# Data Management Policy

August 30, 2017

## Contents

# 1.0 Introduction

The Marine Environmental Observation Prediction and Response (MEOPAR) Network is a Networks of Centres of Excellence of Canada (NCE) network that facilitates partnerships among academia, including natural, human health, and social scientists; government; and the private sector, including NGOs and not-for-profit organizations. The amounts and types of data, information, and knowledge collected and generated by MEOPAR-funded projects vary widely, yet must be managed consistently and effectively to ensure that research funding is being used efficiently and that data collected by MEOPAR researchers are widely accessible to benefit Canada and Canadians: supporting scientific excellence, foster collaboration and innovation, and harness ocean data to inform decision-makers and other stakeholders.

The function of this document is to provide a network-wide data management policy (hereafter, "the Policy") that outlines the objectives, principles, and guidelines for the management, retention, use, and dissemination of data generated and collected by MEOPAR-funded projects. The principles expressed here should inform the development of data management plans by projects. The Policy was created based on current best practices in research data management and in consultation with the Tri-Agency Statement of Principles on Digital Data Management[1].

# 2.0 Objectives

The objectives of the Policy are to:

- Realize the value of data collected and generated through MEOPAR-funded projects by encouraging the re-use of this data to inform additional insights and discoveries, for the benefit of society and research communities.
- Recognize and acknowledge the contributions represented by sharing data by detailing the guidelines for the use of data from MEOPAR-funded projects, including citation, attribution, and acknowledgement requirements.

[1] Government of Canada. (2015). Draft Tri-Agency statement of principles on digital data management. Retrieved from http://www.science.gc.ca/default.asp?lang=En&n=83F7624E-1

- Improve the efficiency of research by ensuring data are easily discoverable and ultimately accessible.

- Foster collaboration among MEOPAR-funded researchers, and the general research community, through clear mechanisms and mandates for data sharing.

- Encourage responsible data sharing by providing guidance to MEOPAR researchers working with sensitive data.

## 3.0 Principles

The overarching principle guiding this Policy is a view of publicly-funded research data as a public good that should be as open as possible to facilitate re-use, while also respecting the rights and responsibilities of the data originator and ensuring that time and energy spent collecting and sharing data is acknowledged and rewarded. Sub-principles extending from this vision are as follows:

- Data cannot be accessed or re-used if no one knows it exists. Because the potential users and beneficiaries of data cannot be identified in advance, knowledge of the existence of data must be public. The nature and type of data collected, including as much metadata as is feasible, should be publicly available. To avoid distributing this information across a variety of websites and portals, where available centralized or aggregated indexes and catalogs should be used to foster awareness of data sets. Data must be made available immediately, though restrictions may be placed on access to the data to ensure researchers or students benefit from their data collection/generation efforts, or to accommodate special considerations including requirements of the partner(s). Access to data must not be unreasonably restricted. Special justification will be required by RMC for cases were data are not used within two (2) years.

- Data collected and/or generated from MEOPAR-funded projects are citeable and publishable, are acknowledged as a valuable contribution to scholarly work and knowledge dissemination, should be included in annual and semi-annual reporting, and will be acknowledged and rewarded.as appropriate.

- Principal Investigators (PIs) retain the right to be the initial beneficiary of data collection, but are expected to: 1) manage data effectively throughout the course of the project; 2) respond to requests for data and not unreasonably restrict access to data; and, 3) ensure that data are made openly available long-term.

- MEOPAR researchers are expected to work with international and national infrastructure for data sharing as it becomes available.

- MEOPAR is committed to connecting centres of excellence including through data sharing within our network, and as a federally-funded network to sharing data to advance Canadian research. While there is special interest in ensuring data are available to these groups, research is a global endeavour, and access to data is not limited to these groups.

- MEOPAR researchers need the capacity, skills, tools, and mandate to manage data effectively. To this end, MEOPAR's Corporate Centre may require Data Management Plans (DMP), may provide DMP templates, and may provide training on data management. Investigators are responsible for ensuring that data are managed in accordance with these policies, and that data management plans and processes are documented.
- Model-generated output is valuable data that is often produced in massive amount and thus difficult to store and make publicly available. Model code used to produce datasets can also evolve as scientific and methodological advances are integrated. In cases where long-term storage of model-generated output can't be maintained, the properly documented model code used to produce data, together with initial and boundary conditions, should be stored and shared as described in this policy, in order to assure reproducibility of results. In cases where reproducibility can't be guaranteed, long-term storage of results and open access is expected.

