

# **Data Valorization for Better Freshwater Management**

**Brief on Freshwater Submitted to the House of Commons Standing  
Committee on Environment and Sustainable Development**

**Presented by the St. Lawrence Global Observatory**

**Written by**

Gabriel Leblanc, Data Management Coordinator

Julien Boucher, Operations Manager

Ève Morin Desrosiers, Coordinator - Communications and Partnerships

## **Contact us**

**Email:** [info@ogsl.ca](mailto:info@ogsl.ca)

**Web:** [www.ogsl.ca](http://www.ogsl.ca)

## **Follow us**

[facebook.com/ogsl.slgo](https://facebook.com/ogsl.slgo)

[twitter.com/ogsl\\_slgo](https://twitter.com/ogsl_slgo)

# **Content**

- 1. SLGO: A recognized federal partner**
- 2. Issue 1: Obstacles to the dissemination of freshwater data**
- 3. Issue 2: Multiple thematic dissemination platforms: a barrier to data discovery and reuse**
- 4. Issue 3: Lack of standardization of freshwater data**
- 5. Presentation of SLGO and the CIOOS**
- 6. Recommendations**

# Introduction

This brief to the House of Commons Standing Committee on Environment and Sustainable Development offers some ideas and recommendations regarding freshwater data and the creation of the Canadian Water Agency. The perspectives in this brief support the vision of the St. Lawrence Global Observatory (SLGO). Our organization's mission is to disseminate freshwater and ocean water data. In this brief, we focused on data valorization, an expertise that we have been developing for fifteen years now.

The brief focuses primarily on three topics:

- (1) Collecting and disseminating freshwater data.
- (2) Improving freshwater data sharing.
- (3) Standardizing data.

A summary of recommendations is provided for each topic, with additional subsections to support those recommendations.

From the outset, SLGO answers some general questions related to the Canadian Water Agency and specifies the role that it wishes to play in the agency's first steps.

# SLGO: A recognized federal partner

SLGO has been working since 2005 to manage, disseminate and valorize data to make those data discoverable and support various stakeholders who care about protecting and managing freshwater resources. Over the years, we have established trusting and partnership-based relationships with several provincial and federal departments, such as Fisheries and Oceans Canada (DFO) and Environment and Climate Change Canada (ECCC). We help departments to organize their data, train them in good data management practices and disseminate various freshwater-related datasets, including with specialized visualization tools. Some examples of our partnerships for data dissemination can be found on page five. In addition, we have created partnerships with other provincial departments, universities, research centers, Indigenous communities, non-governmental organizations and conservation organizations. SLGO has successfully formed a vast network of data providers and also includes many members who support its mission.

## **SLGO is a choice partner for data management, dissemination and valorization.**

SLGO wishes to establish a partnership with the new Canadian Water Agency. Here is how we see our role:

- 1.** Play a formal advisory role to the Canadian Water Agency on data management, dissemination and valorization.
- 2.** Be identified as a partner in the dissemination, management and valorization of data with organizations receiving federal funding.
- 3.** Be recognized as the Canadian Water Agency's freshwater data dissemination platform for data on the St. Lawrence River's ecosystem and watersheds.
- 4.** Train Canadian Water Agency employees who work with data in good management practices.

## Our partners

### *Fisheries and Oceans Canada*

[Freshwater Aquatic Invasive Species Monitoring Program for the St. Lawrence](#)

[Freshwater runoffs of the St. Lawrence at the height of Quebec City](#)

### *Environment and Climate Change Canada*

[Weather buoys network](#)

## Our applications

[Biodiversity](#)

[Atlas of the St. Lawrence Lowlands](#)

# Freshwater data collection and dissemination

We estimate that a significant amount of freshwater data is being collected by scientists and organizations involved in preservation efforts. However, it is difficult to determine how much is being collected and whether that amount is enough. SLGO believes that the main challenge is in the later stages of the data life cycle, i.e., at the dissemination, archiving and valorisation steps. Our strategic position regarding data management allows us to note that gaps persist at these levels. The next sections of this brief propose some ideas and recommendations to overcome the issues that we see every day.

