Enabling broad reuse of Canada's geospatial data and digitized cartographic materials

A response to the NDRIO Call for White Papers on Canada's Future DRI Ecosystem

December 14, 2020

Primary Contacts:

Jay Brodeur (brodeujj@mcmaster.ca), McMaster University

Kara Handren (kara.handren@utoronto.ca), Scholars Portal

Francine Berish (francine.berish@queensu.ca), Queen's University

Martin Chandler (martin.chandler@mcgill.ca), McGill University

Marcel Fortin (marcel.fortin@utoronto.ca), University of Toronto

Amber Leahey (amber.leahey@utoronto.ca), Scholars Portal

Rhys Stevens (rhys.stevens@uleth.ca), U. of Lethbridge

This paper has been shared with and is endorsed by the Association of Canadian Map Libraries and Archives (ACMLA)

Summary

- Researchers from a broad array of disciplines depend on Canada's federally-produced geospatial data and digitized cartographic materials to provide information on environments and changes over time.
- The use of these products in foundational and transformational research is hindered because many of them are not FAIR. Researchers are unable to find, access, and use the resources they need for their research.
- Collections of these materials are distributed across a variety of stewarding government agencies, libraries, and archives. They vary significantly in their discoverability, description, availability, and format-compatibility for modern analytical approaches.
- Canada needs to invest in robust and enduring DRI that preserves these unique and invaluable collections, and enables their broad discovery and reuse in research.
- National-level coordination is required so that distributed stewarding organizations can work together to inventory these materials, describe them, and where needed, digitize and transform them into research-ready products.
- A national DRI strategy should facilitate collaboration between data users, producers, stewards, and providers to develop a centralized clearinghouse for these resources, as well as the underlying standards and technical infrastructure that will make these materials FAIR for future generations of researchers.

Introduction

Geospatial data and digital cartographic materials produced by the Government of Canada are critically important resources for research across a wide range of disciplines. These rich collections of born-digital vector and raster data, as well as digitized urban and rural county maps and plans, topographic maps, aerial photographs, and resource and railway maps provide researchers with granular and unique information about Canada's past environmental and human environments, and their changes over time. The diversity of ecological, socioeconomic, geopolitical, and infrastructural information preserved within these materials supports a wide variety of research areas including architecture, engineering, digital humanities, history, economics, demography, public health and policy, geography, environmental studies, and more[1]. Making the entirety of these resources discoverable and openly accessible for use in research, teaching, and public-interest applications will foster innovation and maximize Canada's net return on investment.

Yet, large swaths of these collections are unavailable to, or unusable by the researchers and other highly qualified personnel (HQP) who need them. Despite efforts by some government agencies, academic libraries, museums, and archives to describe, digitize, preserve, and provide access to their holdings, the lack of common infrastructure, workflows, capacity, and standards to coordinate and support this work results in resources that vary significantly in their FAIRness for research. For example, researchers hoping to access geospatial information at the national level are required to navigate diverse offerings across various departments (e.g., Natural Resources Canada) and agencies (e.g., Geological Survey of

---

[1] Simpkin, S., Brodeur, J., Woods, C., Leahey, A., Beard, C., Janzen, S. "Historical map digitization in libraries: Collaborative approaches for large map series". Bulletin of the Association of Canadian Map Libraries and Archives, 152: 11-28.

Canada), as well as local academic library collections. Leveraging this foundational geospatial data requires an understanding of the governing bodies that create the data, as well as the systems that organize and present them. As such, a broader coordinated and collaborative effort is required to build digital research infrastructure (DRI) that makes the entirety of these resources usable for researchers. In support of such a vision, this paper documents the challenges currently faced by researchers and stewarding organizations alike, and articulates the actions required to realize an idealized national DRI state. While these types of materials are part of a broader conversation that includes researcher-created data and its management, this whitepaper focuses on opportunities that are unique to geospatial collections stewarded by government agencies, academic libraries, and archives.

Current barriers and challenges

From the researcher perspective, there is a cascading set of barriers to reusing government-produced geospatial data and cartographic material in scholarship. In most cases, these barriers are outward manifestations of challenges faced by stewarding organizations in enabling such reuse. These barriers and challenges are described below in alignment with the FAIR Principles:

FINDABLE

Researchers are unable to determine what resources exist, or can't search for them in an efficient and intuitive manner.

