

Conceptualizing a National Approach to Data Curation Services in Canada

Version 1.0

June 2020

Portage Network
Canadian Association of Research Libraries
portage@carl-abrc.ca

www.carl-abrc.ca

portage

SERVICES PARTAGÉS POUR LES DONNÉES DE RECHERCHE
SHARED STEWARDSHIP OF RESEARCH DATA

CARL ABRC
CANADIAN ASSOCIATION OF RESEARCH LIBRARIES
ASSOCIATION DES BIBLIOTHÈQUES DE RECHERCHE DU CANADA

Authors

This report synthesizes thoughts and ideas expressed by participants of the October 2019 Canadian Data Curation Community-Building Forum. Report authors include:

Erin Clary

Portage Network
erin.clary@carl-abrc.ca

Jason Brodeur

McMaster University
brodeujj@mcmaster.ca

Lee Wilson

Portage Network
lee.wilson@ace-net.ca

Jeff Moon

Portage Network
moonj@queensu.ca

Shahira Khair

University of Victoria
skhair@uvic.ca

Drafts were reviewed and revised by:

Jennifer Abel

Portage Network

Ève Paquette-Bigras

Université de Montréal

Corey Davis

University of Victoria

Laure Perrier

University of Toronto

Reyna Jenkyns

Ocean Networks Canada

Kevin Read

University of Saskatchewan

Lisa Johnston

University of Minnesota

Felicity Tayler

University of Ottawa

Caitilín O'Hare

Portage Network

Minglu Wang

York University

Attendees of the Canadian Data Curation Forum, and members of the broader Canadian research data management community.

Acknowledgements

Funding and support for the Canadian Data Curation Forum was provided by the Canadian Association of Research Libraries' (CARL) Portage Network, McMaster University Library, and the Social Sciences and Humanities Research Council (SSHRC). We would also like to acknowledge the Portage Curation Expert Group, and the Program and Local Arrangement Committees who made the Canadian Data Curation Forum possible.

SSHRC  CRSH



Library



Table of Contents

AUTHORS	2
ACKNOWLEDGEMENTS	3
EXECUTIVE SUMMARY	5
INTRODUCTION	7
DATA CURATION IN THE CONTEXT OF RESEARCH DATA MANAGEMENT	9
SUPPORTING ROBUST DATA CURATION AT THE NATIONAL SCALE	10
SUMMARY OF CANADIAN DATA CURATION FORUM	13
DATA CURATION TRAINING EVENT	13
COMMUNITY-BUILDING FORUM.....	15
COMMUNITY-BUILDING FORUM OUTCOMES	16
BARRIERS AND CHALLENGES.....	16
CONCEPTUALIZING A NATIONAL APPROACH TO CURATION SERVICES.....	18
RECOMMENDED ACTIVITIES.....	18
NEXT STEPS AND DISCUSSION	23
ROADMAP	23
FURTHER CONSIDERATIONS.....	26
SUMMARY OF RECOMMENDATIONS FOR NDRIO	28
BIBLIOGRAPHY	30
APPENDIX	34
LIST OF ACRONYMS	34
CDCF COMMUNITY-BUILDING FORUM PARTICIPANT BREAKDOWN.....	35
OVERVIEW OF NATIONAL DATA REPOSITORIES PRESENTED AT THE CDCF	36

Executive Summary

Data curation is the active management of research data as it is created, maintained, used, archived, shared, and reused. It is an iterative process that adds value to scholarship by optimizing datasets for current use, as well as future discovery and reuse, and ensures that datasets developed and deposited by researchers epitomize the FAIR guiding principles. Data curation tasks require a varied skill set, including disciplinary knowledge, familiarity with the research process, knowledge of metadata standards and best practices in data management, and comfort working in a technology-intensive environment. Given the complexities inherent in data curation, there are a variety of gaps in knowledge, skills, tools, and coordination that may limit the value returned on research data management activities. There is a need for training opportunities that expand the capabilities of data curators, and for high-level coordination of data curation services across organizations.

This report summarizes the outcomes of the Canadian Data Curation Forum, a national, three-day event held in October of 2019 to identify perceived barriers and challenges to curation work, and to articulate how a national approach to data curation in Canada could address these challenges. Forum participants included researchers, data and information professionals, officials, and administrators representing a wide range of research institutions and organizations, funding agencies, and support organizations from across Canada. Participants identified the following barriers to curation work:

- Limited capacity for curation given competing priorities
- No formal research data management service, or no dedicated staff to assist data producers with curation
- Lack of messaging to communicate the value of curation to various stakeholders
- A need for shared standards, workflows, best practices, and educational opportunities
- A need for clarity in roles and responsibilities, and evidence of what services would most benefit researchers and data producers.

To address these concerns, participants envisioned a nationally coordinated approach to data curation that would leverage the work of existing organizations, including the emerging New Digital Research Infrastructure Organization. Next steps for the development of a national curation service are outlined in the report (see Roadmap, page 23), and include engagement with researchers, data producers, and other stakeholders to ensure the development of services that are relevant to their needs; investment in infrastructure, including the ‘human layer’ required for curation; strengthening a curation community of practice; and support for services and resources that will

facilitate growth in curation capacity at the local level. To lay the groundwork for a broader national curation strategy, the following actions are recommended:

- Acknowledge the importance of data curation in enabling FAIR data products and open scholarship, and the role of data curators in this process.
- Acknowledge that data curation is a complex, labour intensive activity that requires significant investment in, and development of highly-skilled staff.
- Actively engage with Canadian researchers and data producers to better understand their curation needs and the challenges they face, and work with these communities to ensure that the curation services we develop will support them.
- Contribute to the creation of a national framework for curation services that supports curation needs broadly and addresses the variety of needs of diverse organizations.
- Support the development and provision of training opportunities, learning resources, and guidance materials in both official languages to support the development of data curators across Canada.
- Collaborate with Indigenous communities to ensure the curation services and resources we develop work in harmony with their data stewardship efforts.
- Support the establishment and maintenance of a national curation community of practice.
- Collaborate with the curation, digital preservation, and ethics communities to address complex issues related to the curation and long-term management of biomedical and other health data, legacy and rescued data, multilingual data, and data generated outside of academic institutions.
- Collaborate with digital preservation experts to identify data management and curation actions that will support long-term preservation.
- Develop a mechanism for the curation community to inform governance and policy decisions related to research data management.
- Directly or indirectly, support the creation of curation positions necessary to ensure excellence in datasets deposited into Canadian data repositories.

Introduction

Research data are primary sources of information used to support and validate research results. Data may be generated or collected by observation, experiment, or simulation, or compiled, derived, or processed from existing data.¹ Many data, captured at a moment in time, are irreplaceable, and other data may be cost-prohibitive to recreate. Data sharing catalyzes innovation and discovery by permitting products of research to be reused or repurposed to answer novel research questions, and it enables repetition and verification of analyses, thus helping to address issues related to research transparency and reproducibility.²

Governments, funders, and scholarly communities recognize the value of data sharing as a vehicle for scientific advancement and international collaboration, which has the potential to drive economic growth and enhance social welfare,³ and an increasing number of funding agencies and publishers have adopted policies that require researchers to share data on request, or to archive data in a dedicated repository.⁴ The Government of Canada's *Roadmap for Open Science* recommends that outputs of newly funded scientific research should be "Open by Design and by Default" whenever possible,⁵ but also recognizes that simply sharing data is not enough to realize the full potential of open science. For that, data must be properly managed throughout their lifecycle,⁶ and adhere to the FAIR guiding principles of Findability, Accessibility, Interoperability,

¹ CASRAI, s.v. "Research Data," accessed February 2, 2020, <https://casrai.org/research-data/>.

² Open Science Collaboration, "Estimating the reproducibility of psychological science," *Science* 349, no. 6251 (August 28, 2015): aac4716, <https://doi.org/10.1126/science.aac4716>.

