Building Capacity through Technology: CanWIN's Role in Supporting Indigenous Science

Claire Herbert
Centre for Earth Observation Science
University of Manitoba
Winnipeg, Manitoba, Canada
Claire.Herbert@umanitoba.ca

Wayne Chan
Centre for Earth Observation Science
University of Manitoba
Winnipeg, Manitoba, Canada
Wayne.Chan@umanitoba.ca

Yanique Campbell
Centre for Earth Observation Science
University of Manitoba
Winnipeg, Manitoba, Canada
Yanique.Campbell@umanitoba.ca

Victoria Dearborn
C. H. Riddell Faculty of Environment, Earth and Resources
University of Manitoba
Winnipeg, Manitoba, Canada
Victoria.Dearborn@umanitoba.ca

Dustin Isleifson
Dept. of Electrical and Computer Engineering
and Centre for Earth Observation Science
University of Manitoba
Winnipeg, Manitoba, Canada
Dustin.Isleifson@umanitoba.ca

***Abstract*—Addressing the challenges posed by climate change in recent decades, particularly in Arctic and freshwater ecosystems, requires integrating diverse datasets across disciplines, cultures, and political boundaries. Although Canada has invested significantly in research infrastructure, much of its environmental data continues to be fragmented and difficult to access. The Canadian Watershed Information Network (CanWIN) provides a decentralized open-access platform that fosters research collaboration and knowledge sharing. By implementing CARE and FAIR principles and best practices for data stewardship, CanWIN enables ethical and transparent data sharing, supporting reconciliation efforts with Indigenous communities and enhancing evidence-based decision-making in response to climate and environmental change. An example project in collaboration with the Manitoba Métis Federation is highlighted.**

***Index Terms*—data portal, data sharing, environmental data,climate change, water management strategies, water vulnerability, water governance**

I. Introduction

In the past two decades, climate change and challenges such as water security have become critical issues in both science and society. There is a pressing need to identify emerging concerns and for the advancement of evidence-based strategies for adaptation. Grappling with ecosystem-wide changes in Arctic and freshwater systems demands access to large heterogeneous datasets that span political, cultural, and disciplinary boundaries. This necessitates a specialized research data infrastructure with a spatial component (SRDI) to advance and streamline the sharing, discovery, analysis, and visualization of varied datasets, including sensitive Indigenous knowledge.

A. The Canadian Context

To support the collection of data, Canada has made significant strides in investing in research infrastructure, including research vessels and field stations. Combined with geospatial and remotely sensed data, these infrastructure investments permit data collection at broad geographic scales. Nevertheless, most data remain disjointed and publicly inaccessible or difficult to find, making them essentially hidden from potential users [1]. Furthermore, many diverse groups generate site-level Arctic and freshwater environmental data, with differing approaches to collecting, analyzing, accessing, and sharing the data [2], [3]. Many of these datasets are labor intensive and expensive to collect and not readily replicated [4]. Unique challenges are also presented by Traditional Indigenous Knowledge (IK) and and Inuit Qaujimajatuqangit (IQ), which may be proprietary or sensitive in nature, necessitating particular considerations for their availability and access. Additionally, few datasets are presently delivered back to the initiating Indigenous or Inuit organization in an easy-to-understand and contextualized format, which hampers evidence-based decision-making by these groups. To address ecosystem-wide research questions, critical site-level data are most effective when combined with geospatial data at multiple temporal and geographic resolutions [1]. This form of data harmonization adds significant value beyond the costs of initial data collection [1], and when coupled with clearly documented metadata, ensures long-term data provenance.

As mandated by the United Nations, a fundamental human right is having access to clean water [5]. To create a Canada-wide strategy for the management of water, it is essential that we integrate an approach based on human rights that ensures the rights of Indigenous Peoples, communities, and individuals are met. Data collection in Canada is often disparate and obscured by layers of peripheral or nonessential information. Clear metadata about the authoritative source of a dataset, identifying the location of data collection, by whom it was collected, and the dataset's theme (e.g., nutrients, groundwater, etc.) does not currently exist. Our understanding of where gaps in our knowledge are present is hindered if we do not have

Funding provided in part by Environment and Climate Change Canada, Province of Manitoba, and Natural Resources Canada.

a grasp on the type and amount of current and historical data that are available.

