Role for NDRIO in advancing open data sharing and the use of domain-specific repositories

About DataStream

DataStream is an independent, open access platform for sharing water quality data. It is free to use and designed to make it easy for diverse monitoring and research groups to share, visualize and download data. Our mission is to promote knowledge sharing and advance evidence-based, collaborative decision-making so our waters remain healthy for generations to come.

DataStream was developed to overcome persistent barriers to data sharing and reusability that hamper the use of this information to its full potential. DataStream does this by serving up machine readable, comparable data that is collected by diverse monitoring programs--from academic research projects to Indigenous-led monitoring initiatives--of varying scale, focus, and intent; and across multiple, sometimes overlapping, sectoral and jurisdictional boundaries. This work encourages collaboration and knowledge sharing among initiatives and lays the necessary foundation for the application of advanced modelling and machine learning to our freshwater challenges.

Leadership

DataStream is led at the national level by The Gordon Foundation and carried out in partnership with research and monitoring networks. This rapidly growing platform operates in three hub regions covering two thirds of Canada's landmass and, in 2021 will expand further with the launch of the Great Lakes hub.

Need for Domain-Specific Repositories

In scoping, building and scaling DataStream, The Gordon Foundation has developed a practical understanding of the barriers and opportunities that exist in the management of environmental data. These include complexity and confusion around reporting water data; inconsistent data formats and standards; lack of trust or competition among diverse actors; data security concerns; fragmentation, silos, and lack of coordination across monitoring programs and jurisdictions.

The resulting data management difficulties lead to major inefficiencies and severely hinder our ability to respond to unprecedented environmental change. DataStream and other domain specific repositories are part of the solution when it comes to managing environmental data but NDRIO has a critical role to play in this area.

Role for NDRIO

The following is a list of areas where NDRIO can be instrumental in improving access to environmental data in Canada. For all of the actions below, care is needed to ensure NDRIO engages a wide variety of organizations involved in research. This includes leaders of repositories that provide fundamental supports to the research community.

1. **Convener**
   a. **Potential actions:**
      i. Host forums that bring researchers, librarians, and existing leaders of repositories together to identify members of their respective research communities.
      ii. Forums could include discussions and consensus building in key areas that promote interoperability and (re)usability. This includes championing existing, well-established data standards both within Canada and other jurisdictions.

2. **Funder**
   a. **Potential Actions**
      i. Provide funding to institutions and researchers to ensure they have the technological capacity and training to format and share data
      ii. Provide funding for established repositories to scale their efforts, build data holdings and provide support and capacity building for their user communities (both data contributors and data users).
      iii. In evaluating which repositories to fund, priority should be given to those that engage data holders and researchers across institutions to encourage collaboration and avoid duplication of effort.
      iv. Identify areas where domain-specific repositories are needed and provide targeted resources to encourage their establishment. In these cases, consider providing both financial and technical support.

3. **Direct researchers to appropriate repositories**
   a. **Potential actions:**
      i. Create a minimum standard for repositories (possibly modelled after or in collaboration with Core Trust Seal) in Canada and provide a list of repositories that meet this standard.
      ii. Maintain a list of high-quality repositories broken up by domain (or lack thereof for domain agnostic repositories) that researchers and librarians can quickly reference to determine the best place to put their data.
      iii. Support dedicated Data Curators who work with researchers to identify the most appropriate repository for their data and provide the support they need to structure and openly share their data.

4. **Champion Existing Domain-specific Data Standards**

5. **Drive full implementation of open access policies**
   a. **Potential actions:**
      i. Work with major funders (NSERC, SSHRC, among others) to ensure funding for research is tied to the addition of data to accredited repositories.
      ii. Develop a toolkit for research funders looking to include open sharing requirements as a requirement of receiving funding. This could include:

suggested language to use in funding agreements; lists of appropriate repositories for a given field/data type; options around structuring funding with holdbacks that are released only once researchers provide a link to full, open datasets; basic definitions of what open sharing means… etc.

iii. Develop systems to track/audit whether requirements for open sharing are being met by researchers.

iv. Scope other mechanisms to incentivize researchers to meet their open sharing commitments.

Our main hope is that NDRIO will work not only to support the research and librarian communities in Canada, but also the ever-growing repository community as well. Further, we suggest that NDRIO's definition of "researcher" be broad enough to include those outside the academic community, for instance Indigenous-led initiatives and the not for profit sector. Through some of the roles listed above, we hope that NDRIO could help to provide some guidance, sustainability, and credibility to repositories in Canada and help ensure that researchers have the knowledge of and confidence in these repositories for storing and sharing their data.

Sincerely,

Carolyn DuBois

Director, The Gordon Foundation

Lindsay Day

Program Manager, The Gordon Foundation

Patrick LeClair

Data Specialist, The Gordon Foundation