³ Organisation for Economic Co-operation and Development, "OECD principles and guidelines for access to research data from public funding," 2007, <https://doi.org/10.1787/9789264034020-en-fr>; CODATA, and Paul Uhlir, "The Value of Open Data Sharing: A CODATA Report for the Group on Earth Observations," Zenodo, 2015, <http://doi.org/10.5281/zenodo.33830>; Government of Canada, Innovation, Science and Economic Development Canada, Office of the Deputy Minister, Communications and Marketing Branch, "Tri-Agency statement of principles on digital data management," 2016, https://www.science.gc.ca/eic/site/063.nsf/eng/h_83F7624E.html.

⁴ See for example, 130 policies archived at FAIRsharing, <https://fairsharing.org/policies/>.

⁵ Government of Canada, Office of the Chief Science Advisor or Canada, "Roadmap for Open Science," 2020, https://www.ic.gc.ca/eic/site/063.nsf/eng/h_97992.html.

⁶ For a description of the data lifecycle, see for example, DataONE, "Data Life Cycle," <https://www.dataone.org/data-life-cycle>; Leadership Council for Digital Research Infrastructure. "Data Management Position Paper For Innovation, Science, and Economic Development Canada," 2017, fig. 2, p. 6, <https://wiki.queensu.ca/display/QDSRWG>.

and Reusability.⁷ The importance of sound data management practices for maximizing research value are widely acknowledged,⁸ and data management is included as one of the three key responsibilities of Canada’s New Digital Research Infrastructure Organization (NDRIO).⁹ In 2020, many researchers will be required to include data management plans (DMPs) as part of their grant proposals, and institutions administering Tri-Agency funds will be required to develop a research data management (RDM) strategy to describe how they will support their researchers’ data management efforts.¹⁰

To fully support researchers’ efforts to manage data throughout a project’s lifecycle, and to share them at a project’s conclusion, there are a number of barriers and challenges that must be addressed. A range of upstream decisions and actions can impact the overall utility and value of data, and may determine if and how data can be shared. A lack of time and resources to properly describe and organize data for archiving,¹¹ uncertainty about where and how to share data, concerns for the ethical reuse of their data, and a lack of clarity around concepts such as data ownership,¹² copyright, and licensing¹³ are all documented concerns that may limit data sharing. If data are shared, the potential for reuse is dependent on data integrity, demonstrated in part by the quality and completeness of the data and associated documentation, but this can be time-consuming and difficult to produce, and there are currently limited incentives to do so.¹⁴ Given that

⁷ Wilkinson et al., “The FAIR Guiding Principles for scientific data management and stewardship,” *Scientific Data* 3, 2016, <https://doi.org/10.1038/sdata.2016.18>.

⁸ Government of Canada, “Roadmap for Open Science”; Government of Canada, “Tri-Agency statement of principles on digital data management”; Baker et al., “Research data management in Canada”; Research Data Strategy Working Group, “The 2011 Canadian Research Data Summit: Mapping the data landscape,” 2011, http://longrangeplan.ca/wp-content/uploads/2010/10/Data_Summit_Report.pdf.

⁹ Government of Canada, Innovation, Science and Economic Development Canada, Office of the Deputy Minister, “Digital Research Infrastructure”, 2019, <https://www.ic.gc.ca/eic/site/136.nsf/eng/home>.

¹⁰ Government of Canada, Innovation, Science and Economic Development Canada, Office of the Deputy Minister, Communications and Marketing Branch. “DRAFT Tri-Agency research data management policy for consultation,” 2018, https://www.science.gc.ca/eic/site/063.nsf/eng/h_97610.html.

¹¹ Perrier, Laure, Blondal, Erik, and Heather MacDonald, “The views, perspectives, and experiences of academic researchers with data sharing and reuse: A meta-synthesis,” *PLoS ONE* 15, no. 2, 2020, <https://doi.org/10.1371/journal.pone.0229182>; Stuart et al., “Whitepaper: Practical challenges for researchers in data sharing (Version 1),” Figshare, 2018, <https://doi.org/10.6084/m9.figshare.5975011.v1>.

¹² Perrier, Blondal, and MacDonald, “The views, perspectives, and experiences of academic researchers.”

¹³ Stuart et al., “Whitepaper: Practical challenges for researchers.”

¹⁴ Perrier, Blondal, and MacDonald, “The views, perspectives, and experiences of academic researchers.”

data sharing is a relatively new requirement, researchers also cite the need for training, and resources and services that will support their data sharing efforts.¹⁵

Data curation in the context of research data management

Data curation is an important component of RDM, and involves iterative review and revision of data and metadata to improve their FAIRness -- optimizing them for both current use and future discovery and reuse.¹⁶ While some services and software offer functionality that streamlines or partially automates selected curation activities (e.g. The Scholars Portal [Dataverse Data Curation Tool](#) and [BitCurator](#)'s file format and personal identifier detection), the heterogeneous and multidisciplinary nature of research data typically requires intervention of human curators to assess data and metadata quality, define appropriate FAIR standards, and collaborate with researchers to augment submissions. Data curation does not happen at a single point, and data curators assume a variety of roles and interact with data in a range of capacities.¹⁷ Researchers, data managers, data stewards, academic librarians, and other data practitioners may all curate elements of a dataset as it flows through the research data lifecycle. Relevant data curation activities include:

- Preparing strategies for data collection and documentation;
- Creating documentation and metadata to explain and contextualize data;
- Implementing a secure storage plan, version control for files and code, and protective measures such as file backups and file integrity checks;
- Performing quality assurance during data analysis through metadata inspection, file audit, and code review;
- Preparing datasets for deposit in a data repository or archive, which includes applying appropriate anonymization or access controls;
- Transforming file formats to enhance reuse and long-term preservation; and
- Augmenting metadata and linking datasets to increase discoverability.¹⁸

¹⁵ Perrier, Blondal, and MacDonald, "The views, perspectives, and experiences of academic researchers."

¹⁶ Portage Curation Expert Group. "Primer: Data Curation," Ottawa: Portage Network, 2019, https://portagenetwork.ca/wp-content/uploads/2019/09/Curation_Primer_Aug2019_EN.pdf.

¹⁷ An extensive list of data curation activities and definitions can be found in the appendix of Johnston et al., "How Important Are Data Curation Activities to Researchers? Gaps and opportunities for academic libraries," *Journal of Librarianship and Scholarly Communication* 6, no. 1, 2018, <https://doi.org/10.7710/2162-3309.2198>, or on the Data Curation Network website <https://datacurationnetwork.org/data-curation-activities/>.

¹⁸ Portage Curation Expert Group, "Primer: Data Curation."

Considered in its entirety, data curation will benefit a range of communities and stakeholder groups:

Table 1: Stakeholder groups and potential benefits as a result of data curation

Stakeholder group	Potential benefits as a result of data curation ¹⁹
Funders, government, public	Improved return on investments when data can be found, accessed, understood, and repurposed to answer new research questions.
Institutions	Improved researcher compliance with funder and government mandates. Protection of reputation associated with quality assurance and greater research transparency.
Researchers and research communities	Improved transparency in the scientific process. Improved efficiencies arising from proper data management. Enhanced data quality that encourages collaborative and interdisciplinary research initiatives. Citations for published datasets.
Infrastructure operators (e.g., repository operators)	Higher quality products which are better prepared for long-term preservation. Application of standards to increase discoverability and interoperability. Increased likelihood of meeting repository certification requirements such as the CoreTrustSeal. ²⁰

Supporting robust data curation at the national scale

The emergent data curation profession demands a varied set of skills and expertise, and the roles and responsibilities of a curator may not be consistent across institutions. Data repositories may hire full time curators to work with researchers at the end of a research project to prepare datasets for deposit and long-term preservation storage, while research data management librarians and data stewards may find that aspects of data curation are one of many research adjacent tasks they

¹⁹ Portage Curation Expert Group, “Primer: Data Curation”; Digital Curation Centre, “Digital curation 101: Incentives for digital curation,” Edinburgh: Digital Curation Centre, 2011, <http://www.dcc.ac.uk/sites/default/files/documents/DC%20101%20Incentives%20for%20Curation.pdf>.

²⁰ CoreTrustSeal Standards and Certification Board, “CoreTrustSeal Trustworthy Data Repositories Requirements 2020–2022,” Zenodo, 2019, <http://doi.org/10.5281/zenodo.3638211>.

are asked to engage in throughout the research project lifecycle.²¹ Researchers themselves undertake many curation tasks as they manage their own data,²² although they likely do not label themselves as data curators.