Indigenous Knowledge (IK) is still often ignored in Western scientific analysis, however a common requirement from funding agencies is that Western scientists must work with Indigenous communities to collect data in their traditional areas. This model does not serve communities or relationship building with Indigenous rightsholders. In discussing ten calls to action to work towards reconciliation for natural scientists, Wong, et al. [6] state that scientists must seek reciprocal opportunities that allow the co-production and sharing of knowlege in ways that benefit both Western scientists and the Indigenous community.

B. Canadian Watershed Information Network (CanWIN)

Originally created by Environment and Climate Change Canada (ECCC) and called the Lake Winnipeg Basin Information Network, the Canadian Watershed Information Network (CanWIN) is a Canadian open-access information and data repository that is funded in part by the Lake Winnipeg Basin Initiative under Canada's Action Plan on clean water. The network was initiated to support research on significant water quality issues of concern within the Lake Winnipeg Basin [7].

Since 2012, the University of Manitoba has managed the network after it was transferred from ECCC. It is hosted by the Centre for Earth Observation Science (CEOS), a research centre within the Clayton H. Riddell Faculty of Environment, Earth and Resources. The CanWIN platform has evolved into a distributed network focused on providing access to environmental data from source to sea (freshwater to ocean), in order to address critical climate-related issues facing Canada. CanWIN serves as a platform for sharing research data, and for the creation of information and knowledge products to support science-based decision-making in addressing the challenges caused by our changing climate (Fig. 1).

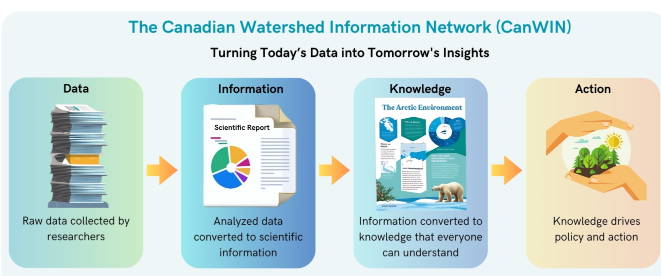

Fig. 1. CanWIN's role in data management.

CanWIN helps connect various stakeholders, including governmental agencies, researchers, environmental organizations, and local communities, to work together in monitoring and managing Canada's water resources effectively. Key pillars underlying CanWIN include: (i) Research Collaboration; (ii) Knowledge Sharing, (iii) Collaborative Governance; and (iv) Public Engagement.

II. The CanWIN Core System

The CanWIN system is steered by a set of principles and services that facilitate our core mandate: to lower the latency for real-time decision-making by supporting the ethical and FAIR (Findable, Accessible, Interoperable, and Reusable) sharing of data across the freshwater-marine spectrum. CanWIN is also aligned with the CARE (Collective Benefit, Authority to Control, Responsibility and Ethics) [8] and OCAP (Ownership, Control, Access, and Possession) [9] principles of Indigenous Data Sovereignty, in addition to the FAIR principles.

Our core principles focus on unified access to data and metadata. We use a decentralized system for data management by acting as a gateway for data creators and publishers to share and access data regardless of its physical location. To accomplish this, we use multiple platforms to allow us to best share the multi-disciplinary datasets we either host directly or federate from external providers (Fig. 2).

