

A Guide to Coastal and Ocean Data Management

Best Practices and Training for Environmental Non-Governmental Organizations



COINAtlantic
Dalhousie University
6414 Coburg Road
PO Box 15000
Halifax, NS B3H 4R2
coinatlantic@dal.ca

1.0 Introduction

Open data is an important resource that assists stakeholders in adapting to changes in the marine environment and achieving shared goals in ocean science, through evidence-based decision making. However without a plan for data management, information can naturally degrade over time. This can lead to lost data and contribute to knowledge gaps. As such, proper data management is essential when conducting research and other projects. A set management process for data collection, analysis, storage, and sharing mitigates the risk of lost data, and supports data reliability for data users; ensuring maximum data value over both a short and long period of time. By adhering to internationally recognized best practices for data management, data-sets can be made more accessible for everyone.

1.1 COINAtlantic

Coastal and Ocean Information Network Atlantic (COINAtlantic) is a non-governmental organization based in Halifax, Nova Scotia. COINAtlantic is dedicated to supporting coastal and Marine Spatial Planning (MSP) and decision-making through collaboration, open data management, and information sharing. COINAtlantic's mandate is to promote, facilitate and influence information management, policies, and programs that enhance integrated coastal and ocean management in Atlantic Canada.

1.2 Purpose and Objectives

COINAtlantic has observed a range of capabilities when it comes to data management practices within environmental non-governmental organizations (ENGOS). Despite data management becoming more commonplace, many ENGOS are not given sufficient guidance and training on creating a proper data management strategy. External efforts to assist ENGOS are limited to consultations, scholarly resources, and the development of data tools and programs. As such, COINAtlantic is engaging in a project aimed at empowering ENGOS to help increase and improve organizational data management capacity for projects in marine and coastal environments. This project is being conducted in collaboration with the Marine Environmental Quality (MEQ) Program and the Clean Foundation.

This report acts as a Data Management Best Practices Guide for ENGOS by providing information assembled from COINAtlantic's Coastal and Ocean Data Management Best Practices Webinar Series, and existing data management tools. Following consultation from the ENGO community, the Guide is tailored to address knowledge gaps and support the development and implementation of ocean conservation and management activities. Here, three resources are presented that can assist ENGOS in the creation of an effective plan for managing research data.

2.0 Community Survey and Interviews

In Spring 2021 during the preliminary research stage of the project, an online survey was circulated to ensure the best practices guide responds to the needs of the ENGO community. Survey participants included maritime-based ENGOs, and Department of Fisheries and Oceans (DFO) employees working with data related to marine environmental quality. The survey results gathered insight on existing knowledge of data management practices and capacity, identified areas of least proficiency and needs going forward, and assessed the level of interest in training material that COINAtlantic can develop.

In Spring 2022, one-on-one follow-up interviews were conducted with select participants to provide an in-depth explanation on survey responses. Interview participants were selected based on their responses to collect a range of perspectives. Of the seven survey participants contacted, three were available for an interview. Interviews were semi-structured in nature to guide discussion while allowing for free flowing conversation. Questions were predetermined and tailored to gather a better understanding of the participant organizations' data management capacity and needs for training materials. Interviews were conducted using the video conference platform, Zoom.

Results of the surveys and interviews can be found in the project final report.

3.0 Data Management Plans

A data management plan (DMP) is a written document that describes how the researcher(s) intend to manage data during and after a research project. There is growing recognition among decision-makers for DMPs in research projects. As a result, many funders and publishers are starting to require DMPs to be included as part of a project proposal. Detailed DMPs typically include all stages of the data life cycle – data collection, data analysis, data storage, and data sharing. DMPs increase data impact and visibility, preserve data long-term, and safeguard data from loss. Data management is best when addressed early; however it is never too late to develop a DMP for a single project or an entire organization. A number of data management resources exist to assist in the creation of DMPs, most notably the [DMP Assistant](#) provided by the Portage Network.