### **Data dissemination: An often overlooked step in the data life cycle**

As mentioned, data dissemination and archiving are typically added as additional steps to the long process of data collection and analysis. These final steps are intended to add value to the data in ways that enable discovery, encourage reuse and ensure sustainability. Nevertheless, these steps are sometimes neglected, particularly because they require time, human expertise, specific knowledge and financial resources. In concrete terms, these steps often call for reorganizing the data into formats for archiving, ensuring that the various entries meet internationally recognized standards, hosting the data files in open-access formats and publishing them on a dissemination platform, which are complex operations that require specific and evolving knowledge and skills.

## **Awareness and support as staples of data sharing**

Prior to dissemination, it is important to make data producers aware of the common benefits of sharing their data, namely, preventing duplication in gathering efforts, ensuring efficient spending of public finances and generating new knowledge from existing data (valorization). This awareness is necessary to build sustainable partnerships with data providers, since organizations are already willing to share their data publicly by design. Building relationships is even more complex since it also requires building trust. From our experience at SLGO, we know that building these relationships can sometimes take a few months and, more frequently, a few years. Developing tools that facilitate data valorization is certainly a step in the right direction, but ensuring partnerships and communication with data producers to acquire the data to feed these tools is equally important. This is one of the reasons why, throughout this report, we are insisting on the need to work with established industry stakeholders.

In addition, to ensure interoperability (an important concept defined under the FAIR principles), it is important to adjust data by giving them a common structure, namely, one that is governed by internationally recognized standards. When data is pooled from various sources, those data must be standardized and reorganized, which can be tedious and long. However, data dissemination for valorization requires a constant and informed liaison with data producers. In fact, data providers are generally unaware of the numerous standards and structures that must be applied to make data shareable and reusable by end users. This is nothing new, since dissemination and archiving require specific knowledge. In other words, data organization for analysis only partially meets the structural requirements of archiving and dissemination. We therefore believe that support from trained and knowledgeable data management experts is fundamental for data sharing. Data standardization and organization will be further discussed in the third section of the brief.

## **Reliance on formal sharing and dissemination agreements**

It is also important to define the framework within which owners and producers share their data on platforms for dissemination. The establishment of an administrative and legal framework makes it possible to reassure data owners, particularly with regard to intellectual property, and clearly sets out the limits of the dissemination platform's mandate. We recommend drafting formal sharing agreements that (1) specify the responsibilities and obligations of the parties involved, (2) guarantee the intellectual property of data-owning organizations, (3) authorize the platform to disseminate data that does not belong to it and (4) establish the scope and process of data dissemination. If these agreements are limited to a certain term, they should set out what happens to the disseminated data when they expire.

## **Formal incorporation of data management into funding programs**

The costs of developing and implementing data management plans should not only be eligible expenses for funding programs, but also mandatory criteria for selecting funded projects to promote the accessibility of data for scientific research communities and pave the way for data valorization. A data management plan is a detailed protocol that specifies how data that is collected in the field will eventually be valorized on a dissemination platform. This plan should not only allocate resources for data standardization and organization, but also identify where the data will be published. Since these steps of the data life cycle

present a certain level of complexity, we believe that it would be wise to have the support of competent organizations in this field.

## **Recommendations**

- 1.** Make federal funding for data collection projects conditional on the prior development of a data management and dissemination plan that includes a partnership agreement with a recognized data dissemination platform.
- 2.** Include existing data platforms and their contact information in funding application forms.
- 3.** Include the costs of developing and implementing data management plans as mandatory expenses in various funding programs.

