# Data collection, sharing, and hosting protocols (ToR 4)

A framework of data collection, sharing and hosting protocols is needed to ensure data collected during the mapping and monitoring phases of the Central Arctic Ocean (CAO) survey and monitoring program are available in a timely manner for the initial post-mapping analysis and ongoing monitoring. The framework needs to strike a balance between making data available for FiSCAO (Scientific Experts on Fish Stocks in the Central Arctic Ocean) assessments and protecting the intellectual property rights of primary investigators, organizations, institutions and countries.

Various examples of existing data sharing and hosting protocols are available from international groups, including ICES, NAFO and the Distributed Biological Observatory (DBO). The DBO provides an excellent example framework for connecting databases from multiple countries and agencies, and is supported by the Sustaining Arctic Observing Networks (SAON) program. Adoption of similar protocols and frameworks would streamline integration of data collected during the mapping phase of the Central Arctic Ocean survey and monitoring program with existing Arctic datasets.

A distributed database is preferred over a single hosted database for several reasons. A distributed database combines metadata on existing data sets and data collection programs within a small hosted database, simplifying data discovery. Data records are hosted by member data centers (e.g. Government of Canada data centre), allowing data owners to maintain greater control over data access. Clear protocols for data sharing, use and publication are essential.

**Recommendations:** Contact DBO data archive developers (e.g. Jacqueline Grebmeier) and managers to assist with establishing the FiSCAO data archive and to discuss the successes and difficulties encountered with their data policies and protocols. Could a shared FiSCAO archive be created within the existing DBO framework?

# FiSCAO Data Policy and Release Guidelines – Draft (October 2, 2017)

[Based on: DBO DATA POLICY AND RELEASE GUIDELINES-­‐Final Version (Feb. 20, 2015)]

## Introduction

The group of Scientific Experts on Fish Stocks in the Central Arctic Ocean (FiSCAO) was established to answer the following questions:

* What are the distributions and abundances of species with a potential for future commercial harvests in the central Arctic Ocean?
* What other information is needed to provide advice necessary for future sustainable harvests of commercial fish stocks and maintenance of dependent ecosystem components?
* What are the likely key ecological linkages between potentially harvestable fish stocks of the central Arctic Ocean and adjacent shelf ecosystems?
* Over the next 10-30 years, what changes in fish populations, dependent species, and the supporting ecosystems may occur in the central Arctic Ocean and the adjacent shelf ecosystems?

To answer these questions the group is developing mapping and monitoring plans for the high seas region of the Central Arctic Ocean (CAO) and adjacent territorial waters. The mapping plan will collect baseline data on marine fishes, phytoplankton, zooplankton, benthos, marine mammals, seabirds and oceanography in the high seas CAO. Data from this mapping phase will be used for initial assessments of species distributions and abundances, and to quantify trophic linkages. Monitoring plans are being developed for the high seas CAO and adjacent territorial waters to collect data that will be used for regular reassessments of populations and ecosystem status; detected changes could trigger new surveys in targeted areas to re-evaluate potential commercial harvesting opportunities in the CAO.

Coordinated multi-national mapping and monitoring programs will require the establishment of an agreement on data management policies that permit the sharing of monitoring and research data related to the FiSCAO program (see the Final Report of the Fourth Meeting of Scientific Experts on Fish Stocks in the Central Arctic Ocean). The FiSCAO group noted that several international bodies have already developed data management policies (e.g., DBO, SAON and IASC, ICES) that might address the current need for data standards and protocols for metadata, quality assurance and data sharing. Development of a distributed data management system was encouraged, in which each nation is responsible for the storage and maintenance of the data it collects, while software provides search and query capabilities across the individual databases; this could be facilitated by leveraging existing networks, such as the DBO. The need for public data sharing limitations and protocols was identified as an important component of any data sharing agreement and policy. This document was developed to provide a draft Data Policy and Release Guidelines for presentation and refinement at the Fifth Meeting of Scientific Experts on Fish Stocks in the Central Arctic Ocean in October 2017. This document was based largely on the Distributed Biological Observatory Data Policy and Release Guidelines.

## Definition of the FiSCAO Data Archive

The FiSCAO distributed data archive will consist of a series of distributed data centers, combining data from multi-national FiSCAO and national sampling programs. A single site (website) will need to be developed for the submission of metadata that meet the standard FiSCAO metadata profile (see Appendix A).

FiSCAO data (see list of potential data fields in Appendix B) will need to be available to FiSCAO researchers in a timely manner for analysis, and to the larger community once initial analyses are completed. The first step in submitting FiSCAO data will be the completion of a metadata profile for the dataset. The data will then be submitted to a national or institutional data archive that is part of the FiSCAO distributed data archive. Metadata should be submitted as soon as possible (i.e. within one month) after completion of a sampling program. Data should be made available as soon as possible after collection and completion of quality assurance programs. A common, password protected shared data archive may be established (e.g. through the Arctic Council Circumpolar Biodiversity Monitoring Program or the pan-Arctic Sustaining Arctic Observing Network program) to facilitate analyses upon completion of the mapping phase of the FiSCAO program and repeated analyses throughout the monitoring phase.