CanWIN's services include a user-friendly interface and easy data discovery. To facilitate data discovery, we use an additive data discovery mechanism consisting of three mutually connected search functionalities – a free-text "smart" metadata search which searches on all available metadata fields, curated search term lists composed of internationally recognized standards (e.g., The National Vocabulary Server (NERC) `https://vocab.nerc.ac.uk/search_nvs/`), and a geospatial search, which allows users to search by creating their own region-of-interest (ROI) or by using a pre-defined ROI. Additive metadata services [10] are the foundation upon which all our other services depend. CanWIN provides users with seven templates to capture multiple types of information, and uses a three layer metadata approach. Users are required to complete either a project or dataset template with core metadata information such as author(s) and title, but are also able to provide additional recommended and research-specific metadata. A unique identifier in the form of a DOI can be assigned to the dataset and an ORCID (Open Researcher and Contributor ID) identifier to the researcher. The research-specific metadata fields provide data creators with the ability to record the sample collection and analytical methodology used in the data collection and analysis process as well as allow publishers to provide supplemental details about data or publications related to the project or dataset that are held elsewhere, permitting cross-linking between datasets and projects regardless of physical location. An auto-generated data dictionary allows data publishers to provide additional detailed descriptions for each column in a tabular dataset, facilitating clarity for end-users.

To support reconciliation in the Canadian context, CanWIN implements multiple tools and processes. The core metadata collected ensure clear data provenance, including data providers, data maintainers and data owners, while project and dataset pages clearly identify the organization that provided the data. CanWIN data curators work with data providers to support the curation of datasets for interoperability, as well as

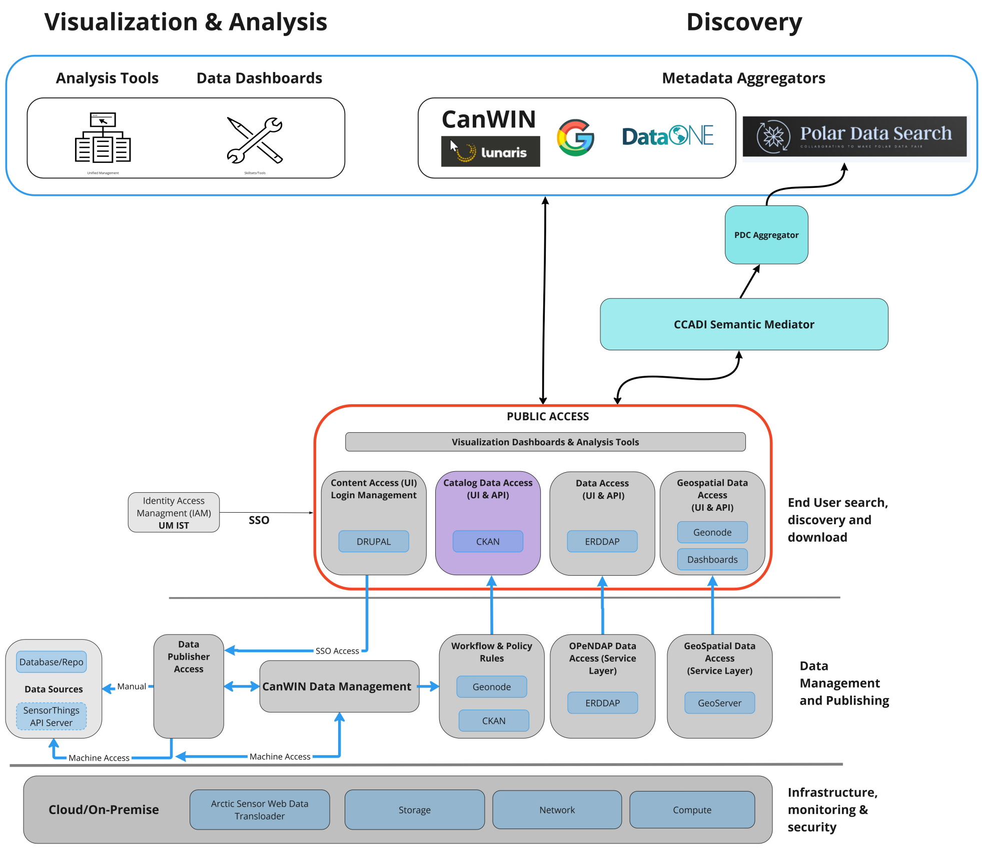

Fig. 2. CanWIN architecture.

creating easy-to-use tools for data analysis. Where relevant associated datasets exist, data curators work with users to create information and knowledge products to support the sharing of complex information in easy-to-understand formats.