# 4.0 Definitions

MEOPAR Data (hereafter, "data"): any and all data that have been collected and/or generated by MEOPAR researchers in the performance of MEOPAR-funded projects and sub-projects, and includes the metadata associated with that data.

MEOPAR Researchers (hereafter, "researchers"): Investigators and highly qualified personnel (HQP; including students, RAs, technicians, and postdocs) working on projects funded in whole or in part by MEOPAR.

# 5.0 Data Quality and Metadata Standards

MEOPAR funds ocean research across a variety of disciplines, including natural, human health, and social sciences. Data covered by this Policy are therefore highly diverse, existing in many different formats with file sizes and management requirements ranging

widely. Researchers must therefore adhere to the data management best practices and generally accepted metadata standards used by their discipline or field. Metadata should include, but not be limited to, records related to the collection, storage, and retrieval of data, as well as steps taken to process, analyze, and visualize data. At minimum, data must include clear supporting documentation and metadata (and, where possible and applicable, computer code) sufficient for re-use and replication of results by other researchers.

## 6.0 Data Retention and Preservation

Researchers are not expected to deposit their data with a centralized MEOPAR repository. NCEs are not structured for the long-term preservation of data; if the MEOPAR Administrative Centre, or any MEOPAR projects or cores, provide data storage, this should be viewed as a temporary solution. Individual projects will be responsible for the identification of appropriate long-term repositories; where possible, national or international repositories; alternatively, institutional or discipline-specific repositories. Plans for data storage (short-term) and retention (long-term) should be considered during the early stages of project planning and be implemented through all stages of the project lifecycle.

## 7.0 Data Access and Sharing

In accordance with the data management principles, data should be easily discoverable (external stakeholders should be able to determine what data are collected or generated, and how to access it) and ultimately accessible (whenever possible, access to data should be provided with no, or with the fewest feasible, restrictions).

Researchers have the right to benefit from the data they collect and generate, and the responsibility for ensuring that data is **discoverable** and **accessible**.

**Discoverable**. Data must be made discoverable as soon as is practical after collection or generation. It can be made discoverable by ensuring that metadata are published in an appropriate international, national, institutional, or subject-specific catalogue during the early stages of a project's lifecycle, or listed in a central, publicly accessible index. Alternative mechanisms for making data discoverable are subject to review by the RMC.

**Publicly Accessible Data**. Data must be made publicly accessible as soon as is practical after collection or generation (for data where this is not possible, see "Not Publicly Accessible

Data"). Data can be made accessible by making the data available through an international, national, institutional, or subject-specific repository, or in general by documenting in the appropriate index (above) a mechanism for accessing data. Reasonable steps should be taken by those using the data to inform the data originator about prospective use, and all data users must provide appropriate citation, acknowledgement, or other appropriate attribution (when applicable, adhering to the request of the originator) in any publications, presentations, or products arising from the use of said data.

Not Publicly Accessible Data. The use of public repositories may conflict with a researcher's right to benefit from the data they collect and generate, or with special considerations regarding data (Section 8.0).

- To protect the rights of researchers, data may be stored privately for a time period of limited duration, and individuals wishing to access the data may contact the appropriate researchers to request access. The duration of private storage should be the shortest period possible, should not exceed the life of the project, and in any case must be no longer than the time required for students to complete and defend their theses or dissertations.
- To ensure special considerations (Section 8.0) are met, data may be stored privately indefinitely, and individuals wishing to access the data may contact the appropriate researchers to request access.

Access requests. A request for access should identify the intended use of the data, how the data will be handled and shared (if applicable), and how the data will be cited, acknowledged, or otherwise attributed. *Such requests must be responded to and must not be unreasonably denied.* It may be necessary to deny, partially deny, or negotiate requests if special considerations apply (Section 8.0). In some cases negotiation may be required. Those requesting access must be willing to consider working in collaboration with the data originator for their envisioned new use of the data. At minimum, citation, acknowledgement, or other attribution for the data originator is expected in all resulting publications, presentations, or products generated from use of said data;

Appeals. Requestors who believe a researcher has unreasonably denied access to data, or not made data discoverable or accessible, may request a review. In the *informal* process, the MEOPAR Executive Director (or designate) will facilitate a conversation with the researcher and the requestor in an effort to reach an outcome consistent with this Policy. In the *formal* process, the requestor may submit a request in writing to the MEOPAR Research Management Committee (RMC). The RMC, or an appropriate sub-committee, will review

the request and will be responsible for deciding the outcome, including any implications to project funding or renewal.

