

DATA MANAGEMENT PLAN FOR PROPOSAL:

BCO-DMO: ACCELERATING SCIENTIFIC DISCOVERY THROUGH ADAPTIVE DATA MANAGEMENT

DATA POLICY COMPLIANCE

Through the course of this project's goals and objectives, various data files resulting from projects funded by the NSF OCE Biological and Chemical Sections, and the Office of Polar Programs (OPP) Antarctic Sciences (ANT) Organisms & Ecosystems Program will be received, documented, assembled along with associated metadata, and served on the BCO-DMO website (<https://www.bco-dmo.org>). This activity facilitates project compliance with the NSF Proposal & Award Policies & Procedures Guide (PAPPG, NSF 17-1, Chapter XI.D.4, 2017) and OCE Sample and Data Policy (NSF 17-037, 2017). This project will also work to comply with various community research programs' policies and guidelines such as JGOFS, GLOBEC, and GEOTRACES. Once data have been processed and publicly served via the BCO-DMO website, data will be packaged, published with a Digital Object Identifier (DOI), and deposited into the Woods Hole Open Access Server, an Institutional Repository of the Marine Biological Laboratory and Woods Hole Oceanographic Institute (MBL-WHOI) Library to facilitate version control of published datasets. In addition, these same data packages will be submitted to an appropriate national data center for permanent archive, e.g., the National Centers for Environmental Information (NCEI).

DESCRIPTION OF DATA TYPES

Data submitted to, and managed by BCO-DMO are the result of research efforts conducted globally in marine, limnological, and atmospheric environments within the physical, chemical, biological, ecological and biogeochemical sub-domains of oceanography. Observations encompass full range of oceanographic research measurement types including: in situ sampling, moorings, floats and gliders, traps; results from laboratory experiments; satellite images; derived parameters and model output; and synthesis products from data integration efforts. Temporal scales of observations range from seconds to years, and size scales from microbial to megafauna.

All datasets are curated along with accompanying metadata including information on instrumentation, acquisition methodologies, and related publications. Instrument types range from standard *in situ* oceanographic sampling instruments (e.g., CTD, Niskin bottles, cores, etc.) to novel experimental apparatus and satellites.

DATA AND METADATA FORMATS AND STANDARDS

Data are submitted to BCO-DMO in a variety of file formats, both proprietary and non-proprietary. Data and metadata files, and all associated submission correspondence are captured and tracked in a ticketing system for management of data processing workflows. While an original copy of each data file is retained, the files are reformatted as necessary into a non-proprietary flat file format such as .csv, facilitating use by a wide variety of end-users and software.

Metadata enabling discovery, interpretation, and use of the data are collected using a standardized form from all submitters. Metadata are then entered, stored, and curated in a local, format-agnostic database which can be written out in a number of standards-compliant formats (e.g., ISO19115-2, Global Change Master Directory, DublinCore, and DataCite).

BCO-DMO data managers work closely with data originators and submitters to ensure observational data are quality checked and reformatted in accordance with community accepted standards, vocabularies and conventions, and that metadata are as complete and accurate as possible.

DATA STORAGE AND ACCESS DURING THE PROJECT

Data are stored on local servers in a flat file format in a canonical directory structure. Metadata are stored locally in databases and flat files. Once data and metadata are assembled, processed, and validated with the submitting investigator, they are linked to the website and made publically available. All data and metadata are backed up locally on the WHOI campus (a system maintained by the WHOI Computer Services Department), with a dedicated, network-accessible storage unit located in Boston to provide secure and off-site backup. This offsite backup in Boston will transition to Amazon's new S3 Glacier Deep Archive in Y1. Data files managed by the office range from kilobytes to terabytes in size, none being larger than 4TB, with a total size holding of approximately 13TB. Based on recent projections for data submission over the next 5 years, we've proposed a capacity of up to 500TB of data and metadata.

MECHANISMS AND POLICIES FOR ACCESS, SHARING, RE-USE, AND RE-DISTRIBUTION

Data submitted to BCO-DMO are discovered and accessed through a web-based system. This system is free and open to the public, with no subscription or login barriers. The BCO-DMO website is the property of the BCO-DMO Project and it maintains the right to determine terms and conditions for website use.

Data submitters who contribute their project's output to the office have an option (in compliance with the NSF OCE Sample and Data Policy) to embargo or restrict access to their data in the BCO-DMO database for a period of two years to allow for publication. During this time, data are still discoverable, and use is negotiated via correspondence with the data originator. After the embargo period, data are made publicly accessible. Those investigators wishing to extend their embargo period must negotiate an accepted agreement with their funding agency and communicate the outcome to the office, where accommodations are made accordingly.

BCO-DMO applies the Creative Commons Attribution 4.0 International License (CCBY4; <https://creativecommons.org>) to all of its data holdings. Under this license, it is understood that any use of the data will properly acknowledge the originators and/or submitters and other related parties. In addition, BCO-DMO recommends a suggested citation format to ensure proper attribution of data originators. This citation is visible and accessible on each dataset page.

The office also obtains Digital Object Identifiers (DOIs) for its datasets, facilitating data discovery, use, and citation in scholarly publications. Where possible, the office documents and tracks linkages between scholarly publications related to the datasets it manages. Potential users of the BCO-DMO system include researchers from all subdomains within the oceanographic and limnological sciences, and cross-disciplinary investigators from related geological and biological sciences. In addition, the BCO-DMO data holdings are also used by teachers and professors in educational courses.

PLANS FOR ARCHIVING

BCO-DMO functions as an intermediate data assembly center for NSF investigators to deposit their data, and for the community at large to use for new science and education endeavors. Therefore, the office ensures that the data it manages are deposited in an appropriate national archive (e.g., the National Centers for Environmental Information, NCEI) for long-term preservation and access should funding on the BCO-DMO project cease. BCO-DMO has collaborated with NCEI to establish automated pipelines for archive of contemporary and legacy data holdings and has a service agreement in place documenting the relationship.

ROLES AND RESPONSIBILITIES

Data and information managed by BCO-DMO represent contributions by hundreds of NSF-funded investigators who submit their project output in compliance with the NSF OCE Sample and Data Policy. BCO-DMO data management staff work closely with each data submitter to ensure they meet Policy requirements, that their data are shared in a timely manner, and the data are as complete and accurate as possible. The BCO-DMO director and technical director assume the responsibility of ensuring that the strategies and processes are in place to execute the project objectives of curating and serving data online, and that all data are archived appropriately.