To support researchers' data management and data sharing efforts, organizations often seek curators who are familiar with the research process and research methodologies, and may require research experience, or disciplinary expertise in the form of an advanced degree.²³ Disciplinary knowledge and an understanding of the research process may be especially important for complex data types, for example in the management of sensitive data, where curators must have an understanding of the role of ethics in data management, disclosure risk, informed consent, and a knowledge of procedures for de-identification. In the Canadian context, respect for, and an understanding of the OCAP© principles²⁴ and other information governance initiatives in Indigenous communities is also important. Curators must have knowledge of metadata standards, as well as other existing and emerging standards, specifications, trends, and best practices in data management and scholarly communications. They must be informed about various licenses, copyright law, funder policies, and institutional policies.²⁵ Curators must be comfortable in a technology-intensive environment, and they may be required to possess knowledge of various operating systems, scripting and programming languages, and software and tools that support curation decisions and actions, preservation, data analysis, data visualization, and data transformation.²⁶ Curators may also be asked to engage in outreach, provide technical support for various tools and softwares, advocate for best practices in data management, and help develop data management policies and workflows at their organizations.²⁷

²¹ Kim, Jeonghyun, Edward Warga, and William Moen, "Competencies Required for Digital Curation: An Analysis of Job Advertisements," *The International Journal of Digital Curation* 8, no. 1, 2013, <https://doi.org/10.2218/ijdc.v8i1.242>; Tamaro, Anna Maria, Krystyna K. Matusiak, Frank Andreas Sposito, and Vittore Casarosa, "Data Curator's Roles and Responsibilities: An International Perspective," *Libri* 69, no. 2, 2019, <https://doi.org/10.1515/libri-2018-0090>; Mons, Barend, *Data Stewardship for Open Science: Implementing FAIR Principles* (New York: CRC Press, 2018), p.28, <https://doi.org/10.1201/9781315380711>.

²² Johnston et al., "How Important Are Data Curation Activities to Researchers?"

²³ Tamaro, Matusiak, Sposito, and Casarosa, "Data Curator's Roles and Responsibilities."

²⁴ First Nations Information Governance Centre, "Ownership, Control, Access and Possession (OCAP™): The Path to First Nations Information Governance," Ottawa: The First Nations Information Governance Centre, 2014, https://fnigc.ca/sites/default/files/docs/ocap_path_to_fn_information_governance_en_final.pdf.

²⁵ Kim, Warga, and Moen, "Competencies Required for Digital Curation."

²⁶ Kim, Warga, and Moen, "Competencies Required for Digital Curation."

²⁷ Kim, Warga, and Moen, "Competencies Required for Digital Curation"; Tamaro, Matusiak, Sposito, and Casarosa, "Data Curator's Roles and Responsibilities."

Given the complexities inherent in the profession, there are a variety of gaps in data curation knowledge, skills, tools, and coordination that may limit the value returned on research data management activities.²⁸ Curators within individual research groups, organizations, and institutions are unlikely to possess all of the skills and expertise required to manage a diverse body of data throughout its lifecycle.²⁹ There is a significant need for training opportunities that expand the capabilities of data curators, and for high-level coordination of data curation development and practices across organizations. Federated curation networks in the United States and the United Kingdom (the Data Curation Network³⁰ and the Digital Curation Centre,³¹ respectively) provide examples of how services, resources, and training can be coordinated at national scales.

It is crucial to address these gaps because there is extant and growing demand for curation services. Repository certification organizations such as CoreTrustSeal require a demonstrated level of data curation.³² These requirements, and other published standards for trustworthy data repositories, such as ISO 16363,³³ will inform publishers and funders as they establish criteria by which to recommend data repositories. A recently drafted set of criteria released by the United States Office of Science and Technology Policy,³⁴ and another set of criteria published by FAIRsharing, DataCite, and a group of publisher representatives³⁵ both list curation as a desirable repository function that will be required as a minimum best practice in the future. Likewise, the *Tri-Agency Statement of Principles on Digital Data Management* asserts that research data should be preserved in an open, curated repository when possible,³⁶ and the *Roadmap for Open Science* recognizes curation as one of the data management activities required to implement FAIR principles.³⁷ Although curation is not mentioned directly, Innovation, Science and Economic

²⁸ Johnston et al., “How Important Are Data Curation Activities to Researchers?”

²⁹ Johnston et al., “Data Curation Network: A Cross-Institutional Staffing Model for Curating Research Data,” *The International Journal of Digital Curation* 13, no. 1, 2018, <https://doi.org/10.2218/ijdc.v13i1.616>.

³⁰ Data Curation Network <https://datacurationnetwork.org/>

³¹ Digital Curation Centre <http://www.dcc.ac.uk/>

³² CoreTrustSeal Standards and Certification Board, “CoreTrustSeal Requirements”

³³ International Organization for Standardization (ISO), “ISO 16363:2012 Audit and Certification of Trustworthy Digital Repositories”, 2012.

³⁴ United States, Office of Science and Technology Policy, “Request for Public Comment on Draft Desirable Characteristics of Repositories for Managing and Sharing Data Resulting From Federally Funded Research,” Federal Register, Doc. 2020-00689, 85 FR 3085, 2020.

³⁵ McQuilton et al, “FAIRsharing Collaboration with DataCite and Publishers: Data Repository Selection, Criteria That Matter,” Open Science Framework, 2019, <https://doi.org/10.17605/OSF.IO/N9QJ7>.

³⁶ Canada, “Tri-Agency statement of principles on digital data management.”

³⁷ Government of Canada, “Roadmap for Open Science”

Development Canada (ISED) recognizes that Canada’s new digital research infrastructure (DRI) will require highly qualified personnel (HQP) with the expertise to support researchers,³⁸ and specific mention of the importance of curation to the DRI re-organization is provided by a number of position³⁹ and background⁴⁰ papers that have informed the process.

In order to better understand the current state and unmet needs of data curation practice in Canada, and to articulate how a national approach to data curation in Canada might address these challenges, the Canadian Association of Research Libraries’ (CARL) Portage Network, in collaboration with McMaster University Library, and the Social Sciences and Humanities Research Council (SSHRC), hosted the Canadian Data Curation Forum at McMaster University in Hamilton from October 16 to 18, 2019. This report summarizes the event, describes outcomes of the Forum, and proposes next steps for expanding data curation services and support in Canada.

Summary of Canadian Data Curation Forum

The Canadian Data Curation Forum (CDCF) brought together representatives from a range of institutions and organizations⁴¹ with the goal of fostering a national community of practice to catalyze the development and adoption of data curation standards, practices, tools, and skills across disciplines and institutions. The event was held in two parts. For the first one-and-a-half days, attendees at the Data Curation Training Event participated in workshops focused on knowledge sharing and hands-on skill-building. This event was followed by a Community-Building Forum, designed to develop an understanding of Canada’s current curation capacity and challenges. Attendees of the Community-Building Forum were active participants in modelling a vision for a national approach to data curation in Canada that would address these challenges.

Data Curation Training Event

The Data Curation Training Event at CDCF was a professional development opportunity that allowed attendees to network and build their data management and curation skills through a

³⁸ Canada, “Digital Research Infrastructure.”

³⁹ Leadership Council for Digital Research Infrastructure, “Data Management Position Paper”; Castle et al., “Data Management Roadmap – 2019-2024: For Submission to Innovation, Science and Economic Development (ISED),” 2019, <https://www.rdc-drc.ca/download/position-paper-data-management-roadmap-2019-2024/?wpdmdl=2482>.

⁴⁰ Baker et al., “Research data management in Canada.”

⁴¹ See Appendix B. for further breakdown of Community-Building Forum participants by organization type, geographic region, and position.

variety of workshops. Each participant attended 3 of the 6 concurrent half-day sessions that were offered. Workshops shared best practices and important resources, and allowed practitioners to develop knowledge and experience through hands-on training activities.

