

Closing Canada's "curation gap": A national approach

White Paper from Portage Curation Expert Group (CEG) for NDRIO, 14 December 2020

Primary Contacts:

Michael Steeleworthy (msteeleworthy@wlu.ca), Wilfrid Laurier University
Jay Brodeur (brodeujj@mcmaster.ca), McMaster University
Erin Clary (erin.clary@carl-abrc.ca), Portage Network

Signatories within Portage CEG:

Marcel Fortin (University of Toronto)
Lelland Reed (NSCAD University)
Sandra Sawchuck (Mount Saint Vincent University)

INTRODUCTION

Data curation is a key part of research data management (RDM, or DM) that improves the quality and value of research data¹. By engaging in activities such as metadata development and inspection, variable standardization, file and system audits, and anonymization, curators ensure that research value is maximized, data can be synthesized, risk can be managed, and research outputs can be discovered. These critical curatorial practices are carried out by researchers, curation specialists, and other highly qualified personnel (HQP) throughout the stages of research, including: research design, collection, quality control, analysis, and archiving. On the whole, curation ensures that datasets are clean, have context, and are usable. It is a critical component of data stewardship that safeguards the data legacy of a research program.

While the value of curation is considerable, it can be difficult to instill curation practices into a research program and the broader research ecosystem. Research teams are often juggling research, organizational and administrative priorities, and curation and broader DM practices may be pushed downstream in the face of significant time or resource constraints. Decisions made early in the research process may affect a dataset's utility for analysis and synthesis--and ultimately, its long-term value. To address this, the power of data curation - its resources, people, and tools - must be leveraged throughout the research life cycle, as a form of active data management that supports individual research programs and the nation's research enterprise.

This document addresses the data curation gap that Canadian research programs and organizations face, including the ability to manage a diverse variety of data, lack of skilled staff, unintegrated curation programs, and the inequity/unevenness in curation engagement/capacity among Canadian research institutions. It envisions an ideal state where curatorial best practices, infrastructure, and people are fully engaged in research communities across Canada. We propose that NDRIO support efforts to engage with researchers in evaluating curation services, strengthen Canadian curation capacity through training, expand and grow curation infrastructure, and cultivate relationships with other DRI stakeholders to develop and maintain a common vision and success.

CHALLENGES

The simple truth is that we have a “curation gap”. Curators are often faced with a wide range of data types from many disciplines, and with greater diversity comes an increasing number of challenges. There is a **need to support curation for all types of research — this requires expertise, but also a community of practice to share knowledge and develop and refine best practices and standards**. Curation takes a lot of time, especially when it is performed at the very end of a research project. There is a **need to find ways to engage with and support researchers throughout the research lifecycle**. Lastly, there is an inequity in curation

¹ Portage Curation Expert Group, “Primer: Data Curation,” Ottawa: Portage Network, 2019, https://portagenetwork.ca/wp-content/uploads/2019/09/Curation_Primer_Aug2019_EN.pdf

capacity and expertise amongst institutions, many of which lack the necessary staff to perform curation. There is a **need for training and resources** to re-skill HQP and a need to promote the value of curation so institutions are willing to invest in **people and infrastructure to support research excellence across Canada**. These challenges are explored below.

Managing diverse types of data

Data that must be restricted in accordance with privacy laws or cultural norms (e.g., dynamic data, big data, third-party data, software, and data in proprietary or specialized formats) all pose unique challenges. Curators may require guidance to make well-informed decisions about long-term preservation, intellectual property rights, ethical reuse, or participant consent. Curators also benefit from the expertise of fellow curators. These include RDM librarians and other HQP such as preservationists, ethics professionals, and advanced research computing and software specialists, but these positions do not exist in every institution, and where they do, there may be limited intersection between departments.

Developing HQP

Although a number of institutional-scale data repositories recognize the benefits of curation, many lack the resources to support a fully functional curation team. Where dedicated curators do exist within individual research groups, organizations, and institutions, one individual is unlikely to possess all of the skills and expertise required to manage a diverse body of data throughout its lifecycle.² As such, in Canada, there is a need for bilingual training opportunities to further develop the skill-sets of existing curators, as well as infrastructure to support high-level coordination of data curators and data curation practices across organizations for the purpose of sharing knowledge, expertise, and curation capacity.

Meeting researcher expectations

Researchers are in the best position to work with curators when it comes to understanding their data and the tools necessary to interact with it, but researchers are juggling many competing priorities. A better understanding of researcher expectations and their needs throughout the research lifecycle is necessary to maximize curator efficiency and ensure the services we develop and provide are those most useful to Canadian researchers and data producers. Data curation can be a time consuming endeavor and decisions must be made regarding expected levels of curation for datasets, metadata, and code, and whether all collections and repositories require an equal level of attention and rigour. If not, how should curation services be prioritized and who should make such determinations?

Providing equitable services

There is inequity and unevenness in curation engagement, capacity, and expertise amongst academic institutions. Clear messaging to promote the value of curation, and incentives for local institutions to invest in HQP to support the curation needs of researchers are necessary. Broad

² 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>.

engagement with stakeholders in the entire research ecosystem, including stakeholders such as disciplinary repositories, government and non-governmental organizations, and Indigenous communities, is necessary in order to understand their strengths and needs, to identify areas for collaboration, and to ensure that we develop equitable, scalable curation solutions.

Recognizing value

Curation adds value to research data, but this value must be demonstrated to stakeholders, including researchers and administrators. Curators need a means of articulating their accomplishments in a way that is recognized by the academy. The recent UNESCO draft [Recommendation on Open Science](#) advocates for member states to “actively engage in removing the barriers and disincentives for Open Science, particularly those relating to research and career evaluation and awards systems.” Better recognition of the work of curation goes hand in hand with emerging systems that better acknowledge and evaluate the contributions of open research data to academic career progression.

