôts de données

FRDR
DFDR

Disseminate
Disséminer

Analyze
Analysez

Figure 8: National Data Stewardship Support - Life-cycle-aligned Expert Groups

Proposed Research Data Management Advisory Committee

As the Alliance assumes national responsibility for supporting research data management, increased engagement and communication with key partners is essential for maintaining awareness of the landscape and ensuring that developing strategy and services respond to community needs. The proposed Alliance RDM Advisory Committee (RDM AC) will convene representatives from key stakeholder groups to develop and align strategy, facilitate communication and feedback, promote awareness and learning, and focus attention on emerging issues of national importance.

The RDM AC will connect to Alliance governance structures and decision-making. It will provide advice and direction to the Alliance's Strategy and Operations teams regarding the development and maintenance of the Alliance's RDM roadmap, strategic planning, and annual priorities; the development and implementation of Alliance RDM programs and funding opportunities; and best practices, communications, and other engagement with the community. It will support inclusive involvement of the Canadian RDM community in Alliance decisions and activities, and promote active engagement and information exchange between the Alliance and the wider RDM community. Finally, it will identify opportunities for collaboration among Canadian organizations and international groups.

First Nations, Inuit, and Métis Relationship Building

The Alliance continues to develop and expand communications, relationship building, and outreach with First Nations, Inuit, and Métis Nation communities as part of a broader Indigenous engagement strategy.

The Alliance is in early stages of relationship-building with Indigenous partners. First Nations, Inuit, and Métis Nation communities have a strong interest in the development and usage of DRI. A broad set of priorities in relation to Alliance work and Indigenous engagement have been identified, including: equitable access to DRI; Indigenous Data Sovereignty and the roles and responsibilities of data repositories in relation to Indigenous data; access to RDM and DRI training and education resources, particularly for youth; co-development of RDM resources and services and the development of processes to support co-development; and interoperability between Indigenous and settler DRI systems.

Advancing work in the above-noted areas is contingent upon the establishment of trustworthy relationships between the Alliance and First Nations, Inuit, and Métis Nation communities.

The Alliance has made First Nations Principles of OCAP® training available to all staff to help enable broad organizational understanding of the importance of data sovereignty and information self-governance and how First Nations research and data-related principles intersect with RDM and DRI.

International Outreach and Engagement

Engaging with international organizations in various contexts is key to the Alliance's (and by extension Canada's) success and role as an innovator in the DRI ecosystem. Strengthening and enlarging national and international partnerships not only positions the Alliance in a leadership role on the global stage, but also facilitates the adoption of best practices on national, regional and local scales. It also enables integration of both national and international processes and requirements facing researchers (e.g., data security, privacy, Indigenous Data Sovereignty) into service offerings. In the same way that the Alliance can see substantial benefit from the international community, its own innovation over the coming years will inform and enhance best practices globally. International partnerships can be leveraged to enhance not only the Alliance's service catalogue, but also funding and research partnerships. This "two-way street" will engender the greatest benefit for the research community.

The Alliance continues to maintain and grow international connections through membership and active participation in International Organizations such as: Research Data Alliance (RDA), RDA

Region of the Americas, CODATA International, Research Software Alliance (ReSA), World Data System, GO FAIR, and domain-specific organizations of importance to the Canadian research community. These efforts contribute to extending and socializing the goals and deliverables of both the Alliance and these international groups, align with the work of members of the RDM Network of Experts, and inform broader outreach and training initiatives. The Alliance Director of International Relations is a close partner in this work, and international partnerships are governed by the Alliance International Relations Policy and the associated International Partnership Review Framework.

The following section reviews the RDM-focused activities that are underway currently and those that are proposed in the MYFP Business Cases. These stem directly from the RDaF self-assessment and other dialogue with significant community partners (e.g., the RDM Network of Experts).