III. Data Stewardship

CanWIN supports community data stewardship by offering accessible visualizations and intuitive tools that simplify data access, analysis, and interpretation. Recognizing the importance of data integrity, we have incorporated data curation workflows that ensure the provenance of the data collected. Having reliable data provenance bolsters trust in the data and methods used, and fosters community confidence in the ownership and management of the data.

A. Documentation

Comprehensive documentation in our data workflows extends beyond technical clarity; it becomes a vital tool for empowering Indigenous communities. By hosting accessible guides (e.g., for installation, troubleshooting, and equipment testing), we foster transparency and trust. This approach supports sustainable data sovereignty, enabling communities to independently manage and validate their data and safeguard any traditional knowledge they choose to share publicly, while upholding scientific integrity.

To promote the reusability and transparency of data, we create data “cookbooks” and “codebooks”, serving as detailed guides that document the essential steps involved in data processing/analysis and scripting, respectively. From a data management perspective, documenting these key steps is invaluable, as it allows for a streamlined, consistent approach to working with a specific type of data. This makes it easy for sampling or analytical methods to be accurately replicated, which may promote their reuse. Reproducible data and methods demonstrate the scientific rigor of researchers, increasing their credibility and potential citations.

It is important that, where appropriate, research documentation also embodies Indigenous cultural knowledge, methods, and principles. Inclusive documentation ensures that valuable Indigenous knowledge is not only protected, but recognized by Western science. Documentation should bridge traditional wisdom with modern research practices, supporting communities in sharing their knowledge alongside Western methods in order to understand the data.

1) Codebooks: Data codebooks document salient functions and modules in scripts that were used to process or analyze the data. Codebooks should have:

- A short overview of what the scripts do.
- Setup: computing environment/folder setup; any input files required for the scripts to run; and packages or

modules that need to be imported to run the scripts.
- Location of the script.
- A description of key functions/sections in the scripts.
- Steps to actually run the script.

2) Cookbooks: Data cookbooks document the flow of data from its raw state to its final, processed state. They can be tailored to suit the specific needs of a project and adapted to various levels of data transformation. The CanWIN approach is to simplify this documentation workflow by creating a straightforward, reusable template. Key data management elements in the cookbook that ensure data integrity are described as follows:

1) Critical Summary
    - The critical summary provides an overview of the data, outlining its purpose and the methods of collection. It also encourages descriptions of the data that reflect Indigenous knowledge systems.
2) Raw Data
    - Location of the raw data: Core CanWIN datasets are stored within our GitLab infrastructure, where the data are secure and backed up. Data producers are given access as needed and can work on their data while maintaining a standardized repository structure.
    - Data type: The raw data type is essential to the sampling and processing techniques employed. A future change in the raw data type could impact the processing method, such as scripts tailored to read and handle a specific data format.
    - Scripts or software used to process the data: Whether an internal script or external software is used, identifying the tool used to process the data is essential for ensuring reproducibility. Scripts can also be shared directly from the CanWIN site.
3) Data Processing
    - The data processing section is intentionally designed to be adaptable; the data producer should essentially delineate the steps taken in the data processing workflow. When multiple versions of the data are produced during processing, they may categorize each segment to reflect the different intermediate steps.
4) Final Data
    - Location of the processed data: The processed data should be published in an open-access repository, such as CanWIN.
    - Location of the codebook: If a script was used in data processing, including a codebook that describes key modules or functions in the script is encouraged.
    - File type: The ideal, processed file type should be machine-readable; that is, it should be ingestible by multiple types of software or programming languages. For tabular data, we recommend using CSV files, as they are non-proprietary, widely compatible, and easy to read across different systems.
    - Data Dictionary: To promote understandable and interoperable data, a data dictionary for all variables present in the processed files should be provided. When variables are ambiguously named, the data may be misinterpreted or not understood. To mitigate this, we recommend clear and consistent naming conventions for variables. Additionally, the data dictionary should offer detailed descriptions of each term. This includes providing common names (as well as the associated standardized name from a common vocabulary, like NERC) for variables, definitions, units of measurement, media from which data were collected, and any statistical transformations applied.