IDEAL FUTURE STATE

All Canadian researchers must have equal access to high-quality curation services and resources within Canada’s research ecosystem. We see this support being offered through: **distributed HQP** with diverse expertise and domain knowledge to meet researchers’ curation needs, **a network** to serve all the groups involved in data curation, and **higher-level governance and coordination** to provide assessment and strategic direction. This distributed network of curation experts will help our digital research infrastructure meet researcher requirements, enable services to scale, and build resilience against any future loss of funds (either national or local) for research data management.

Ideally, researchers and other stakeholders will understand the value of data curation for improving research quality and outcomes. Curation services will be designed to meet the most pressing needs of researchers, and researchers will be supported throughout a project by curators and other HQP, such as RDM librarians, research ethics boards, and preservation, metadata, ARC, and RS specialists. The benefits of curation and the potential for data reusability and study reproducibility will be maximized when HQP are available to assist with active data management and when curation is integrated into all research support processes.

Curation knowledge and expertise must also be improved across the entire DRI enterprise. Curation capacity should be nurtured at the institutional level so that opportunities are provided to smaller projects and connections are made with other local service providers. Training opportunities and national infrastructure that supports a curation community of practice should be supported and will allow curators to network, share knowledge, expertise, tools and best practices.

Continued support for existing national infrastructure, such as the Federated Research Data Repository (FRDR) and Scholars Portal Dataverse, will also be important. These national initiatives allow all Canadian researchers to leverage services in support of their long-term data

stewardship efforts. In an ideal future, any researcher who wants assistance with active data management or long-term data storage will have their needs met, and Canada will see an increase in the number of data assets that are shared, deposited for long-term archiving and preservation, and reused by research communities and the public.

BRIDGING THE GAP

We propose a 4-part strategy to close the national curation gap:

Engage researchers

- NDRIO should actively consult with data producers, researchers, and the scientific community to understand the challenges they face in preparing data for long-term storage and sharing. If their data management and curation needs are not being met, what services should be expanded or added to best support them?
- NDRIO should provide researchers with continued training opportunities and develop resources that will further strengthen researchers' active data management skills. Helping researchers incorporate RDM best practices into existing workflows should ease the burden of managing data and increase the number of datasets that are 'deposit ready' at the end of a research project.

Strengthen curation capacity in Canada

- NDRIO should support data specialists through continued development and provision of training opportunities, learning resources, and guidance materials in both official languages to expand the skill sets and capacity of data curators across Canada. We see NDRIO playing a key role in building a strong curation community of practice with infrastructure that allows curators to network and share knowledge, expertise, tools and best practices.
- NDRIO needs to both develop coordinated messaging that promotes the value of data management and curation, and incentivize local institutions to invest in HQP to support curation and active data management. Investments in HQP will help ensure excellence in datasets deposited into Canadian data repositories.

Support and expand infrastructure

- NDRIO should continue to support the development of repository options that provide mechanisms for data curation, such as Scholars Portal Dataverse and FRDR.
- In partnership with other stakeholders, NDRIO should support a curation commons with secure computing space, shared tools, and software that curators and preservationists can use to access, appraise, and manage research data once it has been deposited.
- In partnership with other stakeholders, NDRIO should support a research commons that will allow researchers to easily access the tools and resources they need to actively manage, analyze, and visualize research data, including access to curators and other HQP.

- NDRIO should evaluate mechanisms to support collaboration across institutions and organizations so curators and other HQP can easily connect to share resources, best practices, and the knowledge and expertise necessary to best support Canada's researchers in a seamless manner.

Cultivate relationships among various stakeholders

- NDRIO should continue to foster the type of cross-institutional relationships that CARL Portage has built among research support staff, and should facilitate opportunities for curators to establish new working relationships with other HQP.
 - Connecting curators with RS and ARC professionals provides an opportunity for curators to share knowledge on active data management, while learning more about managing big data deposits and models, and implementing code or software necessary to generate, clean, process, or analyze data.
 - Connecting curators with ethics professionals will allow the groups to build a shared understanding of the complexities inherent to managing diverse types of sensitive data in the short and long term.
- It is essential for NDRIO to continue to engage with academic libraries that have committed to, and invested in data management. Library-based HQP have direct ties to their researcher communities and knowledge of institutional solutions to address their data management needs.
- NDRIO should foster connections and dialogue between academic institutions and other members of the wider research ecosystem (including Indigenous communities, disciplinary repositories, non-governmental organizations, and research groups) to ensure Canada's curation services and resources benefit them and work in harmony with their data stewardship efforts.

CONCLUSION

As Canada reimagines its DRI ecosystem, this paper presents a strategic approach for closing the *curation gap* and ensuring that our nation's research data is usable and valuable to generations of researchers to come. Engaging with researchers to evaluate how current curation services meet their needs is a high priority. Their feedback should be used to inform the growth of existing curation services and the development of new services and resources. To ensure that curators are prepared to meet researcher needs, a national service should support the development of educational opportunities to re-skill existing HQP and incentivize local institutions to invest in HQP to meet current and future demand. The continued cultivation of relationships between curators and other HQP who support researchers is also necessary to build a seamless service with support for researchers in any phase of the research project lifecycle. Finally, national governance and coordination is required to provide ongoing assessment, direction, and strategic investment as the Canadian DRI and research ecosystems evolve.

This proposal has been shared with and is endorsed by the Canadian Association of Research Libraries (CARL) and the CARL Portage Network.