Data centers that are part of the FiSCAO distributed data archive will need to coordinate their data management activities, including developing consistent metadata generation, curation and interoperability. When data submitted directly to an institutional or national archive are deemed ready for long term storage and distribution, a final version of the data and metadata will be uploaded or linked to the shared-archive, if one has been created.

## FiSCAO Data Policy and Release Guidelines

### Data Policy in Compliance with IASC and other Collaborative Arctic Activities

Any policy for the dissemination of FiSCAO data should be consistent and compliant with International standards and agreements such as the IASC Statement of Principles and Practices for Arctic Data Management and any data sharing commitments made during the International Polar Year (IPY). That is, free, timely, and unrestricted exchange of essential data and products to the maximum extent possible. The FiSCAO data policy approach is fully compatible with the World Meteorological Organization (WMO) CLIVAR Data Policy. The FiSCAO data archive will follow the WMO Core Profile of the ISO 19115: Geographic Information -­‐ Metadata standard.

This FiSCAO policy is not meant to conflict or supersede any national or international agency policy related to public access to these data, such as the U.S. Public Access to Research Results (PARR).

### Broad Community Access to Data

It is in the best interests of both data providers and potential users to maintain only the latest version of fully audited data and metadata in FiSCAO archives. This will allow the FiSCAO archives to alert users of data revisions or updates. Unrestricted copying of original data from sources outside FiSCAO archives may lead to the propagation of errors in data analyses, confusion regarding data versions, incomplete metrics and loss of recognition of FiSCAO as the data source. The sharing of data through a common password protected FiSCAO shared archive would allow for the distribution of preliminary data among the FiSCAO group before it is submitted to national or institutional archives and made fully accessible to the broader community.

### Acknowledgement and Citation

Whenever FiSCAO data are downloaded from the archive and used in publications, the data’s origin should be acknowledged and referenced. Every user is responsible for referencing the Principle Investigator (PI) responsible for creating the dataset that is used and identifying that the dataset was obtained through the FiSCAO data archive. If multiple sources have been used, acknowledgement must be provided for each dataset used.

International agencies, professional societies and research organizations are moving towards the formal citation of data and sources that led to a given analytical result or conclusion. Consequently, there has been increased use of DOIs (Digital Object Identifiers) as a simple, standard way to reference datasets. DOIs allow for linkages between datasets and respective publications, providing the ability to track the use of individual datasets in the literature and to provide metrics of their use or influence. DOIs are considered perpetual and provide proper attribution, even if a dataset is moved between archives. Standards have been established for the creation of data DOIs and have been supported by international coordination groups such as the Research Data Alliance (RDA).

Recommendation: It would be beneficial if an agreement with an existing organization could be developed to help data providers develop DOIs if their institutional or national data archive cannot provide the service.

### Co-Authorship for FiSCAO Principal Investigators

Research programs that contribute data to FiSCAO use sophisticated, state-of-the-art instrumentation and comply with strict requirements for maintenance, exposure of instruments, calibration, quality assurance procedures and the like, in order to achieve the highest attainable standards of measurement, accuracy, representativeness, stability and repeatability. To ensure that this goal is reached, PIs who are leading experts for their instruments will take responsibility for individual instruments operated on the respective research program.

Users of FiSCAO data are encouraged to establish direct contact with FiSCAO Scientific Point of Contact for each data set used; this contact is included in the metadata for each data set. The FiSCAO Scientific Point of Contact will discuss the planned use of the dataset and, if necessary, put the data user in contact with the data set PI as the data provider for the purpose of complete interpretation and analysis of data for publication purposes. Co-authorship of FiSCAO PIs on publications that make extensive use of FiSCAO data is highly recommended if their work has contributed to the study in question, or if they have been involved in directly contributing to the publication in other ways. It is highly recommended that any data user contact the responsible PI and discuss whether the PI’s data collection and Quality Assurance (QA) or Quality Control (QC) work warrants co-authorship or an acknowledgement.

### FiSCAO Publication List

Users of FiSCAO data are strongly encouraged to submit citations for any publications or products that make use of FiSCAO data to the FiSCAO shared archive. The FiSCAO shared archive will develop a citation list of FiSCAO publications from the submitted citations. Whenever possible, the FiSCAO archive will use DOIs to link to a publication to its data source(s). The FiSCAO shared archive will make the FiSCAO citation list public via the archive website to provide a continuous record of applications and analyses of FiSCAO data and FiSCAO scientific achievements.