# **Improving freshwater data sharing**

Our fifteen years of experience in data management expertise has given us a keen eye for current data management requirements. We believe that freshwater data sharing can be improved and that this can be achieved by supporting existing dissemination platforms that often have developed dissemination tools and sustainable partnerships with data producers. In addition, SLGO is of the opinion that the way to achieve improved data sharing should not be creating an abundance of new platforms, which are costly in terms of time and public finances. The following sections of this brief support our argument in favour of improved data sharing.

## **The challenge of feeding and sustaining data platforms**

Ensuring the sustainability of databases, applications, visualization tools and metadata catalogs involves significant maintenance costs. The above are supported by complex IT infrastructure that requires ongoing maintenance and incur various expenses (e.g., salaries paid to IT teams). We believe that adding more distribution platforms could exacerbate IT maintenance issues and prove to be counterproductive to improving data sharing.

Furthermore, operating and maintaining data storage and distribution infrastructure requires considerable resources. In addition to the human resources required for updating our industry's technological toolkit, the main issue that we are confronted with is the high cost of storing large amounts of data. Regardless of which storage option is chosen, they are all costed at several thousand dollars per terabyte of data per year. The issue of recurring funding for existing data hosting and dissemination platforms appears to be a determining factor in: (1) ensuring the continued availability of the hosted data inventory and (2) allowing existing platforms to continue to host new data, especially in large amounts.

From our point of view, the notion of sustainability has not yet been fully integrated into the funding mechanisms of existing programs for organizations that build and host dissemination tools. Furthermore, the data disseminated through these platforms



must be stored, which adds to the maintenance burden. If the goal is to provide a data repository, one factor worth considering is the sustainability of hosting it in the systems, which results in significant data management expenses. Hosted data accumulates over time, and increased storage space is a constant requirement.

Our understanding of the complexity of data management leads us to argue that the development of existing data archiving and dissemination systems should take precedence over the construction of new platforms. For example, SLGO has powerful tools that are capable of responding to immediate needs, but we face constraints related to data storage. Requests to host and distribute data in video format, for example, sometimes have to be declined due to storage issues, even though our tools to distribute this type of data are available and fully functional. We believe that public resources should be allocated in consideration of advanced data management protocols and therefore go toward ensuring that established systems are improved.

#### **Promotion of existing platforms to improve data sharing**

SLGO believes that improving freshwater data sharing goes hand in hand with enhancing existing dissemination platforms, including through the allocation of additional financial resources to ensure the development and maintenance of an IT infrastructure that is capable of meeting the demand for data hosting and dissemination. However, ensuring greater visibility for platforms with a goal of increasing contributions from data producers remains a factor.

As previously mentioned, those platforms' success depends on making partnerships with data providers to obtain the necessary data. Therefore, we believe that efforts should be made with data producers to raise awareness of the importance of sharing and pooling knowledge and data. These efforts should be supported by previously established industry stakeholders, who have already built partnerships and whose efforts to increase their reach are currently under way.

We also believe that it is important to provide technical assistance to data providers, especially since locally hosted databases rarely meet the rigorous requirements of proper data management for dissemination and archiving. Mastering the many standards and structures that promote data interoperability and reusability requires special knowledge. We therefore believe that improving the sharing of freshwater data should build on expertise that has already been developed by established industry stakeholders in data dissemination and use.

#### **Recommendations**

- 1.** Financially support existing data platforms' outreach efforts to find new data dissemination partners.
- 2.** Provide recurring financial support to existing thematic data platforms to guarantee sustainability.
- 3.** Make data producers aware of the benefits of sharing and valorizing data.