Current RDM Activities and Initiatives

As described above, the RDaF framework describes RDM as spanning the whole Research Lifecycle, and states that "the [RDaF] Functions should be performed concurrently and continuously to create a dynamic operational culture that addresses… research data management needs."[11] This focus on *concurrent* and *continuous* applies well to the RDM activities, initiatives, and pilots developed, delivered, and promoted through the Alliance and various partners. This interconnectedness is a hallmark of RDM, and contributes to a more holistic view of the research data life cycle and the diverse and growing range of support the Alliance and its partners provide to both institutions and researchers.

Partnering in Service Delivery

The Alliance relies on a range of stakeholders and other partners to support the ongoing development, delivery, and promotion of a range of RDM tools, services, platforms, standards and best practices, and training in Canada. These include, but are not limited to: the Data Management Planning (DMP) Assistant, Borealis, the Federated Research Data Repository (FRDR), the pilot Preservation Service Provider Network (PSP), Lunaris, and programs to support Persistent Identifiers (PIDs), standards and best practices, and training. Some of the partners whose financial and/or in-kind contributions are substantial are University of Alberta,

---

[11] Hanisch. J. R., Kaiser. L. D and Carroll. C. B. NIST Special Publication 1500-18. RdaF. Retrieved from: Https://doi.org/10.6028/NIST.SP.1500-18

OCUL Scholars Portal (University of Toronto), University of Saskatchewan, University of British Columbia, Simon Fraser University, and the Canadian Research Knowledge Network (CRKN).

DMP Assistant

The DMP Assistant[12] is a national, online, bilingual data management planning tool developed by the Alliance in collaboration with the University of Alberta Library (UAL) to assist researchers in preparing effective data management plans (DMPs). This tool is freely available to all researchers and end-users and assists with the development of DMPs through a series of key data management sections and questions spanning the research lifecycle, supported by best-practice guidance and examples. Canadian postsecondary institutions may also create institutional accounts that allow them to directly support their researchers using the tool, customize templates with branding and institution-specific guidance, and access metrics on DMP Assistant usage at their institution. DMP tools offer many advantages both to researchers and administrators: To researchers, they provide a streamlined, accessible, and easy-to-use interface for guiding the development of DMPs; to postsecondary institutions and their administrators, the local capture of individual research project plans provides insight for capacity planning as well as the ability to create customized templates and guidance to support local needs across a range of research disciplines and methodologies.

Developing and implementing effective data management plans is a foundational component of sound RDM, an international best practice, and a growing requirement of both institutions and funders, including the Canadian Tri-Agencies as outlined in their Research Data Management Policy.[13] A well-supported national DMP tool is essential for the success of the Tri-Agency Policy, particularly in light of the recent release of the Tri-Agencies' first set of funding calls requiring DMPs,[14] beginning in Fall 2022. A nationally-hosted DMP tool, such as the DMP Assistant, saves institutions the cost and trouble of establishing, maintaining, and supporting their own platform. It also allows for more efficient and strategic investments in consistent guidance, training, and platform development, and better governance through nationally-representative advisory bodies, such as the DMP Assistant Steering Committee.

Since its launch in 2015, the DMP Assistant has seen strong adoption by the Canadian research community, with 13,762 user accounts, 75 institutional accounts, and 8,435 DMPs created to date. In 2021, the number of DMPs created increased by 27.8%, and the number of institutions using DMPs increased by 35.2%. The predicted rise of users and DMPs requires an expansion of the team to maintain the quality of service the DMP Assistant provides.

---

[12] DMP Assistant - https://assistant.portagenetwork.ca/
[13] https://www.science.gc.ca/eic/site/063.nsf/eng/h_97610.html
[14] https://science.gc.ca/eic/site/063.nsf/eng/h_547652FB.html

Based on findings from the RDaF and the direction provided by the DMP Assistant Steering Committee and the Data Management Planning Expert Group, we aim to extend the functionality of the tool to support machine actionability that will allow the tool to connect with other systems in the ecosystem (e.g., Canadian Common CV), and to support information sharing between funder platforms and researchers. Other community-identified improvements include embedding DMPs into existing national service provision, such as the Alliance Federation Resource Allocation Competition; continuing to expand the user base and increase adoption by both researchers and organizations; refining the platform UX and UI for increased accessibility and usability; and migrating the service to more sustainable infrastructure (e.g., the Alliance Federation National Host Sites), thereby providing increased security and scalability.