Table 2: Workshops and Description

Workshop	No.	Description
Introduction to R for Social Scientists	1a	Introduces the R syntax and the RStudio interface, importing data, data frames, factors, adding/removing rows and columns, calculating summary statistics, and plotting.
Making Ethical and Effective Data Management Work for You!	1b	Introduces RDM supports and services, such as Portage DMP Assistant tool, and discusses best practices in data management to safely and ethically manage data throughout their lifecycle.
Data Wrangling and Versioning with OpenRefine and GitHub	2a	Introduces GitHub as a tool to acquire, document, and version data and code, and OpenRefine, a tool to explore, clean, and transform data for interoperability.
Curating Data for Reproducibility	2b	Introduces best practices for reproducible research, the challenges of creating and curating reproducible research, and tools such as OSF, Jupyter Notebooks, and Binder that can be used to organize, document and share data, code, and computational environments.
The DCN CURATE Model	3a	Introduces a set of standardized curation steps and checklists that help ensure all datasets receive consistent treatment. Data curation primers are introduced, which detail the framework in practice for specific file types or disciplines.
Curating Data in Repositories	3b	Introduces the FAIR principles, and discusses current curation practices in domain and general repositories from across Canada that help improve the findability, accessibility, interoperability, and reproducibility of research data.

Forum materials, including slides and workshop materials, have been made available online at <https://data-curation.github.io/materials/>

Community-Building Forum

A one-and-a-half day Community-Building Forum followed the training event. Participants included researchers, data and information professionals, officials, and administrators representing a wide range of research institutions, non-governmental and research organizations, government and funding agencies from across Canada. The objectives of the Forum were to develop a collective understanding of the role of data curation within the larger data management and data stewardship landscape, to articulate the challenges of curation, to identify resources that would support best practices in data curation, and to specify ways in which national coordination of data curation efforts might support local curation needs. Participants attended three plenary sessions, and through a series of breakout sessions, were active participants in modelling a vision for a national approach to data curation in Canada.

- Jeff Moon, Director of the Portage Network, provided an overview of developments in the Canadian Digital Research Infrastructure (DRI) landscape and the state of national support for RDM, and articulated the important role of data curation in this ecosystem. The role of data curation in institutions was further discussed by panel members Jeff Moon; Alexander Clark, Associate Vice President Research, Faculty of Nursing, University of Alberta; and Mark Leggott, Executive Director of Research Data Canada (RDC). In the Q&A session that followed, panel members advanced the complementary role that domain, institutional, and national repositories each play in a national approach to data management, and acknowledged the importance of incorporating Indigenous initiatives and perspectives into the RDM landscape.
- Lisa Johnston and Cynthia Hudson-Vitale provided attendees with an overview of the development and growth of the US-based Data Curation Network (DCN), a coordinated cross-institutional network of data curators and data curation services. Network partners collaborate to create and share data curation procedures and best practices, and provide in-kind curation services. Datasets submitted to the network are matched to a curator with expertise in the file type or research discipline, and data are curated using a standardized set of steps (C-U-R-A-T-E-D).⁴² Recommendations are provided back to the local institution, which coordinates any necessary changes with the researcher. The DCN's origin, growth, and sustainability were particularly relevant to the visioning activities that followed, as they provided a tangible prospective model for establishing and evolving a data curation network in Canada.

⁴² C-U-R-A-T-E-D steps and checklists, <https://datacurationnetwork.org/resources/>.

- Representatives from a selection of national data repositories provided updates on their services and discussed the curation-related challenges they face in providing FAIR data through their platforms. The session featured presentations about both generalist [Federated Research Data Repository (FRDR), Scholars Portal Dataverse (SPDV)] and domain-specific [Canadian Integrated Ocean Observing System (CIOOS)] federated repositories that aim to integrate data across institutions, organizations, and research programs. The talks explored the important differences and similarities in data curation approaches and needs among Canadian data repositories. See [the Appendix](#) for an overview of repositories discussed.

Community-Building Forum outcomes

In breakout sessions, participant groups discussed the main barriers and challenges to data curation activities within their organizations and their work. Each group was asked to conceptualize a national approach to data curation that would address the identified challenges. Groups outlined the resources and support necessary to implement their models, and identified which components should be coordinated nationally, regionally, or locally. The outputs of this exercise are summarized below, alongside recommendations for a Canadian data curation network.

Barriers and challenges

Limited capacity

Limited capacity for curation work is a critical concern and was identified as a barrier by all groups. There is a scarcity of time due to competing priorities, and a lack of funding for curation-related work. Many organizations lack a formal RDM service or dedicated staff to perform curation work, and some participants worried that if demand for curation services were to increase at their local institutions, it would go unmet.

Need for training and guidance

The need for training and resources to support curation work is another pressing concern. Curators will encounter data from a diverse array of disciplines and sources, including research data, purchased assets, cultural resources, Indigenous specific data, biomedical and other health data, legacy and rescued data, data and metadata in French and English, and data generated outside of academic institutions. Participants do not yet feel prepared to handle the unique set of requirements and challenges presented by each data type. There is an urgent need for training and

other educational opportunities to build curation expertise, and participants cited the need for toolkits, workflows, standards, and policies to support curation work and inform best practices.

Intrinsic need for a multilingual/international approach

Canada is a multilingual nation, with many French, English, and Indigenous language speakers. Participants recognize the need for curation services, resources, and educational opportunities in multiple languages, and acknowledge that while it may be a challenge to support data management and curation needs in a multilingual nation, it is imperative that we do so.

Lack of clarity about existing services, roles, and expectations of rigour

Participants would like a clear picture of the current curation landscape. A better understanding of what services, resources, and opportunities exist, and how they can be leveraged at the local level is needed. Given limited resources at many institutions, some participants questioned how to prioritize datasets for curation, and wondered whether all datasets required the same level of curation. Others wondered who should assume responsibility for curation work, and how curation tasks should be divided. Although the answer will vary across institutions, organizations, archives, and repositories, and will likely be dependent on the number of available staff and their individual skill-sets, some clarification around the roles and responsibilities of the researcher or research team, and the responsibilities of the data steward or data curator may be useful.

Difficulty communicating value

Attendees also observed that the value of curation can be difficult to convey. Advocates lack clear messaging to communicate the value of curation to diverse stakeholder groups, including researchers, local administrators, and funding bodies. There are limited metrics to assess and demonstrate the impact of curation, which make it difficult to encourage buy-in given competing priorities for funds and staff time.

Engagement with researchers and other curators outside of our self-identified community of practice

Participants also recognize the need to build bridges between curators and research data management librarians at academic libraries, and curators and data stewards from domain repositories, data centres, and other organizations that manage data. Identifying various stakeholders who could contribute to a curation network and benefit from a curation network may be the first challenge. Researchers and other data producers are one stakeholder group that participants indicate must be actively engaged in data curation efforts. Their input will be

invaluable to developing and offering services and resources that will support their data management and data archiving needs.

Conceptualizing a national approach to curation services

To address these challenges, participants envisioned a nationally coordinated approach to data curation that would leverage the work of existing organizations and systems such as CARL Portage, Research Data Canada (RDC), First Nations Information Governance Centre (FNIGC), CIOOS, established regional consortia, and the emerging New Digital Research Infrastructure Organization (NDRIO).⁴³ A national curation network must represent a diversity of organizations, perspectives and voices. Its work should be community-driven and the end-goal should be to grow and enhance curation capacity at the regional and local level. Participants called for action as soon as possible and articulated that much of the groundwork can be laid now, even as a more formal curation network is being established. To that end, the first steps in modelling a national strategy that will empower individuals and institutions to build and grow curation capacity are proposed below. It is important to note that each of the proposed objectives inform and relate to the others, and they should be considered comprehensively.

Recommended Activities

Build curation capacity at the local level

- Invest in HQP to support research data management and data curation
- Create opportunities for training and skill-building
- Support professional development and education
- Develop and share standards, workflows, best practices, policies, and other guidance to inform practice at the local level

Limited capacity for curation work given competing priorities was cited as a critical concern. Curation services that meet the needs of researchers and other stakeholders will not be possible without dedicated HQP. To ensure the workforce is prepared to meet their needs, a national service should develop educational opportunities to re-skill existing HQP, and incentivize local institutions to invest in HQP. Professional development opportunities, and other opportunities to network and build hands-on curation skills are necessary in the short-term. Longer term, certificate programs and curricula enhancements could provide formal training opportunities for practicing data and information professionals, data stewards, new students, and early-career researchers, to

⁴³ <http://engagedri.ca/>

equip them with the tools and knowledge necessary to meet the challenges facing contemporary scholarship.