IV. Example: the Manitoba Métis Federation's Weather Keeper Program

An example of a CanWIN workflow in supporting Indigenous data sovereignty is the co-creation of the Manitoba Métis Federation Weather Keeper (MMF) program (https://canwin-datahub.ad.umanitoba.ca/data/project/mmf-weather-keeper). In this project, 3-meter high weather stations were installed on MMF citizen property. The data from the weather stations are telemetered via the Iridium satellite network and streamed into an Arctic sensor web data transloader application (https://github.com/GeoSensorWebLab/data-transloader), where the data are standardized using the Open Geospatial Consortium (OGC) SensorThings API. The data are stored on an instance of FROST-Server (Fraunhofer Opensource SensorThings-Server), an open-source server implementation of the OGC SensorThings API (https://github.com/FraunhoferIOSB/FROST-Server). This provides a common endpoint for access.

On the CanWIN webpage for the Weather Keeper program, the project description was translated into the two key languages of the Métis Nation, Michif and French. Cookbooks and codebooks were created to describe in detail the processing steps and scripts used. To enable simple visualization of the data, the endpoint is accessed by GeoNode, an open source mapping platform (https://geonode.org/), and as a simple digital dashboard using Grafana (https://grafana.com/), an open-source analytics and interactive visualization web application. A smart mirror was also constructed using a Raspberry Pi single-board computer. It was designed to be an ambient data display that would blend with typical office decor and be an unobtrusive way to show weather station data in near real-time (Fig. 3). The mirror displayed the Grafana dashboards and was provided to the MMF to use at their head office in Winnipeg, Manitoba.

We also developed a web application (https://canwin.short.gy/mmf-weather) to facilitate the visualization, interpretation, and validation of weather station data from the MMF Weather Keeper Program. With its intuitive and accessible interface, this tool is designed to accommodate users across all skill levels. It allows for two options:

Fig. 3. Smart mirror showing weather data from MMF Weather Keeper station.

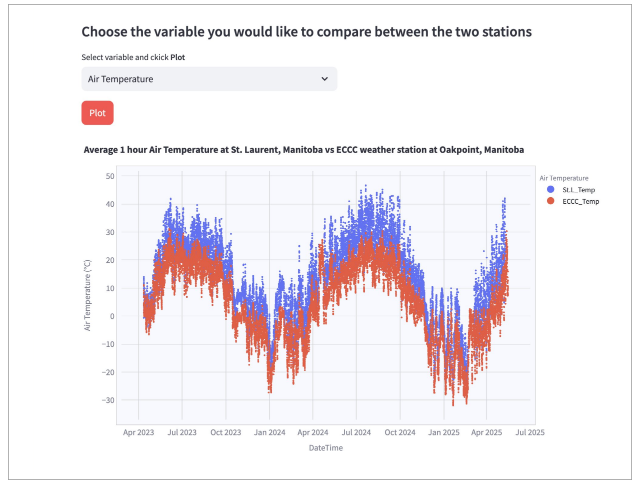

Fig. 4. Graph showing comparison of air temperature between St. Laurent and Oak Point, Manitoba.

- Visualizing and manipulating the different parameters over the entire data period.
- Comparing the data from any MMF Weather Keeper station to the closest Environment and Climate Change Canada (ECCC) station.