National Repository Services

Data repositories are an essential part of our national Digital Research Infrastructure. They allow researchers to publish curated datasets as a valued research output in compliance with growing funder and journal policies. They enable the preservation, discovery and reuse of data to support new research. They enable verification and reproducibility of existing research. The 2021 Tri-Agency Research Data Management (RDM) policy includes data deposit into trustworthy digital repositories as one of its three main pillars. Canadian researchers are well-served by the two complementary, national, bilingual repository options (Borealis and FRDR)[15] supported by the Alliance, but continued investment to maintain and strategically grow these services is essential to ensure the DRI ecosystem is prepared to meet increased demand for data publication, curation, and preservation driven by the new Tri-Agency policy. In particular, we need to grow our national services to support diverse disciplinary needs so that researchers do not need to conform to a one-size-fits-all model of data publication.

Borealis

Borealis[16] is a bilingual, multidisciplinary, secure, Canadian research data repository, supported by the Alliance and academic libraries and research institutions across Canada. The service is provided by Scholars Portal based at the University of Toronto. Borealis supports open discovery, management, sharing, and preservation of Canadian research data using the Dataverse repository platform. Researchers can release and share data openly or privately, and visualize and explore data. Borealis supports discovery of various data types with extensible metadata blocks. Borealis provides a common repository platform to research institutions, backed by robust shared storage infrastructure and common platform development and maintenance. Its distributed service and sliding scale costing model is equitable and sustainable

---

[15] Federated Research Data Repository (FRDR) - https://www.frdr-dfdr.ca/repo/

[16] Borealis - The Canadian Dataverse Repository - https://borealisdata.ca/

over the long term and contributes to securing the match required as a part of the New Service Delivery Model.

As of June 2022, over 60 Canadian postsecondary institutions participate in Borealis, with over 8,000 datasets deposited and published by individual researchers, research projects, organizations, institutions, groups, and departments.

Federated Research Data Repository (FRDR)

FRDR is a bilingual, secure, general-purpose data repository that accepts deposits from Canadian researchers in any discipline. FRDR is custom built to support the publication and preservation of big data via integration with Globus File Transfer and parallelized Archivematica instances. FRDR allows researchers to collaborate on deposits, and currently supports special collections by research groups with organizational branding and links to external web content (e.g., Global Water Futures, SuperDARN, Szechtman Lab Collection). Each dataset is reviewed by a member of FRDR's Curation Team, who work directly with researchers to enhance the FAIRness (findability, accessibility, interoperability, and reusability) of new deposits. Hosted on Canada's national ARC infrastructure, with repository storage spread across multiple hosting sites, FRDR is well-suited to support the data publication needs of researchers working at the HPC scale that individual institutions are not well equipped to meet.

As of June 2022, FRDR has nearly 650 registered users and 319 datasets published totalling over 110 TB of Canadian research data curated and openly available for discovery and access. Underscoring the importance and value of this repository, data files housed in FRDR have been downloaded over 6.5 million times, to date.

Lunaris

Lunaris, formerly the FRDR Discovery Service, provides powerful functionality to search for Canadian research data using metadata harvested from repositories across Canada. This federated tool aggregates harvested metadata and provides a single point of search for datasets from repositories in universities, government organizations, research organizations, and national multidisciplinary initiatives (including FRDR's own repository). The platform integrates both text- and map-based searching to offer a dynamic and intuitive discovery experience.

As of June 2022, Lunaris indexes metadata records from over 100 Canadian data repositories, making over 75,000 datasets discoverable through a single, national search platform, and, importantly, supporting broader discovery through international aggregation portals. An integral part of the suite of RDM services offered by the Alliance, the Lunaris discovery service is a valuable national resource for finding and accessing Canadian research data, thereby supporting their global visibility and enabling dataset reuse.