Curators will encounter a wide variety of data and digital resources, including purchased assets, cultural resources, Indigenous-specific data, health data and other data collected from human participants, legacy and rescued data, multilingual data and description, and data generated outside of academic institutions. Resources such as shared standards, workflows, and policies are also needed to guide curators and inform best practices at the local level. Although this body of resources could be managed at the national level, there should be mechanisms in place to allow curators and data professionals at the local level to participate in the development process and to provide feedback. Resources must be developed with input from a broad community of practice, to meet community needs. Curators at local institutions and organizations may use best practices and workflows to inform and standardize their work, but will maintain ultimate control over the decisions made on the datasets they ingest.

Foster a curation community of practice

- Create a framework that encourages individuals and institutions to network
- Establish informal and formal ways for curators to communicate with one another and participate in resource development
- Account for diversity in types of data, languages, organizations, and perspectives

Curators, research data managers, and data stewards would benefit from both formal and informal networking opportunities to advance curation practice. A national approach to curation will require a framework that enables collaboration, and a community of practice that reaches beyond traditional academic institutions to engage the diverse group of individuals and institutions working with data. Listservs, slack, virtual meet-ups, and other informal communication channels will allow any interested individual, repository, organization, or institution to connect, ask questions, share insights, and participate in the network. Working groups provide a more formal way for individuals and institutions to build relationships and collaborate to devise solutions to complex questions, and organized events, similar to the curation training event at the CDCF, will enable curators to network in person, and to share and learn practical data curation skills.

A national data curation network must incorporate principles of equity and inclusion at its core. There will be diversity in the variety of data that are managed, and there must be diversity in the perspectives and voices that contribute to policy and resource development. Traditional stakeholder communities include academic institutions, funders, administrators, researchers, and domain and institutional repositories; however, there are a wide range of communities and

organizations involved in curation, research, data management, or data stewardship that fall outside of the traditional academic research setting, including Indigenous communities, not-for-profit organizations, government ministries, public libraries, and other community groups. A national curation network should connect and engage these stakeholder communities so that newly developed curation policies and resources are representative of, and applicable across, a variety of institutions and organizations. These stakeholder groups must also be resourced in a manner that allows equal opportunity for contribution, and full participation in the network.

A national network should also explore ways to both leverage and contribute to work being done by international organizations such as the Research Data Alliance, the DCN, the Digital Curation Centre (DCC), the Australian Research Data Commons (ARDC), the National Coordination Point Research Data Management (LCRDM) curation task group, and the European Open Science Cloud (EOSC), for example. A professional association for data curators would give curators a unified voice to communicate with entities such as NDRIO, and at national RDM meetings, and could spearhead curricula development and shape opportunities to interface and network with international curation professionals and organizations. The development of multilingual resources and services at the national level also provides a unique opportunity to leverage our work at the international level, and to engage with international communities of practice.

Communicate the value of curation

- Establish metrics to measure the value added by curation
- Promote the value of curation and advocate for the profession
- Engage with researchers, data producers, and other stakeholder groups who will benefit from robust curation services

To build curation capacity and create buy-in at the local level, the value of curation must be clear. Buy-in from researchers, data producers, and other stakeholders who will benefit from curation services is necessary, and these stakeholder groups should be actively consulted as new services are developed, so the services are meaningful and serve their needs. Clear messaging will allow curators, information professionals, data stewards, and others to communicate the value of curation to various stakeholder groups at the local level, including researchers and other data providers, end-users, information technology departments and professionals, funders, administrators, and institutional Offices of Research. Data management professionals could recruit and engage champions from within these various stakeholder groups to help promote the value of curation more widely.

Metrics to measure the impact of curation should be evaluated and adopted, and an evidence base that demonstrates the importance of data curation in RDM should be established. The results of these efforts can be used to advocate for the profession, and to position curation as a first order concern of the data management component of NDRIO. In addition to metrics that measure the value added by curation, data usage metrics, impact studies, and success stories will all be important to encourage growth in curation capacity at local institutions.

Develop and maintain shared infrastructure

- Develop and maintain software and hardware essential to data curation and preservation
- Establish a framework within which a network of curators can communicate and operate
- Invest in the “human layer” required for effective data management
- Develop and coordinate curation services

At the national level, the continued development and maintenance of repository options that provide mechanisms for data curation, such as Dataverse and FRDR, is important. Resources should be devoted to shared platforms, tools, software, advanced computing environments, network connectivity and technical services such as DOI minting, all of which can be leveraged by institutions and organizations at the local level. National repository platforms should continue to seek funding for the development of automated tools to assist curators in their work, to create tools that will enable connections with digital preservation infrastructure and services, and to develop infrastructure necessary to support multilingual metadata for discovery.

Curators and data stewards must be able to apply and manage rules for data access. Investment in repository infrastructure that supports the ability to control or restrict access to data according to privacy laws (health data) or cultural norms (data with Indigenous ownership) will be required. If curation will happen across institutions and organizations, curators will also require a secure method to share and access data and metadata from distributed locations.

National coordination of best practices, standards, workflows, and policy development may also benefit local institutions. Although the development of these resources should be community-driven, and informed by the needs of various stakeholders at the local level, a national service could serve as a knowledge-base to host and manage these resources, and other curation software and tools. An example of this type of coordination is occurring within CIOOS, where regional

representatives determine metadata best practices and standards, which are consistent with the international oceanographic community.⁴⁴

It will be equally important to invest in the human component of data management. It is individuals who will build, connect, and maintain systems and networks, but institutions will have varying capacities to contribute, with many institutions currently lacking dedicated RDM staff. An initial effort to build local capacity, and create a flexible, scalable framework within which curators can operate and communicate is required. Participants were open to the idea of cross-institutional networking of data curators to share expertise, but there must be institutional buy-in, and a clear means of balancing local needs with the needs of the network before a more advanced model for shared staffing can be considered.

Represent curators and inform governance

- Establish and maintain a national curation advisory group that interfaces with the community of practice and national organizations on matters of data curation governance and support
- Collaborate with national funding agencies and digital research infrastructure organizations to develop curation policies and guidance in complex areas such as ethics, data rights management, and long-term preservation
- Develop funding models and infrastructure that ensure sustainability of the network

A connection to governance (via a committee or advisory council) should be established to provide strategic direction and scope on matters of curation, and position it as a first-order component of effective data management. An advisory council or similar entity will allow the curation community of practice to interface with NDRIO, and other data management organizations at the national and international level, with a unified voice. As with other research-supporting endeavours, the work of the advisory council should be community-driven, and its members should be representative of various stakeholders in the research ecosystem, including researchers, librarians, and data stewards from various domain repositories, national and institutional repositories, and Indigenous information governance centres.

Organizations will benefit from clear strategy and goals as they implement curation services. The advisory council could provide leadership around complex topics such as ethics, sensitive data, liability protections, copyright, licensing, complex data sharing agreements, and digital

⁴⁴ Stewart et al., "The Development of a Canadian Integrated Ocean Observing System (CIOOS)," *Frontiers in Marine Science* 6, 2019, <https://doi.org/10.3389/fmars.2019.00431>.

preservation, and help develop policy to inform local practice. This entity could help promote the value of curation, and model a curation network that is sustainable over the long-term.

Federal investment in a national data curation services will be necessary, but a sustainable network cannot rely solely on federal funds. Curation staff will be required to assist researchers who deposit data in national repositories such as FRDR, but resources must also be devoted to building curation capacity at local institutions. A skilled, distributed workforce could be networked to share expertise, and may mitigate against the loss of funds or personnel at the national level. Further work will be necessary to define the scope and strategic direction of a data curation network, and to model how a national data curation service could be sustained in the long-term.