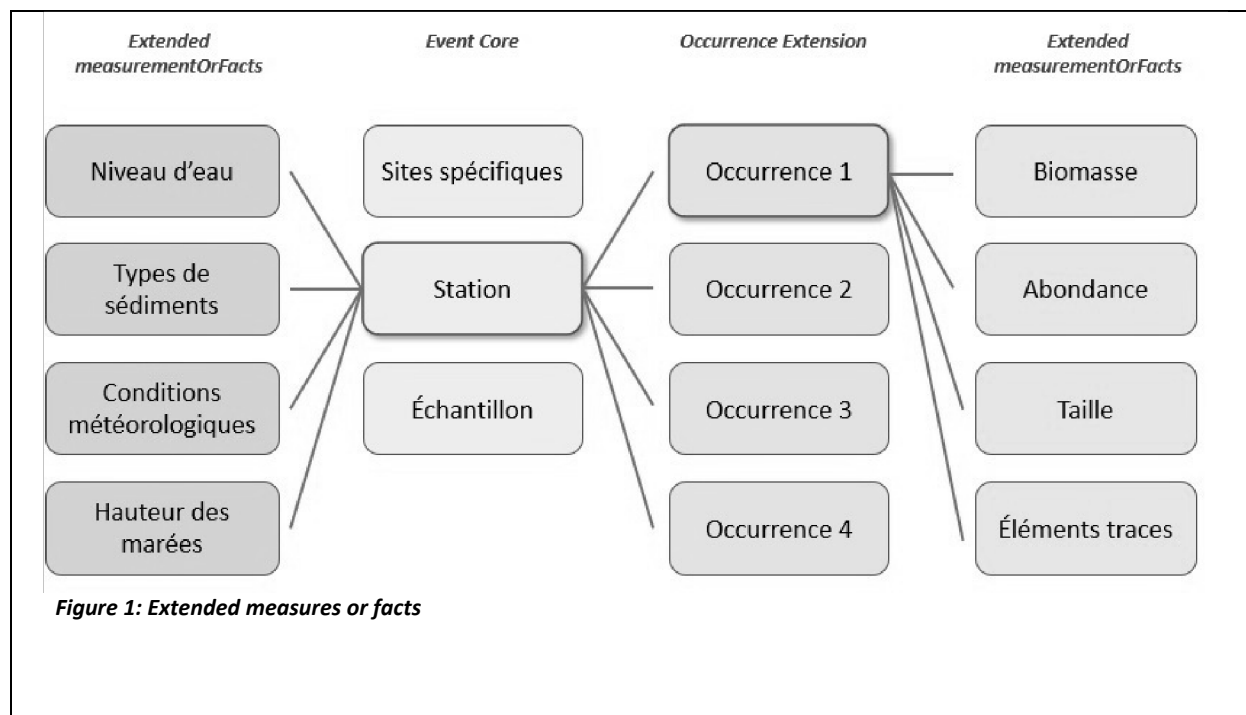


# Data standardization and organization

Data received at SLGO is generally sent to us using formats, structures and standards that meet the needs of data producers. However, while these files are usable by some people (for example, the team that shared them with us), they usually do not meet the conditions for immediate dissemination. In addition, shared data needs to be standardized and reorganized, which can take from a few days to a few weeks.

There is a difference between data that was organized by researchers, for example, and data that is standardized for dissemination: while some structures allow producers to handle data, they generally do not meet international requirements for archiving and dissemination, particularly because they do not allow for reuse and interoperability. For example, data that describe the abundance of animal populations will sometimes tag species names by representing them as variables, whereas the standard requires that the scientific names of observed taxa be noted as values. Standardizing the data in these cases may require programming work: a skill that is sometimes not present in data-producing organizations. Another good example is the frequent use of common names to tag observed species. This practice poses challenges in terms of data sustainability, since the common names of species sometimes tend to change depending on cultural affiliations, where the data was produced and when it was collected. Furthermore, common names do not allow end users to ensure that the species in the database are actually the ones that are being studied. Standard practices exist to overcome this issue, and a data management team is often (if not always) needed to guarantee proper archiving and dissemination.

That is why we have a data integration process at SLGO that involves a great deal of expertise, including from data coordinators, data entry specialists and IT operators.