Persistent Identifiers (PIDs)

Support for and promotion of Persistent Identifiers such as DOIs (Digital Object Identifiers) and ORCID (unique researcher identifiers), and engagement internationally with the Research Data Alliance[17] on a variety of standards and best practices, are other areas of activity that support the Canadian research community. With some financial support from the Alliance, the Canadian Research Knowledge Network (CRKN) supports two membership-supported PID consortia: 1) the DataCite Canada Consortium; and 2) ORCID Canada. DataCite provides member institutions with the ability to register digital object identifiers (DOIs) for datasets and other research outputs, while ORCID iDs are PIDs for individual people. These two core PIDs identify Canadian scholars, scholarship, and collaborations to ensure that research outputs (e.g., datasets, publications) are FAIR. Each consortium has a community-representative governance body, with the PIDs ecosystem as a whole represented through the Canadian Persistent Identifier Advisory Committee (CPIDAC). These consortia are still in a growth phase with exceptional opportunities to increase DOI and ORCID iD uptake at Canadian institutions.

Standards and Best Practices

All platforms and services, and our work with domestic and international partners and initiatives, follow and contribute to advancement of RDM standards and best practices which in turn support achievement of the FAIR,[18] CARE,[19] and TRUST[20] Principles. The DMP Assistant is a good example of this effort. There is growing awareness nationally and internationally of the importance of machine-actionable data management plans (maDMPs). The internationally-deployed DMP Roadmap codebase generates maDMPs aligned with emerging international best practices and paves the way for integration with other research-support endeavours (e.g., ethics approval processes, research management information systems [RMI], granting council compliance), all aimed at improving efficiency and lowering administrative burden on researchers.

Training

The Alliance is a key player in the advancement of RDM training in Canada. Work will continue to focus on a variety of RDM training initiatives, including but not limited to historically

---

[17] Research Data Alliance https://www.rd-alliance.org/

[18] Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Bouwman, J. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data, 3*.

[19] CARE Principles for Indigenous Data Governance - https://www.gida-global.org/care

[20] TRUST Principles - https://www.rd-alliance.org/trust-principles-rda-community-effort

successful tool development initiatives, the Data Champions funding opportunity, and collaborations.

Previously, the following types of training products have been supported by the Alliance. These types of tool development initiatives will continue:

- Guides and Primers covering a range of RDM-related topics,
- Videos on topics such as creating Institutional RDM Strategies, Data Management Planning, FRDR, and Borealis,
- Webinars, both from the Alliance and in partnership with other DRI organizations,
- Symposia and Workshops, and
- Expansion of innovative tools such as the successful 'Workshop-in-a-Box' toolkit developed to help local RDM specialists more easily develop and launch RDM training.

To help advance awareness, understanding, development, and adoption of research data management tools, best practices, and resources in Canada, a national Data Champions Pilot[21] project was designed and implemented, based on models that have been instituted globally.[22] While some of the efforts of Data Champions are focused locally (with the potential to be scaled nationally), one of the goals of the Pilot is to support development and sharing of specialized, discipline-specific training and/or other RDM-aligned resources across Canada.

The Alliance RDM Data Champions Pilot, launched in January 2022, funds 18 experts and their teams to take on the role of 'Data Champions' at their post-secondary institutions, research institutions, or nonprofit organizations, and within their disciplinary or interdisciplinary contexts. The Data Champions Pilot provides essential funding and specialized support through a national Community of Practice model connecting Data Champions with each other and the RDM Network of Experts.

Broadly speaking, this Pilot promotes sound RDM practices and contributes to a culture change in the conduct of digital research, to ensuring Canadian research data are FAIR, and ultimately to fostering excellence in Canadian research. As a tangential benefit, the inaugural Data Champions cohort will provide valuable feedback on how the Alliance can sustain, improve, and grow the Data Champions model, and the associated Community of Practice initiated by this Pilot, into a Data Champions Program.