Next steps and discussion

Roadmap

The following tables propose specific activities to be undertaken in the short-, medium-, and long-term to address each of the objectives described in the previous section. Note that responsible parties for each activity are not defined, but it is anticipated that they would be coordinated and carried out by components of NDRIO.

Objective 1: Build capacity for data curation, local and national

Short term (1Y)	Middle term (2Y)	Long term (5Y+)
Gap analysis to assess current curation capacity, existing resources, and researcher and other stakeholder needs	Develop content and host workshop(s) to provide further hands-on skill building and networking opportunities	Develop and expand formal educational opportunities, such as curricula enhancements for information professionals and researchers, and certificate programs for practicing professionals
Draft general curation workflows using best practice guidance and seek community feedback	Grow and populate a curation knowledge-base with community input	
Seek community input on what types of curation training and resources would be useful, and seek funding opportunities	Hire and coordinate HQP to curate for national repositories and support local curation needs	
	Explore advanced training options, such as train the trainer, to grow capacity exponentially	

Objective 2: Foster community of Practice

Short term (1Y)	Middle term (2Y)	Long term (5Y+)
<p>Identify ways that curators can connect informally (e.g., monthly virtual 'meet-ups', curator reference desk, listserv, Slack channel)</p> <p>Create opportunities for curators to collaborate, work, and learn in their first language</p> <p>Seek to identify data stewards, data managers, researchers, and others involved in curation work who may be operating outside of the library or repository, or outside of traditional academic research settings</p> <p>Establish connections with related communities of practices, such as the preservation community of practice</p>	<p>Create networking opportunities for curators, both virtually and in person</p> <p>Outreach to and engage with data stewards, data managers and others who are operating outside of traditional academic research settings</p> <p>Establish working groups, or ensure curator voice is represented in existing working groups, to address difficult curation questions in areas such as sensitive data, ethics and liability, outreach, and inclusion</p>	<p>Develop a professional association for data curators, or a curator sub-group in an existing professional association</p> <p>Continue to support networking opportunities for the curation community</p> <p>Continue to outreach and expand a network of curators</p>

Objective 3: Communicate Value of Curation

Short term (1Y)	Middle term (2Y)	Long term (5Y+)
<p>Start to build an evidence base for value added by curation. Identify and evaluate existing metrics.</p> <p>Identify success stories and data champions in the research community who may help promote the value of data curation services</p>	<p>Develop a set of assessment measures, and pilot their application to curation processes in an existing repository such as FRDR</p> <p>Develop and promote clear messaging that can be used to promote the value of curation to various stakeholder groups</p> <p>Actively engage with researchers and data creators to promote the value of curation services, and to ensure that the services provided meet present and evolving needs</p>	<p>Implement a standard set of measures to illustrate value of curation in local, national, institutional and domain repositories</p>

Objective 4: Develop and Maintain Infrastructure

Short term (1Y)	Middle term (2Y)	Long term (5Y+)
<p>Continue to seek funding for development of automated curation tools</p> <p>Set up infrastructure to manage informal curator communication channels</p> <p>Assess <u>C-U-R-A-T-E-D</u> steps⁴⁵ and other models of data curation</p> <p>Establish points of connection between curation and preservation workflows</p> <p>Establish a central knowledge-base as a go-to stop for curators to access tools, resources, documents, guidelines, links to external resources, etc.</p>	<p>Incentivize local investment in data curation HQP</p> <p>Further explore cross-institutional or regional staffing models for curators</p> <p>Explore infrastructure for data sharing amongst curators</p> <p>Explore infrastructure that will allow curators to manage diverse types of data, such as restricted access data, or data with complex rights management policies in place</p> <p>Establish repository integration with digital preservation infrastructure and services</p> <p>Identify methods to support multilingual metadata in repositories across Canada</p>	<p>Pilot a cross-institutional service model for curation</p>

Objective 5: Represent Curators and Inform Governance

Short term (1Y)	Middle term (2Y)	Long term (5Y+)
<p>Establish an advisory council or working group, representative of the multitude of organizations that curate data and manage digital resources, to provide oversight and strategic direction</p>	<p>Advisory council to position curation as key component of RDM, and provide strategic direction in key areas such as policy development, ethics, liability, and sustainability</p> <p>Work with community to develop curation policies to inform local practice</p> <p>Model sustainability in networked curation services</p> <p>Seek funding opportunities to expand curation services and resources</p>	<p>Interface with international data management organizations that have a curation component, and seek opportunities to collaborate</p> <p>Ensure there is a mechanism in place to review and update curation related policies to reflect changes in culture, technology, and best practices</p>

⁴⁵Johnston et al., "Data Curation Network"

Further Considerations

Curation services are an integral part of RDM. Institutions must be able to support the data management efforts of researchers, and they must either offer solutions for data curation and storage themselves, or direct researchers to organizations and repositories that can provide those solutions. Investment in data curation services and personnel will improve the FAIRness of Canada's research output, thus accelerating research, reducing researcher burden, and increasing the overall return on investment for publicly-funded research.

There are a number of challenges to implementing and growing curation services, many of which are mirrored in the challenges of research data management as a whole. Roles and responsibilities are not clearly defined and may vary across stakeholder institutions. There is a need for further skill-building and training opportunities, development of shared services, tools, and resources, as well as policies and standards to guide best practices. Additionally, a culture shift that emphasizes the importance and value of research data and its effective management is necessary.⁴⁶ Through breakout sessions at the Forum and subsequent discussions, the community has articulated many necessary components and functions of a national curation approach and identified a number of areas where further development and discussion is necessary. As has been the case with recent national RDM and DRI reorganization, having broad involvement from diverse stakeholder groups has been critical to articulating a robust, inclusive approach to data curation in Canada. Such involvement will be equally important moving forward.

As it coheres, a national data curation network must strive to support federated infrastructure while growing curation capacity and supporting curation needs at the local level. At the same time, input and engagement from individuals at the local level is imperative to sustaining a Canadian network of curation expertise. Thus, a well-defined, national-level strategy is required to develop new curators and re-skill existing HQP, while supporting their work in institutions, organizations, and repositories, and enabling them to participate in the creation of curation policies, standards, and workflows for diverse data types and contexts. A well-informed, highly-skilled, and distributed network of curation experts will allow Canada's digital research infrastructure to meet the needs of researchers in their current and future work, and can help build resilience against any future loss of funds (either national or local) for research data management. Some degree of local autonomy is important, as well: While best practices and

⁴⁶ Baker et al., "Research data management in Canada."

standards should be used to guide local decisions, institutions should retain a degree of authority over how they are implemented when curating the datasets they ingest or manage.

Although it is not currently clear how such a national network should be managed and sustained in the long-term, there is work that can be undertaken now to inform future decisions. While aspects of this service could be supported by the nascent NDRIO, there are likely opportunities and approaches that could provide support at the regional and local (institutional/organizational) levels. An effort should be made now to benchmark Canada's current curation capacity and needs—including services, training, personnel, and supporting resources—and model approaches to sharing costs, services, and development work across stakeholders. Data producers, researchers, and the scientific community must be actively consulted and involved throughout the process. A survey and small group or individual consultations can be used to assess how current curation services meet their needs. The results and feedback from these efforts can be used to inform the growth of existing services, and the development of new services and resources. An environmental scan that accurately maps the landscape will illustrate potential connections, identify gaps in resources and services, and ensure limited resources are applied judiciously. The ongoing work to connect institutions and data at national data repositories such as FRDR, SPDV, and domain-specific repositories such as CIOOS will also provide valuable insights, as will international efforts to create networks of curation resources and services, such as the DCN, the DCC, and the ARDC.

As a unified approach is being developed, it is important to consider how it promotes inclusiveness, connects diverse institutions and organizations, and creates equity between institutions that have different abilities to participate. For example, many institutional-scale data repositories operating across Canada recognize a present or imminent need for data curators, but lack the resources to support a fully functional local curation team. In such cases, shared staffing models for curation expertise—similar in nature to the DCN—could provide a cost-efficient solution. Similar opportunities exist for national-scale repositories such as FRDR and the federation of Dataverses hosted by Scholars Portal. Options range from a shared staffing model where institution-based curators are seconded for a small fraction of their time (e.g., 5 to 10%), to a centrally-funded curation support team that provides services to all institutions, to a hybrid of both approaches. Shared staffing approaches would allow the network to scale up as local capacity is strengthened and data deposits increase, but would likely be complicated to fund and maintain. More work must be done to assess current capacity in Canada, explore funding options, and model how local institutions, including those without dedicated RDM staff, might participate in such a network before a pilot is implemented.