The comparison option serves as a tool for validating community-collected data and uncovering potential discrepancies that may arise from differing measurement conditions. For instance, examining measurements from the MMF weather station in St. Laurent, Manitoba, alongside those from nearby Environment and Climate Change Canada stations, such as the station in Oak Point, Manitoba, reveals how variations in location and elevation above ground influence recorded values like temperature (Fig. 4). Understanding these differences enables the community to interpret data more accurately and avoid incorrect conclusions. Moreover, this comparison encourages further analysis, using ECCC stations as a validating source. It also encourages continued collaboration between CanWIN and the MMF to improve data accuracy, ensuring the community's observations are contextualized within a broader framework.

For each weather station deployment, comprehensive documentation covering installation, troubleshooting, and testing is available on the CanWIN site under the ownership of the MMF. This ensures the weather keepers are able to adopt a standardized approach for data collection, which further supports the integrity of the data.

V. Summary

Canada has made significant investments in research infrastructure, yet challenges remain in making environmental data more accessible and usable. The Canadian Watershed Information Network (CanWIN) offers a solution by creating a decentralized CARE- and FAIR-aligned platform that connects researchers, policymakers, and Indigenous communities, making it easier to share and analyze environmental datasets. By prioritizing ethical data stewardship and knowledge co-production, CanWIN fosters collaboration while supporting reconciliation efforts. As concerns about climate change continue to evolve, CanWIN's role in facilitating transparent, interdisciplinary data sharing will be essential in supporting evidence-based decision-making and sustainable resource management.

References

[1] P. A. Soranno, L. C. Bacon, M. Beauchene, K. E. Bednar, E. G. Bissell, C. K. Boudreau *et al.*, "LAGOS-NE: a multi-scaled geospatial and temporal database of lake ecological context and water quality for thousands of US lakes," *GigaScience*, vol. 6, no. 12, p. gix101, 10 2017. [Online]. Available: https://doi.org/10.1093/gigascience/gix101

[2] T. Koivurova and L. Heinämäki, "The participation of indigenous peoples in international norm-making in the Arctic," *Polar Record*, vol. 42, no. 2, p. 101–109, Apr 2006.

[3] G. Whiteman, C. Hope, and P. Wadhams, "Vast costs of Arctic change," *Nature*, vol. 499, no. 7459, p. 401–403, Jul 2013.

[4] P. F. Fogal, L. M. LeBlanc, and J. R. Drummond, "The Polar Environment Atmospheric Research Laboratory (PEARL): Sounding the atmosphere at 80°North," *ARCTIC*, vol. 66, no. 3, Sep 2013.

[5] *Transforming Our World: The 2030 Agenda for Sustainable Development*. United Nations, 2015. [Online]. Available: https://sdgs.un.org/sites/default/files/publications/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf

[6] C. Wong, K. Ballegooyen, L. Ignace, M. J. Johnson, and H. Swanson, "Towards reconciliation: 10 calls to action to natural scientists working in Canada," *FACETS*, vol. 5, no. 1, p. 769–783, Jan 2020.

[7] I. Wong, W. G. Booty, P. Fong, and S. Hall, "Development of an information portal for the Lake Winnipeg Basin initiative," *IFIP Advances in Information and Communication Technology*, p. 667–674, 2011.

[8] S. R. Carroll, I. Garba, O. L. Figueroa-Rodríguez, J. Holbrook, R. Lovett, S. Materechera *et al.*, "The CARE principles for Indigenous data governance," *Data Science Journal*, vol. 19, 2020.

[9] *Ownership, Control, Access, and Possession*. The Alberta First Nations Information Governance Centre, May 2014. [Online]. Available: https://obrieniph.ucalgary.ca/sites/default/files/Nav%20Bar/Groups/GRIP/ocap-faq-may2014.pdf

[10] I. I. Enescu, G.-K. Plattner, L. E. Pernas, D. Haas-Artho, S. Bischof, M. Lehning, and K. Steffen, "The ENVIDAT concept for an institutional environmental data portal," *Data Science Journal*, vol. 17, 2018.