<b>French</b>	<b>English</b>
Niveau d'eau	Water level
Types de sédiments	Types of sediments
Conditions météorologiques	Weather conditions
Hauteur des marées	Tidal height
Sites spécifiques	Specific sites
Station	Station
Échantillon	Sample
Occurrence 1-4	Occurrence 1–4
Biomasse	Biomass
Abondance	Abundance
Taille	Size
Éléments traces	Trace elements

### **Purpose of data management standards**

The end goal is to provide a standardized format to facilitate integration into different platforms. Once the standardization is complete, it is then possible to create dissemination interfaces that meet specific needs. In other words, visualization tools (such as an interactive application reporting on the georeferenced distribution of certain species) require databases. If new data are to be integrated into these databases, they must use a standard structure; otherwise, new programming code will have to be created for each new file in order to meet its specificities. That is why standardized structures promote the interoperability and, eventually, the reusability of the data, which justifies the need for good data management and specialized assistance.

### **Data standardization as a prerequisite**

We believe that it is necessary to collectively follow internationally recognized data standards and structures. Some organizations, such as the Ocean Biodiversity Information System (OBIS) and the Darwin Core (DwC), have been developing a set of standards for years now to govern the formats for archiving and disseminating biodiversity data. In fact, these two entities have been working with the international scientific community since the early 2000s to propose a data structure that meets ongoing needs and issues.

For example, OBIS provides a structure for linking occurrence data (distribution of species and different taxa) with environmental data, i.e., data that characterize the locations where specimen observations were made. OBIS also links different types of measurements (biotic and abiotic) to occurrences and events by using unique identifiers and Extended Measurements or Facts (EMoF) principles.

From SLGO's perspective, this kind of structure, which is governed by specific standards, facilitates the development of visualization tools that meet the needs identified by data users. Having a standardized data structure makes it easier to create automated integration systems from files to databases that feed applications. In this sense, this standardization meets the notion of interoperability as identified by the FAIR principles.

## **Recommendations**

1. Foster co-operation between data producers and data management organizations to reorganize data into standardized structures that meet archival and dissemination needs.
2. Follow existing standards to ensure that data structures meet the guiding principles of interoperability and reusability.

# Recommendations

1. Make federal funding for data collection projects conditional on the prior development of a data management and dissemination plan that includes a partnership agreement with a recognized data dissemination platform.
2. Include existing data platforms and their contact information in funding application forms.
3. Include the costs of developing and implementing data management plans as mandatory expenses in various funding programs.
4. Financially support existing data platforms' outreach efforts to find new data dissemination partners.
5. Provide recurring financial support to existing thematic data platforms to guarantee sustainability.
6. Make data producers aware of the benefits of sharing and valorizing data.
7. Foster co-operation between data producers and data management organizations to reorganize data into standardized structures that meet archival and dissemination needs.
8. Follow existing standards to ensure that data structures meet the guiding principles of interoperability and reusability.

## About SLGO

The St. Lawrence Global Observatory (SLGO) is a non-profit, non-partisan organization that was created in 2005 through the collaboration of several federal/provincial, academic and community organizations.

Its purpose is to maximize the benefits of data collection by integrating multidisciplinary data from many different partners into a single web portal and promoting the discovery and valorization of data on the St. Lawrence ecosystem, from the Great Lakes to the Gulf. SLGO currently has over 60 members who share their data or support its mission and enjoys international recognition. [www.ogsl.ca](http://www.ogsl.ca)

## About the CIOOS

SLGO is a founding member of the Canadian Integrated Ocean Observing System (CIOOS), which is a national and international initiative developed in partnership with other integrated systems (e.g., IOOS in the U.S.).

This initiative provides Canada, like many other countries, with an integrated ocean observatory system. The CIOOS is a consortium of partners spread out among three regions—Pacific, St. Lawrence and Atlantic—who work with regional data providers and users to make these data accessible and available on their portals.

**[www.siooc.ca](http://www.siooc.ca)**