Important decisions must also be made with regard to expected levels of curation for datasets, metadata, and code, and whether all collections and repositories require an equal level of attention and rigour. If this is not the case, how should they be prioritized and who should make such determinations? Answering such questions requires in-depth discussions amongst broader DRI stakeholder groups to identify points of need, characterize existing capacity, and define stakeholder expectations.

Finally, unanswered questions persist around organizational structure and governance as it pertains to a national data curation approach. It is currently unclear how a national curation service would fit into the emerging DRI landscape, and how it might be managed within or beyond the NDRIO. While this document has suggested some means by which the Canadian curation community could contribute to RDM governance and policy development, much of this will depend upon the structures and mechanisms developed within the NDRIO.

Summary of Recommendations for NDRIO

Data curation is an important component of sound research data management. Curation improves the FAIRness of data, promotes transparency in the research process, and improves funder returns on investment in research.⁴⁷ Although some curation processes can be automated, the complex and heterogeneous nature of research data often requires the intervention of human data curators with a unique complement of disciplinary knowledge, information management skills, collaborative ability, and software expertise. While this report has outlined a set of recommended activities that could be initiated at any point in the future, there is a strong sense of urgency within the community. Data are being generated, managed, and archived now, and thus require curation now. As we work collaboratively to re-envision Canada's DRI ecosystem, there are tangible and effectual actions that may be taken today to meet a number of current needs and lay the groundwork for a broader national curation strategy. These actions are outlined below:

- Acknowledge the importance of data curation in enabling FAIR data products and open scholarship, and the role of data curators in this process.
- Acknowledge that data curation is a complex, intensive activity that requires significant investment in, and development of highly-skilled staff.

⁴⁷ Canada, "Tri-Agency statement of principles on digital data management"; Canada. Innovation, Science and Economic Development Canada, "Seizing Canada's Moment: Moving Forward in Science, Technology and Innovation 2014," 2014, http://www.ic.gc.ca/eic/site/113.nsf/eng/h_07655.html.

- Actively engage with Canadian researchers and data producers, to better understand their curation needs and the challenges they face, and work with these communities to ensure that the curation services we develop will support them.
- Contribute to the creation of a national framework for curation services that supports curation needs broadly and addresses the variety of needs of diverse organizations.
- Support the development and provision of training opportunities, learning resources, and guidance materials in both official languages to support the development of data curators across Canada.
- Collaborate with Indigenous communities to ensure the curation services and resources we develop work in harmony with their data stewardship efforts.
- Support the establishment and maintenance of a national curation community of practice.
- Collaborate with the curation, digital preservation, and ethics communities to address complex issues related to the curation and long-term management of biomedical and other health data, legacy and rescued data, multilingual data, and data generated outside of academic institutions.
- Collaborate with digital preservation experts to identify data management and curation actions that will support long-term preservation.
- Develop a mechanism for the curation community to inform governance and policy decisions related to research data management.
- Directly or indirectly, support the creation of curation positions necessary to ensure excellence in datasets deposited into Canadian data repositories.

Bibliography

Baker, David, Donna Bourne-Tyson, Laura Gerlitz, Susan Haigh, Shahira Khair, Mark Leggott, Jeff Moon, et al. "Research data management in Canada: A backgrounder." Zenodo, December 13, 2019. <http://doi.org/10.5281/zenodo.3574685>.

CASRAI, s.v. "Data curation," accessed February 2, 2020, <https://casrai.org/term/data-curation/>.

CASRAI, s.v. "Research Data," accessed February 2, 2020, <https://casrai.org/research-data/>.

Castle, David, Laura Gerlitz, Susan Haigh, Mark Leggott, Lori MacMullen, Jeff Moon, Benoit Pirenne, Robbin Tourangeau, and Bo Wandschneider. "Data Management Roadmap – 2019-2024: For Submission to Innovation, Science and Economic Development (ISED)." March 29, 2019, accessed February 2, 2020, <https://www.rdc-drc.ca/download/position-paper-data-management-roadmap-2019-2024/?wpdmdl=2482>.

CODATA, and Paul Uhlir. "The Value of Open Data Sharing: A CODATA Report for the Group on Earth Observations". Zenodo, November 16, 2015. <http://doi.org/10.5281/zenodo.33830>.

CoreTrustSeal Standards and Certification Board. "CoreTrustSeal Trustworthy Data Repositories Requirements 2020–2022". Zenodo, November 20, 2019. <http://doi.org/10.5281/zenodo.3638211>.

Digital Curation Centre. "Digital curation 101: Incentives for digital curation." Edinburgh: Digital Curation Centre, 2011, accessed January 10, 2020, <http://www.dcc.ac.uk/sites/default/files/documents/DC%20101%20Incentives%20for%20Curati%20on.pdf>.

FAIRsharing. "FAIRsharing policies: A catalogue of data preservation, management and sharing policies from international funding agencies, regulators and journals." Accessed January 13, 2020. <https://fairsharing.org/policies/>.

First Nations Information Governance Centre. "Ownership, Control, Access and Possession (OCAP™): The Path to First Nations Information Governance." Ottawa: The First Nations Information Governance Centre, 2014. ISBN: 978-0-9879882-8-7. Accessed February 17, 2020, https://fnigc.ca/sites/default/files/docs/ocap_path_to_fn_information_governance_en_final.pdf

Government of Canada. Innovation, Science and Economic Development Canada. “Seizing Canada's Moment: Moving Forward in Science, Technology and Innovation 2014”. Ottawa: Innovation, Science and Economic Development Canada, 2014, modified July 5, 2018, accessed February 26, 2020, http://www.ic.gc.ca/eic/site/113.nsf/eng/h_07655.html.

Government of Canada. Innovation, Science and Economic Development Canada. Office of the Deputy Minister. Communications and Marketing Branch. “Tri-Agency statement of principles on digital data management”. Ottawa: Innovation, Science and Economic Development Canada, December 21, 2016, accessed January 10, 2020, https://www.science.gc.ca/eic/site/063.nsf/eng/h_83F7624E.html.

Government of Canada. Innovation, Science and Economic Development Canada. Office of the Deputy Minister. Communications and Marketing Branch. “DRAFT Tri-Agency research data management policy for consultation.” Ottawa: Innovation, Science and Economic Development Canada, May 25, 2018, accessed January 10, 2020, https://www.science.gc.ca/eic/site/063.nsf/eng/h_97610.html.

Government of Canada. Innovation, Science and Economic Development Canada. Minister of Innovation, Science and Economic Development. “Digital Research Infrastructure.” Ottawa: Innovation, Science and Economic Development Canada, August 7, 2019, accessed January 10, 2020, <https://www.ic.gc.ca/eic/site/136.nsf/eng/home>.

Government of Canada. Office of the Chief Science Advisor or Canada. “Roadmap for Open Science.” Ottawa: Office of the Chief Science Advisor, February 26, 2020, accessed April 22, 2020, https://www.ic.gc.ca/eic/site/063.nsf/eng/h_97992.html.

International Organization for Standardization (ISO). Space Data and Information Transfer Systems. “ISO 16363:2012 Audit and Certification of Trustworthy Digital Repositories”. 2012. <https://www.iso.org/standard/56510.html>.

Johnston, Lisa R., Jake Carlson, Cynthia Hudson-Vitale, Heidi Imker, Wendy Kozlowski, Robert Olendorf and Claire Stewart. “How Important Are Data Curation Activities to Researchers? Gaps and opportunities for academic libraries.” *Journal of Librarianship and Scholarly Communication* 6, no. 1 (April 26, 2018): eP2198. <https://doi.org/10.7710/2162-3309.2198>.