In addition, the Alliance and its National Training Expert Group (a group within the RDM Network of Experts), are continuing to develop rich training focused-initiatives, while seeking connections with ARC, RS, and Cybersecurity colleagues, as well as other DRI stakeholders, to

---

[21] Data Champions - https://alliancecan.ca/en/funding-opportunities/data-champions-pilot-project-call

[22] See Savage, JL and Cadwallader, L. 2019. Establishing, Developing, and Sustaining a Community of Data Champions. Data Science Journal, 18: 23, pp. 1–8. DOI: https://doi.org/10.5334/dsj-2019-023

collaborate on training initiatives that support all key elements of DRI for audiences across Canada's research and research support communities.

CoreTrustSeal Repository Certification Pilot

Data repository certification is increasingly recognized as an important next step in the maturity of repository services that enables compliance with the FAIR and TRUST principles. Consequently, the national Data Repositories Expert Group (DREG), supported by the Alliance RDM team, launched a Pilot program aimed at increasing the awareness and adoption of CoreTrustSeal Certification in Canada. CoreTrustSeal is an international, community-based, non-governmental, and non-profit organization promoting sustainable and trustworthy data infrastructures and aims to provide a core level of certification. This core certification may serve as a first step in achieving extended levels of certification (i.e., nestor-Seal DIN 31644 and ISO 16363).[23] The *CoreTrustSeal Certification Canadian Support Cohort and Funding Initiative* Pilot was established in the Spring of 2021, with members participating in workshops and other collaborative activities to help them achieve, or make progress toward achieving, CoreTrustSeal certification for their repository. Active support continues for the Cohort members through this Pilot initiative to increase capacity building for certification in Canadian repositories. This program has been welcomed by the community and we are exploring possibilities for similarly framed future initiatives.

Curation Initiatives

The Alliance follows and contributes to best practices and standards to ensure data, metadata, and code are FAIR. FRDR, for instance, operates under a curated deposit model, with all deposits reviewed by curators before publication to ensure data are independently understandable and usable by others. Ongoing work with the RDM community through the Network of Experts is advancing policy recommendations that encourage data review and other curation activities at the institutional level, particularly for institutions that use the Borealis repository platform.

The national Curation Expert Group, with support from the Curation Coordinator, advises on and develops curation initiatives, identifying and deploying best practices and tools designed to increase service capacity nationally, institutionally, and within disciplinary communities.[24] Three FRDR Curation Officers, working closely with the Curation Coordinator and FRDR Service Team, provide ongoing curation support to Canadian researchers depositing data into FRDR as well as participating in Expert Group activities. One of the initiatives includes the development of

---

[23] CoreTrustSeal - https://www.coretrustseal.org/about/

[24] Conceptualizing a National Approach to Data Curation Services in Canada. https://doi.org/10.5281/zenodo.3894934; CEG Roadmap (External)

a community-driven, national network of data curation experts. National level data stewardship support is essential to providing pan-Canadian coordination of these efforts, and creates opportunities for knowledge sharing and alignment of service delivery efforts at all levels across Canada.

Preservation Initiatives

From 2018 to the present, Alliance RDM and its predecessors have worked with institutional and community partners to establish and coordinate digital preservation activities based on best practices and international standards. This work is guided by FRDR policy and disciplinary best practices, as the FRDR Preservation Coordinator works to preserve data published in the repository with assistance from the larger Service Team.

FRDR's Preservation Policy highlights a key aspect of the repository's preservation strategy as "the development of a distributed digital preservation network, which focuses on building capacity in Canadian memory institutions and preservation partners," noting that "the long-term preservation of Canadian research data requires broad community support and participation, which is encouraged and maintained by the Alliance."[25]

To this end, FRDR has participated in the development of a pilot Preservation Service Provider (PSP) network in collaboration with the national Preservation Expert Group, SciNet / University of Toronto, and Internet Archive Canada / University of Victoria. While SciNet storage is currently operational, governance structures and policies supporting wider community participation are still under development.

Additional Preservation Expert Group initiatives supported by the Preservation Coordinator include the development and publication of *Appraisal Guidance for the Preservation of Research Data* ([EN](#) / [FR](#)), a complimentary Appraisal Checklist and Workshop, and the recently convened Research Software Preservation Working Group.