# 8.0 Special Considerations for Data Sharing

## 8.1 Research Overlap

A researcher may determine if a request for data, and the intended use identified, significantly overlaps with their own work, or would jeopardize the work of a student. It is important that students' work, including theses and dissertations, be protected. In such cases, data may be justifiably withheld until such a time that its use does not jeopardize their own work or that of a student. However, it must also be considered that many types of research can be conducted in parallel using the same set of data without overlap. The determination of overlap is subject to the appeals process described in Section 7.0.

## 8.2 Data of a Sensitive Nature

Certain types of data (e.g., data containing personally identifiable information [PII], Traditional Knowledge, or data related to commercially valuable or endangered species) may be considered sensitive, with the release of such data resulting in potential harms. Where data released may cause harm, specific aspects may need to be omitted (e.g., the locations and nests of endangered species or locations of sacred sites[2]) (Shearer, 2015). In addition, the rights and privacy of individuals must be protected at all times. Any data made available publically must therefore be free of PII and other variables that could lead to the deductive disclosure of the identity of individual subjects[2]. This can be achieved through the use of standard anonymization and access restriction procedures already described in the Tri-Council Policy Statement on ethical conduct for research involving humans (TCPS 2)[3], and various institutional ethics review processes. Any research and its corresponding data involving human subjects must conform to TCPS 2 guidelines.

Research projects working with Traditional Knowledge must not compromise the rights of the knowledge holders. Research involving Traditional or Indigenous knowledge must

[2] Shearer, K. (2015). *Comprehensive brief on research data management policies*. Available: http://www.science.gc.ca/1E116DB8-E7F3-4B6F-BB44-83342BAAA030/Comprehensive%20Brief%20on%20Research%20Data%20Management%20Policies.pdf

[3] Government of Canada. (2014). *Tri-Council policy statement: Ethical conduct for research involving humans*. Available: http://www.pre.ethics.gc.ca/pdf/eng/tcps2-2014/TCPS_2_FINAL_Web.pdf

adhere to the concepts of respect, reciprocity, and responsibility[4]. This includes the "appropriate engagement of Indigenous People, communities or organizations throughout the entire data life cycle, formal attribution of contributed knowledge, establishment of informed consent for use of knowledge and derived products, and the maintenance of contributor control of data and information resources"[5]. Restrictions to data access may be made on the basis of ethical concerns arising from violating the rights of Traditional Knowledge holders.

## 8.3 Intellectual Property Rights

The Intellectual Property Rights of data are defined by the terms of the MEOPAR Research Funding Agreement. Data made available to researchers by third parties is not subject to data sharing requirements. Researchers may not enter agreements with a third-party that would restrict (due to commercial or proprietary considerations) the use or sharing of data collected or generated by researchers, without approval from the RMC.

## 8.4 Model Code and Model-Generated Output

It is understood that not all model-generated output is suitable for preservation and sharing, and in some cases only subsets of model results can be practically stored and made accessible. Model-generated output should be considered as data to be managed and shared as described in this policy in cases where research outcomes and/or impact outcomes involve research partners and/or end-users. Otherwise, the reproducibility principle for model-generated output (Section 3) must minimally be adopted and implemented.

[4] International Arctic Science Committee (IASC). (2013). The state of principles and practices for Arctic data management. Available: http://iasc.info/images/data/IASC_data_statement.pdf
[5] Ibid., page 3.

## For Information

Other policies consulted when authoring this document:

ArcticNet. (2011). *Data policy*. https://www.polardata.ca/pdcinput/public/data-policy.pdf

Biological and Chemical Oceanography Data Management Office (BCO-DMO). (2008). *BCO-DMO Data Management Guidelines Manual*. http://www.bco-dmo.org/files/bcodmo/BCO-DMO_Guidelines.pdf

Canadian Healthy Oceans Network (CHONe). (2013). *Data policy*. http://chone.marinebiodiversity.ca/data/data-policy/chone-data-policy

Jones, S. (2009). A report on the range of policies requires for and related to digital curation. *Digital Curation Centre, University of Glasgow*. http://www.dcc.ac.uk/sites/default/files/documents/reports/DCC_Curation_Policies_Report.pdf

Research Councils UK (RCUK). (2015). *Guidance on best practice in the management of research data*. http://www.rcuk.ac.uk/documents/documents/rcukcommonprinciplesondatapolicy-pdf/