3.0.1 Data Collection

Data management is nothing without data collection. Employing the use of established standards and guidelines for data collection can assist in producing high quality data and reduce the likelihood of errors. The goal for all data collection is to gather sufficient information to address a research question. Reliable and up-to-date data can provide valuable information for regional, national, and global research efforts to fill knowledge gaps and develop accurate theories.

3.0.2 Data Analysis

Data analysis is the process of interpreting huge amounts of data so that researchers can discover meaningful insights, make informed conclusions, and support decision-making. Analyzed data is far more useful than raw data and enhances the significance of the research. Once data is analyzed, it is important to consider the format(s) for reporting the data. Data visualizations such as graphs, tables, and charts, are a valuable tool to help communicate the key messages of the data.

3.0.3 Data Storage

Effective data storage management is more important than ever as COVID-19 increased remote work. It is important to consider data storage requirements for all data forms, such as raw and final data, gathered throughout the project timeline. With proper storage and backup, data can be available to researchers regardless of time or location. It also ensures that the data is protected from loss via human error, system failures or outside threats. Over time, data storage can provide increasingly valuable historical data.

3.0.4 Data Sharing

Making data discoverable by publishing in data portals is an essential part of data management best practices. Data sharing can greatly increase the value of the data set by encouraging connection and collaboration between researchers and stakeholders. It incentivizes researchers to use data management best practices to ensure high quality data, and produce high quality research. It also increases the researchers credibility. Data sharing reduces the likelihood of repeating research studies, thereby saving time, money, and unnecessarily duplicating the data. Data sharing can increase recognition, stimulate interest, and provide direction for future studies, as researchers are able to build upon the work of others to support new findings. Ultimately, widely spread information can assist in informed decision making for environmental planning and policy, and create a sustainable community.

3.0.5 Metadata

Metadata is information about the dataset - for example where, when, and how the data was collected. Metadata provides context and description to the dataset. It is an essential detail to record in a DMP as it helps future researchers determine if the data is relevant for their work and makes the dataset interoperable with other databases both nationally and internationally. Good metadata ensures researchers will be able find data, preserve data, and re-use data in the future.

3.1 Creating a Data Management Plan

Before creating a DMP, the research team should consider the following general guidelines. Begin by outlining basic identification and logistics for the research project. This includes the project title, contact information, and the anticipated start and end dates of the project. The research team, and any other partners involved with the project need to be identified. All members of the research team should have access to the DMP so everyone may contribute to it. Lastly, any resources that are needed to implement the DMP should be outlined, including an estimated budget. The length of a DMP is dependent on the research project; however typically DMPs are 2-3 pages. It is strongly recommended that the DMP be a living document, so it can be edited consistently throughout the project as needed.

The following table outlines the section and contents of a complete DMP.

Section	Contents
Data Collection	<ul style="list-style-type: none"> ● Specify the source(s) and type(s) of the data. ● Describe the method of data collection. ● Provide an estimate of the amount of data to be collected. ● Identify any software or platforms being used for data collection and processing. ● Outline the file formats for data organization. ● Determine what procedures are needed to structure, name, and control the data files to help future researchers better understand the data.
Documentation and Metadata	<ul style="list-style-type: none"> ● List any documents and/or outline information needed for the data to be read and interpreted correctly. ● Choose a metadata standard suitable for your organization. ● Specify a required data citation for acknowledgement purposes. ● Describe how documentation will be recorded throughout the project.
Storage and Back-up	<ul style="list-style-type: none"> ● Identify storage space requirements (i.e. gigabytes, megabytes, etc.) and the length of time the data needs to be stored. ● Outline a schedule for backing-up data. ● Consider using preservation-friendly file formats, anonymization, and inclusion of any supporting documents ● Describe where the data will be stored
Sharing and reuse	<ul style="list-style-type: none"> ● Outline the form in which the data will be shared. Data can be raw, processed, analyzed, or in its final form, and shared through data catalogs or on social media. ● Plan for what type of end-use licenses should be used when data is replicated. ● Identify a data repository for sharing data.