Sensitive Data Initiatives

RDM initiatives also align with the Federal Government's 'Roadmap for Open Science'[26] call for transparency in scientific research outputs -- making data "open by design and by default." The Alliance is addressing issues of stewarding sensitive data in a number of ways through the work of the Sensitive Data Expert Group and facilitated by members of the Alliance RDM team. The Sensitive Data Expert Group developed 'deposit-friendly' language for ethics approval and

---

[25] [Digital Research Alliance of Canada. (2021). Preservation Policy. The Federated Research Data Repository.](#)

[26] Roadmap for Open Science, 2020 - https://www.ic.gc.ca/eic/site/063.nsf/vwapj/Roadmap-for-Open-Science.pdf/$file/Roadmap-for-Open-Science.pdf

informed consent,[27] a glossary of sensitive data terminology,[28] and a research data risk matrix [29] to help researchers and ethics boards navigate challenges surrounding the deposit of sensitive data. A Policy, Privacy, and Sensitive Data Coordinator was hired to facilitate these and other efforts to advance principles and practices that will allow research data to be as open as possible while respecting ethics, policy, and legal requirements that often vary across jurisdictions.

Another important initiative is a Sensitive Data Repository Project, in partnership with Simon Fraser University and the FRDR team, and continued development of a range of researcher-facing resources and training.

Following the release of the Tri-Agency Research Data Management Policy, which institutes both DMP and data deposit requirements for researchers, and the release of Guidance on Depositing Existing Data in Public Repositories[30] by the Tri-Agency Panel on Research Ethics, which outlines the responsibilities of researchers and Research Ethics Boards in the context of sensitive data deposit, the Sensitive Data Expert Group has undertaken the development of resources and guidance on the implementation of these policies within sensitive data contexts.

The Sensitive Data Repository Project has three key components: 1) a lightweight, easy-to-use zero-knowledge encryption application for researchers to encrypt, decrypt, and share access to sensitive datasets, designed to be platform-agnostic and adaptable for use by FRDR and other platforms; 2) a national policy framework developed collaboratively with participating institutions from jurisdictions across Canada; and 3) service delivery supports (including resources, training, and services) to facilitate sensitive research data management workflows within institutions and repositories.

The harmonized policy framework for sensitive research data management will be co-developed by institutions, stakeholders within the research community, and the Alliance. Such a framework will facilitate further innovations in sensitive research data management by creating a common foundation upon which new projects, services, resources, and guidance to support sensitive data management may be built and supported.

---

[27] Deposit-friendly language - https://zenodo.org/record/4088959#.X43A3tDYo2w

[28] Glossary - https://zenodo.org/record/4088946#.X43BMtDYo2w

[29] Risk Matrix - https://zenodo.org/record/4088954#.X43BCtDYo2w

[30] Guidance on Depositing Existing Data in Public Repositories - https://ethics.gc.ca/eng/depositing_depots.html

Research Intelligence and Ecosystem Assessment

The Alliance continues to work with the broader DRI community to guide the development and direction of policies, practices, systems and infrastructure within RDM and the broader DRI ecosystem, which is evolving both nationally and internationally. The Research Intelligence Expert Group (RIEG) gathers, distills, and reports on the development of best practices in RDM in Canada and informs both the Alliance and stakeholder communities about existing and emerging issues in related policies and practices. The RIEG conducts independent studies and analyzes results to provide evidence-based recommendations regarding RDM to the Alliance. Outputs include but are not limited to Insight Reports, Institutional Capacity surveys, and an Institutional RDM Strategy survey, as a part of a broad suite of research intelligence and assessment activities.

This section has focused on the current RDM initiatives, but it bears reminding that this focus includes the five (5) Business Cases included in the MYFP. These Business Cases are available for review in the MYFP. The next section offers a look forward towards future visioning.