Johnston, Lisa R., Jake Carlson, Cynthia Hudson-Vitale, Heidi Imker, Wendy Kozlowski, Robert Olendorf, Claire Stewart, et al. “Data Curation Network: A Cross-Institutional Staffing Model for

Curating Research Data.” *The International Journal of Digital Curation* 13, no. 1 (December 26, 2018): 125-140. <https://doi.org/10.2218/ijdc.v13i1.616>.

Kim, Jeonghyun, Edward Warga, and William Moen. “Competencies Required for Digital Curation: An Analysis of Job Advertisements.” *The International Journal of Digital Curation* 8, no. 1 (June 14, 2013): 66-83. <https://doi.org/10.2218/ijdc.v8i1.242>.

Leadership Council for Digital Research Infrastructure. “Data Management Position Paper For Innovation, Science, and Economic Development Canada.” August 31, 2017, accessed February 26, 2020, <https://wiki.queensu.ca/display/QDSRWG>.

McQuilton, Peter, Susanna-Assunta Sansone, Helen Cousijn, Matthew Cannon, Wei Mun Chan, Ilaria Carnevale, Imogen Cranston, et al. “FAIRsharing Collaboration with DataCite and Publishers: Data Repository Selection, Criteria That Matter”. Open Science Framework, October 19, 2019, accessed January 6, 2020, <https://doi.org/10.17605/OSF.IO/N9QJ7>.

Mons, Barend. *Data Stewardship for Open Science: Implementing FAIR Principles*. New York: CRC Press. 2018. <https://doi.org/10.1201/9781315380711>.

Open Science Collaboration. “Estimating the reproducibility of psychological science.” *Science* 349, no. 6251 (August 28, 2015): aac4716. <https://doi.org/10.1126/science.aac4716>.

Organisation for Economic Co-operation and Development. “OECD principles and guidelines for access to research data from public funding.” Paris: OECD Publishing, April 12, 2007, <https://doi.org/10.1787/9789264034020-en-fr>.

Perrier, Laure, Erik Blondal, and Heather MacDonald. “The views, perspectives, and experiences of academic researchers with data sharing and reuse: A meta-synthesis.” *PLoS ONE* 15, no. 2 (February 27, 2020): e0229182. <https://doi.org/10.1371/journal.pone.0229182>.

Portage Curation Expert Group. “Primer: Data Curation.” Ottawa: Portage Network, 2019, accessed January 28, 2020 https://portagenetwork.ca/wp-content/uploads/2019/09/Curation_Primer_Aug2019_EN.pdf.

Research Data Strategy Working Group. “The 2011 Canadian Research Data Summit: Mapping the data landscape.” December, 2011, accessed February 26, 2020, http://longrangeplan.ca/wp-content/uploads/2010/10/Data_Summit_Report.pdf.

Stewart, Andrew, Brad deYoung, Mike Smit, Karen Donaldson, Alexa Reedman, Andréane Bastien, Ray Brunsting, et al. "The Development of a Canadian Integrated Ocean Observing System (CIOOS)." *Frontiers in Marine Science* 6 (Aug 9, 2019): 431.
<https://doi.org/10.3389/fmars.2019.00431>.

Stuart, David, Grace Baynes, Iain Hrynaszkiewicz, Katie Allin, Dan Penny, Mithu Lucraft, and Mathias Astell. "Whitepaper: Practical challenges for researchers in data sharing (Version 1)." Figshare (March 21, 2018). <https://doi.org/10.6084/m9.figshare.5975011.v1>.

Tammaro, Anna Maria, Krystyna K. Matusiak, Frank Andreas Sposito, and Vittore Casarosa. "Data Curator's Roles and Responsibilities: An International Perspective." *Libri* 69, no. 2 (May 31, 2019): 89-104. <https://doi.org/10.1515/libri-2018-0090>.

United States. Office of Science and Technology Policy. "Request for Public Comment on Draft Desirable Characteristics of Repositories for Managing and Sharing Data Resulting From Federally Funded Research." Federal Register, Doc. 2020-00689, 85 FR 3085, (January 17, 2020), accessed February 26, 2020, <https://www.federalregister.gov/documents/2020/01/17/2020-00689/request-for-public-comment-on-draft-desirable-characteristics-of-repositories-for-managing-and>.

Wilkinson, Mark D., Michael Dumontier, IJsbrand Jan Aalsbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. "The FAIR Guiding Principles for scientific data management and stewardship." *Scientific Data* 3 (March 15, 2016): 160018.
<https://doi.org/10.1038/sdata.2016.18>.

Appendix

List of Acronyms

Acronym	Term
ARDC	Australian Research Data Commons
CARL	Canadian Association of Research Libraries
CDCF	Canadian Data Curation Forum
CIOOS	Canadian Integrated Ocean Observing System
DCC	Digital Curation Centre
DCN	Data Curation Network
DMP	Data Management Plan
DOI	Digital Object Identifier
DRI	Digital Research Infrastructure
FAIR	Findable, Accessible, Interoperable, Reusable
FNIGC	First Nations Information Governance Centre
FRDR	Federated Research Data Repository
LCRDM	National Coordination Point Research Data Management
HQP	Highly Qualified Personnel
ISED	Innovation, Science and Economic Development Canada
NDRIO	New Digital Research Infrastructure Organization
OCAP©	Ownership, Control, Access, and Possession
RDC	Research Data Canada
RDM	Research Data Management
SSHRC	Social Sciences and Humanities Research Council
SPDV	Scholars Portal Dataverse

CDCF Community-Building Forum participant breakdown

Table 1: Breakdown by organization type

Organization Type	Count
Academic Institution - Canada U15	21
Academic Institution - US	13
Non-Governmental	10
National / Provincial Research Service Provider [‡]	9
Government Agency	5
Academic Institution - Canada non-U15	2
Public Library	1

[‡] Includes CARL Portage, Compute Canada, Scholars Portal, Research Data Canada

Table 2: Breakdown by geographic region

Region	Count
Ontario	29
Canada - West	15
Canada - East	11
Quebec	4
International	2

Table 3: Breakdown by position

Position	Count
Librarian / Library Staff	31
Director	5
Manager	5
Analyst	4
Associate University Librarian / Associate Director	3
Research Ethics Manager / Coordinator	3
Project Manager / Officer	3
Data Curator / Specialist	3
Associate Vice President, Research	1
Bibliometrician	1
Computing Specialist	1
Digital Curation Archivist	1

Overview of national data repositories presented at the CDCF

Canadian Integrated Ocean Observing System (CIOOS)

CIOOS will integrate Canadian oceanographic data from regional associations to make interoperable datasets accessible to end-users through a national web platform. Regional associations will collaborate with regional data providers and users to adopt international standards and tools to ensure interoperable data management and dissemination. Challenges addressed in the creation of the CIOOS network included integrating complex and diverse data from the three regional associations; accounting for variance in local policies, procedures, and infrastructure; and establishing a framework to centralize distributed project management and working group activities. These challenges and the solutions implemented by CIOOS will provide valuable insight if curation services will be networked.

Federated Research Data Repository (FRDR)

Created in a partnership between CARL Portage and the Compute Canada Federation, FRDR promotes findability by harvesting metadata from 45+ repositories nationally and exposing results in a federated search interface. It is also a curated repository for large datasets, leveraging Globus file transfer for data upload and download. FRDR is currently in limited production, with plans to transition to full production in the Fall of 2020. At that time, all Canadian faculty members, or their designates, at eligible postsecondary institutions will have access to a large default allocation of curated repository storage, with a process to request more storage if needed.

Scholars Portal Dataverse (SPDV)

Launched in 2012, SPDV currently has over 50 participating institutional members across Canada, and has seen a rapid uptick in the number of published datasets over the past 2 years. In an effort to strengthen the Dataverse community and improve accessibility for researchers and the public alike, Scholars Portal is working with the community to build a national instance of Dataverse available to all Canadian postsecondary institutions. As a part of its service expansion, Scholars Portal is partnering with CARL Portage to enhance storage infrastructure for Dataverse, helping to lower service costs for institutions nation-wide. Through this partnership, Scholars Portal will host and operate this service, with national RDM community oversight and coordination facilitated through CARL Portage. Current work is focused on developing partnership agreements with regional library consortia, and defining national governance structure and policies.