In the current ecosystem, these collections are stewarded by a wide range of government agencies, research institutions, academic libraries, museums, and archives with greatly differing access to expertise, tools, and infrastructure to enable reuse. Many of these collections are poorly described and indexed (or not at all), and the existing finding aids and catalogues vary widely in terms of their upkeep, functionality, and efficacy. Few collections are indexed and searchable by location, which is the most common and intuitive method of finding appropriate geospatial materials. Without a common clearinghouse[2] to search for these resources, researchers need to know beforehand where to look for these materials or require help from map and geospatial data specialists (often in academic libraries). Otherwise, they resort to broad searches, or give up altogether. For example, a researcher interested in land-use in Canada at the national level could start with the Forest Resource Inventory (FRI), Canada Land Inventory (CLI), or Annual Crop Inventory found on the Open Canada Portal, before exploring comparable geospatial data at a provincial or municipal scale. There are a number of completely separate interfaces for searching the foundational collection of maps available through Library and Archives Canada (LAC). The need to navigate complex naming conventions and different interfaces is a defining researcher experience when working with geospatial data.

---

[2] See the USGS Earth Explorer and topoView systems as examples.

ACCESSIBLE

In cases where researchers discover materials of interest, they are often inaccessible or unavailable for the intended type of reuse.

A significant proportion of Canadian cartographic collections exist only in print format, and are unavailable for digital research purposes. Digitization of these collections is impeded by a collection of factors, including: funding shortfalls, insufficient information and inventories to meet the policy requirements for scanning, and a lack of clarity on Crown Copyright restrictions. Organizations have limited personnel, time, training, and tools to digitize materials, describe them appropriately, and develop discovery systems that make them accessible to users. Furthermore, lacking access to robust and affordable discovery, repository, and preservation infrastructures means that there is limited capacity to ensure the long-term viability of digitized and born-digital materials alike.

INTEROPERABLE

Many of the available resources are difficult to integrate into research because they are poorly-described or not in research-ready formats.

In the absence of a federated approach to enabling reuse of these digital and digitized materials, many organizations are doing their best to fill this gap with limited resources. This decentralized, uncoordinated approach is incredibly inefficient; efforts are duplicated and products are difficult to index or combine across systems due to a lack of common standards for metadata, description, and file formats. For example, member libraries in both the COPPUL and OCUL consortiums have undertaken separate projects to inventory and describe NTS print maps in their collections, despite these containing many of the same sheets. Strategies and infrastructure for providing research access also vary greatly across the country, even for the same materials, thus creating duplicated and inefficient workflows for sharing. For digitized media specifically, researchers commonly require value-added formats (e.g., georeferenced and vectorized maps) that are time- and expertise-intensive to produce, and require specialized infrastructure to display, share, and preserve. Without shared approaches and infrastructure to coordinate work, this patchwork approach will persist to the detriment of Canada's research enterprise.

REUSABLE

Researchers are unable (or unclear on permissions) to use these materials as intended for their research.

While the Government of Canada has made progress in applying the Open Government License to recently-created data products, there remains a significant proportion of materials -- historical maps in particular -- for which copyright status and reuse permissions are unclear and confusing. Without explicit declaration from the Government of Canada on the copyright status of these materials, stewards will be reluctant to share them openly, and researchers will be reluctant to use them to create new research products and share them alike. These important historical data require specific attention in order to meet current expectations for research data; concerted efforts should be made to connect these resources to the current national DRI landscape in order for them to become more readily reusable.

Towards an ideal DRI state

Canada's digital research infrastructure should enable individual researchers and institutions to share, find, access, and reuse geospatial data and digitized cartographic materials in a manner that is frictionless, barrier-free, intuitive, and seamless. Building and maintaining DRI that meets these aspirations requires investment in, and coordination of, the many components that comprise it. These include:

- The digital materials and tools that researchers and other HQP use in their research;
- The systems and information that allow for these materials to be discovered, accessed, and preserved;
- The people, tools, and processes that support creation, access to, and long-term stewardship of resources and systems; and,
- The communities who establish workflows, best practices, and standards.