A Look Forward: Future RDM Initiatives

This future visioning section is a continuation of the evidence-based approach and leverages the analysis of the Research Data Framework (RDaF) as the Alliance continues to build upon the current RDM initiatives. This vision also capitalizes on the momentum set by the current MYFP Business Cases defined in the above sections of this document. Most significantly, it embeds a collaborative approach to future planning in that RDM was not considered as one element, but as part of the larger DRI (and Alliance) whole, alongside the needs of Advanced Research Computing and Research Software, and in alignment with Cybersecurity. This visioning approach to future RDM planning is designed to ensure the Alliance meets its ambitious mandate "to transform how research across all academic disciplines is organized, managed, stored and used."[31]

Considering RDM as one interconnected pillar within the Alliance and DRI ecosystem demonstrates common issues across the strategic reports generated by the Alliance. Key

---

[31] Digital Research Alliance of Canada. https://alliancecan.ca/en/about/alliance. Retrieved August 2022.

findings in both the ARC[32] and RDM[33] Current State Reports acknowledge that similar issues exist around systems: namely, that storage is insufficient, systems remain siloed from each other, and there is a lack of specific features, including interoperability and optimized storage management (i.e., moving content from one solution to another automatically). While work is underway to remedy this gap, the issue persists within the pillars. A clear recommendation in these Current State Reports is to specifically increase system capacities and capabilities in terms of host sites and storage solutions.

Given that systems and storage are both highlighted as concerns, our forward focus must include innovative solutions for these issues. As we consider the future, we begin to reframe these systems issues as interconnected solutions. One proposed future solution is the National Data Storage Grid model. This storage solution **"**separates compute and storage, which allows for the easy addition of storage without the need to augment compute power. This creates cost-effective scaling with the simple use of low-cost data storage,"[34] and provides the Alliance with one solution to issues identified within the DRI pillars. Not only does it provide necessary storage capacity with features that support Artificial Intelligence, Machine Learning, and mid-term storage capabilities to support researcher needs, it also includes features that specifically support RDM. For RDM, this solution will also support and facilitate interoperability of a range of current storage platforms and capabilities, such as institutional repositories, Borealis, and domain-specific repositories (it assumes current RDM initiatives will continue to be supported to facilitate the harvesting of available content from one portal).

As we continue to explore future innovations for RDM, three additional new initiatives build from integrating the pillars and expand on the National Data Storage Grid concept.

Preservation Service Provider Network

Expanding on the existing pilot, an investment in the Preservation Service Provider (PSP) Network will increase preservation capacity across Canada through the hiring of dedicated preservation personnel, the co-development and harmonization of shared preservation policies and platforms, and an expansion of distributed storage to ensure the longevity and accessibility of valuable Canadian research data over time. The PSP network project not only encourages knowledge and resource sharing among institutions as they develop complementary policies

---

[32] Alliance's Advanced Research Computing Working Group et al. (2021). Current State of Advanced Research Computing in Canada: An Update to the LCDRI Advanced Research Computing Position Paper. https://alliancecan.ca/sites/default/files/2022-04/2021-05_ARC_Current_State_Report_1.pdf

[33] Khair, S. et al. (2020). The Current State of Research Data Management in Canada: An Update to the LCDRI Data Management Position Paper. https://zenodo.org/record/6564659

[34] Oracle (2022). Retrieved from: https://www.oracle.com/ca-en/data-lakehouse/what-is-data-lakehouse/

and preservation workflows, but also provides an opportunity for repositories with less access to resources to participate in vital data preservation activities.

Since 2018, Alliance RDM, its predecessors, and stakeholders such as the national Preservation Expert Group have affirmed the importance of establishing a "distributed coordinated network of archives" modeled on the Open Archival Information System (OAIS).[35] A joint publication from the Preservation Expert Group and the Canadian Association of Research Libraries in 2020,[36] as well as an NDRIO Storage Working Group report from 2021,[37] reiterate the critical role of preservation in current and future research activities in Canada. The PSP network is directly informed by the extensive research and community consultation represented in these publications. Given the strong foundation developed during the pilot phase, existing community buy-in and participation, and potential opportunities for alignment with the proposed National Data Storage Grid, this initiative is well placed to move forward.