Responsibility and Resource	<ul style="list-style-type: none"> ● Identify an individual who will be responsible for data management during and after the project. ● Identify a second individual who can take-over any data management tasks in the event the primary data manager cannot complete their tasks.
Ethics and Legal Compliance	<ul style="list-style-type: none"> ● Outline any restrictions on the data. ● Describe how sensitive data is securely managed. ● List any legal, ethical, and intellectual property issues.

3.2 Data Management Principles

A key component to a DMP is the comprehensive description of the data and contextual information so that future researchers can discover, use, and interpret the data after time has passed. This is achieved by adhering to the [FAIR Guiding Principles](#) for scientific data management and stewardship. The FAIR principles provide guidance for ensuring data is Findable, Accessible, Interoperable, and Re-usable. Data should be easily findable by having it indexed in a searchable resource. Once data is found, it should be accessible following usage licenses. The data should be interoperable so it may be integrated with different applications and repositories. Following these practices supports the ultimate goal of optimizing the reuse of the data.

The FAIR principles are data specific, and thus do not fully engage with Indigenous issues. As such, the [CARE Principles](#) for Indigenous data governance were developed to complement the FAIR principles. The CARE principles provide guidance for the Collective benefit, Authority to control, Responsibility, and Ethics of the data. The CARE principles state that data sets should be designed and function such that there is collective benefit for all Indigenous peoples involved in the research. Indigenous people’s interest must be recognized, and their authority to control Indigenous data must be acknowledged. Researchers working with Indigenous data have the responsibility to share how the data is being used to support Indigenous peoples. Lastly, Indigenous peoples rights and well-being should be the primary concern at all stages of the data life cycle.

4.0 Coastal and Ocean Data Management Best Practices Webinar Series

COINAtlantic hosted a series of information/training webinars aimed at explaining open data sharing platforms and data management plans. The Webinar Series was based on the progression of topics covered in this Best Practices Guide and the results of the pre-survey. Training sessions included an interactive component to walk attendees through concrete steps to implement and/or augment data management practices within the participants’ own organization. COINAtlantic collaborated with subject matter experts to ensure that the appropriate material is covered in each virtual training session.

Webinar 1: CIOOSAtlantic - Data Discovery and Contribution

As Canada's nucleus for ocean observing, the Canadian Integrated Ocean Observing System (CIOOS) makes connections for a sustainable ocean future by fostering partnerships and growing a powerful online platform. Through the nation-wide open-access Data Catalogue, CIOOS Atlantic generates information, knowledge, and place-based solutions to advance our understanding of the ocean. CIOOS Atlantic uses Essential Ocean Variables aligned with the Global Ocean Observing System (GOOS) to help improve the monitoring of key ocean processes. You can contribute your data to CIOOS Atlantic by contacting info@cioosatantic.ca or submitting a form to the Metadata Entry Tool.

Watch the recording.

Webinar 2: OBIS - From Nature to Numbers

The Ocean Biodiversity Information System (OBIS) is a global open-access data and information clearing-house on marine biodiversity for science, conservation, and sustainable development. OBIS works to build and maintain a global alliance that collaborates with scientific communities to facilitate free and open access to ocean biological data. The growing network currently houses over 4000 data sets, and 79,000,000 records, with contributions from 56 countries. You can contribute to OBIS through OBIS Canada.

Watch the recording.

Webinar 3: Digital Research Alliance of Canada - Research Data Management from Plan to Preservation

The Digital Research Alliance of Canada (the Alliance) is a national non-for-profit with the objective of advancing Canada's position as a global leader in research data. By integrating, championing and funding the infrastructure and activities required for advanced research computing, research data management, and research software, the Alliance provides the platform for the research community to access data tools and services faster than ever before.