While there is a great deal of work required to reach this aspirational state, we believe that it is attainable through the following actions:

1. INVEST IN LOCAL CAPACITY

Considering that collections and expertise are currently distributed across a variety of federal agencies, academic libraries, and archives, it is imperative that these organizations are provided with the support they need to contribute to national collection consolidation and system development initiatives. Strategic investments in time, training, equipment, and tools will enable local organizations to participate in national efforts, while also building their capacity to support the specific research needs of their users. Investing in resources at the local level will also augment initiatives like the [National Heritage Digitization Strategy](#), which aims to digitize and preserve Canada's documentary heritage, including historical maps.

2. INVEST IN NATIONAL SYSTEMS AND TOOLS

Centralized infrastructure is vital to realizing the reuse potential of Canada's expansive but distributed collections of federal geospatial data and cartographic materials. Such infrastructure should federate discovery, distribution, and preservation of these materials to benefit researchers, while allowing stewarding organizations to contribute their collections and expertise to national consolidation efforts.

Through the development of shared systems and tools for digitization, georeferencing, and metadata creation, organizations from across Canada can contribute to a federated digital collection using streamlined, standardized workflows. These products should, in turn, be indexed and presented in a federated discovery, access, and visualization interface that allows researchers to find data using spatial and textual search, and download data products in formats that meet their needs. Such a system may be similar to the [Geodisy map search](#) currently in beta as part of the Federated Research Data Repository ([FRDR](#)), which allows for spatial searching of research data housed in participating repositories. In the case of these geospatial data and digitized cartographic materials, any system should also be connected and integrated as much as possible with scholarly publishing mechanisms and long-term preservation storage that ensures these materials are accessible well into the future.

3. FACILITATE COLLABORATION AND COORDINATION

Before work can begin on national systems that serve both data users and providers, there is a critical need to bring together members of these communities to define requirements, articulate a shared vision, and lay the groundwork for collaboration and coordination that will realize it. National leadership is needed to facilitate collaboration among these groups to define and develop national standards for file and metadata formats, value-added products, and digitization and preservation workflows. Greater collaboration will also promote the formation of networks of expertise that can support ongoing training, administration, and support needs[3]. These networks of expertise could be built around existing community and professional organizations such as the ACMLA, which support these activities among Canadian map librarians, map archivists, and geospatial data specialists.

National-level coordination is also required so that the distributed organizations that currently steward these materials can work together to build a truly unified national collection. A coordinated approach to collections development and digitization would help to expose the many collections housed at LAC, while allowing other institutions to focus their digitization efforts around unique collections, and collaborate to make FAIR geospatial products[4].

Concluding Remarks

As outlined in this paper, there are many challenges to overcome in enabling broad reuse of these materials. A cohesive, national DRI program depends on all aspects of digital infrastructure, including geospatial tools and products. The current dispersed system lacks coordination and exacerbates regional disparities in research, policy, and digital infrastructure; the development of a more cohesive DRI for geospatial data and digitized cartographic materials will help to improve these issues and advance geospatial literacy among institutions across the nation. There is a longstanding history of collaboration that has led to Canadian academic libraries acting as repositories for some of these materials. Academic libraries are familiar with the information-seeking behaviour of researchers and are well positioned as stakeholders in future partnerships.

Though there is still much work to be done, strategic investments by NDRIO into local capacity, shared infrastructure, and coordination and collaboration between content users and providers can lay the groundwork for realizing our aspirational goals. Building connections between nationally coordinated digitization funding opportunities, local workflows and infrastructures, and national discovery, access, and preservation infrastructure will provide institutions and researchers with invaluable digital research tools and services to support their research capacity moving forward. Collaboration and coordination across institutions in a range of activities including inventories, digitization, georeferencing, descriptive metadata, resource licensing and sharing, discovery and access, and preservation are required in order to effectively connect these valuable resources for use in digital research today and in the future.

---

[3]The engagement of the Ontario map library community (via OCUL) in the development and use of the Scholars GeoPortal is a successful example of this model.

[4] As in the case of the Historical Topo Maps Digitization Projects undertaken by the Ontario Council of University Libraries (OCUL) Geo Community in 2014. This was a 2.5 year initiative to digitize and georeference early-to-mid 20th century historical topographic maps of Ontario, and provide centralized access to the collection via OCUL's GeoPortal. https://ocul.on.ca/topomaps