## References

Arctic Council Circumpolar Biodiversity Monitoring Program (CBMP) http://www.arctic-­‐council.org/index.php/en/

Distributed Biological Observatory (DBO) http://www.arctic.noaa.gov/ International Arctic Science Committee (IASC) http://www.iasc.info/ Research Data Alliance (RDA) https://rd-­‐alliance.org/

Final Report of the Fourth Meeting of Scientific Experts on Fish Stocks in the Central Arctic Ocean https://www.afsc.noaa.gov/Arctic\_fish\_stocks\_fourth\_meeting/pdfs/FourthFiSCAOreportfinalJan26\_2017.pdf

Sustaining Arctic Observing Networks (SAON) Program http://www.arcticobserving.org/

Public Access to Research Results (PARR) http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp\_public\_access\_memo\_20

13.pdf

World Meteorological Organization (WMO) Climate and Ocean: Variability Predictability and Change (CLIVAR Project) Data Policy http://www.clivar.org/resources/data/data-­‐policy

## Appendix A – FiSCAO METADATA PROFILE

|  |  |  |
| --- | --- | --- |
|  | **Metadata Field Name** | **Definition** |
| **General** | title | A name given to the data set. |
|  | dataset authors | The person(s) receiving credit for the data set, as in a citation (usually the PI). |
|  | FiSCAO scientific point of contact | The person(s) to contact when using the data, to make link with dataset authors, for data requests and interpretations. |
|  | organizations | The organization(s) for which the PIs worked at the time of data collection |
|  | description | A summary of data set content |
|  | language | Language of the data set (e.g. English, Japanese, Korean) |
| **Dataset Details** | cruise number  funding agency | Cruise ID or Number  The Agency providing funding (e.g. JAMSTEC, NSF, NOAA) |
|  | grant, project or award number | Agency assigned Grant or Award number |
|  | temporal coverage start | Begin date of full data set |
|  | temporal coverage end | End date of full data set |
|  | continuation | Likelihood that the data collection will continue and be useful for ongoing monitoring (categorical) |
|  | temporal resolution | The sampling or reporting frequency of an instrument or platform |
|  | northernmost latitude | Northern extent of data collection in decimal degrees |
|  | southernmost latitude | Southern extent of data collection in decimal degrees |
|  | westernmost longitude | Western extent of data collection in decimal degrees |
|  | easternmost longitude | Eastern extent of data collection in decimal degrees |
|  | regions occupied | List of large marine ecosystems in which data were collected. |
| **Contact Info** | point of contact | Person who is responsible for the content of the metadata and data. |
|  | principal investigator | The PI responsible for leading the project |
|  | publisher | The Institution where the data resides and responsible for distributing the data set (e.g. KOPRI, JAMSTEC, PRIC, UCAR/NCAR, CCIN) |
|  | Weblink to dataset | Electronic link to location of dataset (e.g., at KOPRI, JAMSTEC, AOOS, EOL) |
| Data Details | platform | The vessel or vehicle from which instruments are deployed, or name of ice station |
|  | instrument | The name of the instrument used to acquire the data |
|  | keywords | Suite of data type collected, with pull down menu of parameters from FiSCAO data matrix: e.g., CTD, ADCP, bottle data for chlorophyll, nutrients; abundance, biomass and composition of Ice algae, phytoplankton, zooplankton, benthic fauna (infauna and epifauna) and fish; seabird and marine mammal surveys; Mooring data (T, S, Currents, fluorescence, nutrients, sediment trap); Satellite data (surface T, S, winds) |
| \ | science keywords | GCMD Science Keywords |
|  | data version | Version number of the data set available |
|  | dataset last revision date | Date the data set was last revised |
|  | distribution format | Distributed file format of the data set (e.g. excel, ascii, multiple) |
|  | data set progress | Amount of progress through to data publication (i.e. in work, or completed) |
|  | access restriction | Password protection required |
| **Citation** |  | Any additional citations |
| **DOI** |  | List if known |

## Appendix B – Definition of FiSCAO Data

The following is a suite of potential FiSCAO data types and parameters:

* conductivity
* temperature
* depth
* current direction
* current speed
* turbidity
* fluorescence
* chlorophyll
* dissolved nutrients
* species specific abundance and biomass
  + ice algae
  + phytoplankton
  + zooplankton
  + benthic infauna
  + benthic epifauna
  + fishes
  + sea birds
  + marine mammals
* measurements and biological samples collected from individuals
  + morphometrics
  + ageing structures
  + organ weights and volumes
  + stable isotope samples
  + genetics samples
  + sample storage location
* Sediment
  + grain size
  + organic carbon content
  + chlorophyll *a* content
* Satellite data
  + chlorophyll pigment concentration
  + sea surface temperature
  + sea ice concentration
  + cloud fraction
  + winds
  + sea level pressure
* analytical results
  + stable isotope values
  + fatty acid values
  + genetic sequences