FRDR/Borealis Feature Extension

The FRDR/Borealis Feature Extension is a joint project between Borealis and the Federated Research Data Repository (FRDR) to broaden support for data collections in Canada. This initiative involves identifying and documenting large data collections in institutions across Canada that are not easily findable or usable to identify gaps in disciplinary repository options for various fields of research, and the prioritization of those collections and fields of research requiring the most urgent support. The Feature Extension will expand our existing national repository services to better support and manage collections, to better accommodate a variety of data types, and to better meet the needs of research groups that generate and use those data.

Core Domain Repositories

The Canadian data repository landscape is rich and diverse, with data repositories in operation at national, regional, and local/institutional levels and of varied categories (e.g., generalist and domain specific). However, domain-specific repositories are supported through a patchwork of grant-based funding, leading to potential funding gaps and sustainability issues for these important data infrastructure. This may lead to data loss, and the loss of innovative approaches to data re-use, in the long term. The Core Domain Repositories project supports strategy for

[35] Qasim, U. et al. (2018). Research Data Preservation in Canada : A White Paper. [White paper]. https://open.library.ubc.ca/soa/cIRcle/collections/ubccommunityandpartnerspublicati/52387/items/1.0371946 p. 2.

[36] The Portage Preservation Expert Group and The Canadian Association of Research Libraries Digital Preservation Working Group. (2020). *Digital preservation and NDRIO: a white paper* [White paper]. https://alliancecan.ca/sites/default/files/2022-03/ndrio-preservation-white-paper.pdf

[37] Sahrakorpi et al. (2021). NDRIO Storage Working Group: Overview and Recommendations for Research Data Storage in Canada. Final draft, unpublished.

bringing domain-specific repositories operating at regional and national levels into the Alliance's service ecosystem, with the aim of identifying consistent and sustainable support ongoing. This initiative builds directly on the National Data Storage Grid capacity.

These future-focused RDM initiatives present an innovative solution to a cross-DRI pillar issue and layer into each other, strengthening pan-national RDM supports and services. This future planning enables the Alliance to provide Canadian researchers with a "much more centralized, coherent, and predictable funding for the DRI ecosystem as a whole."[38]

Conclusion

This RDM Strategy will continue to evolve in the coming months, informed by continued application of the RDaF framework and ongoing consultations with the RDM and broader DRI communities. These efforts will help ensure this strategy aligns with and has the support of the RDM and DRI ecosystems, including meeting the varied needs of Canadian researchers and the professionals who expertly support them.

From its library-funded inception as CARL Portage to its government-funded transition and integration into the Alliance, the development and delivery of national RDM services has been strategic and deliberate. It has entailed both proactive and responsive initiatives, each informed by needs of researchers, institutions, and other stakeholders identified by a variety of means: the expertise of those serving researchers at the local level (e.g., RDM Network of Experts), the Alliance's Researcher Council and Needs Assessment processes, and many of the submissions from the RDM and library communities.

In the two-year period covered by this Strategy, the plan is to stay the strategic course, while scaling up to meet growing need and demand, appropriately integrating RDM best practices and tools into ARC workflows, and supporting development of a coherent data storage network. It is equally important that Alliance RDM have the expertise, credibility, and network reach to collaboratively shape and coordinate the growth and evolution of standards, practices, expert capacity, and infrastructure of the entire Canadian research data ecosystem. National leadership and coordination remains absolutely vital.

At this time, the requests for new/additional RDM funding in the Multi Year Funding Proposal target necessary growth and enhancement initiatives, building on a strong foundation of well-established RDM supports and services. As has been done previously, we will work to better identify, define, and prioritize initiatives that will best serve data-intensive disciplines and the

---

[38] Digital Research Alliance of Canada. https://alliancecan.ca/en/about/alliance . Retrieved July 2022.

longer-term data storage needs of researchers. It is anticipated that Alliance support for national RDM initiatives, including those described in this Strategy, will continue to evolve and grow.

Acknowledgements