Watch the recording.

5.0 Resource Index

This index is a collection of data management resources, from training materials to repositories.

Atlantic DataStream

The Atlantic DataStream is an open access hub for sharing data on freshwater. It allows users to access, visualize, and download full water quality data sets collected from New Brunswick, Newfoundland and Labrador, Nova Scotia, and Prince Edward Island. The Atlantic DataStream promotes knowledge sharing across watersheds and advances collaborative decision making.

DataONE Best Practices database

DataONE is a community driven program providing access to data across multiple member repositories. DataONE provides recommendations and promotes best practices on how to effectively work with data management through educational resources and materials.

Dataverse

The Dataverse Project is an open source web application to share, preserve, cite, explore, and analyze research data. It facilitates making data available to others, and allows you to replicate others' work more easily. Researchers, journals, data authors, publishers, data distributors, and affiliated institutions all receive academic credit and web visibility.

Federated Research Data Repository

The Federated Research Data Repository (FRDR) addresses a longstanding gap in Canada's research infrastructure by providing a single platform from which research data can be ingested, curated, preserved, discovered, cited and shared. Anyone can use FRDR to search for and download data across Canadian repositories and Canadian post-secondary institutions may use FRDR to publish their data.

Ocean Best Practices

The Ocean Best Practices System (OBPS) is a global, sustained system comprising technological solutions and community strategies to improve data management and support the development of ocean best practices. OBPS provides access to relevant and tested methods from observation to application so that the ocean community may share strategies and develop best practices. OBPS hosts a secure private document repository containing community accepted existing ocean best practices. Find a best practice that works for you? Submit it to the Ocean Best Practices Repository.

Ocean Exploration and Research Data Management

NOAA's Office of Ocean Exploration and Research (OER) performs data management for projects to provide rapid and easy data access, facilitate efficient and effective long-term data

preservation, and inspire further ocean exploration and research. OER ensures that ocean data collections are FAIR - findable, accessible, interoperable, and reusable, to the public.

Ocean Networks Canada

Ocean Networks Canada (ONC) delivers continuous data in real-time for scientific research that assists communities, government, and industry in decision making on ocean management, disaster mitigation, and environmental protection. ONC is based out of the University of Victoria.

OceanTeacher

OceanTeacher provides a comprehensive web-based training platform that offers classroom, online, and blended learning. Courses cover a wide range of topics related to the International Oceanographic Commission programs. The Global Academy has established a network of training centers to deliver customized training for ocean experts and professionals to increase national and regional capacity in coastal and marine sciences, services, and management.

Ocean Tracking Network

The Ocean Tracking Network (OTN) is a global aquatic animal tracking, technology, data management, and partnership platform based at Dalhousie University. OTN informs sustainable management and stewardship of aquatic animals by providing knowledge on the scientific foundation via data warehousing and technological innovation. OTN unites marine scientists to better understand and manage pressing global ocean concerns in the face of climate change.

Portage Network and Curation Expert Group

The Portage Network is dedicated to shared stewardship of research data in Canada through developing a national research data culture, fostering a community of practice for research data, and building national research data services and infrastructure. Portage works with stakeholders to develop tools and resources to help with data management planning, access, preservation, and discovery so that all researchers in Canada have the support they need. The Curation Expert Group involves the process, practices, techniques, and tools used to improve the quality and best practices of data and metadata to facilitate dissemination and reuse with other data sources. You can learn how to write a DMP using the Portage [DMP Assistant](#).

Research Data Canada Best Practices for Scholar Metadata

Research Data Canada (RDC) works with organizations to ensure research data is properly managed in an infrastructure that provides long-term preservation and access. The RCD facilitates communication and partnership between organizations, promotes education and training, and highlights knowledge gaps. The RCD also acts as a point of contact for international initiatives and organizations. The RCD recommends best practices with respect to standards and interoperability of research data in Canada.