SUBMISSION AGREEMENT
BETWEEN
THE BIOLOGICAL AND CHEMICAL OCEANOGRAPHY DATA MANAGEMENT OFFICE
AND
THE NATIONAL CENTERS FOR ENVIRONMENTAL INFORMATION
FOR BIOLOGICAL AND CHEMICAL OCEANOGRAPHY DATA MANAGEMENT OFFICE

2018-05-09

Introduction

This document represents the agreement that the Biological and Chemical Oceanography Data Management Office (BCO-DMO) (the "Provider") and the National Centers for Environmental Information (NCEI) (the "Archive") have reached for submitting the Provider's data, Biological and Chemical Oceanography Data Management Office, to the Archive for long-term preservation. It represents a joint effort between the Provider and the Archive to accurately document the agreement and the expectations between the two groups.

In order to ensure that the quality and integrity of the archived data is not compromised, the Provider and the Archive agree to maintain this agreement with accurate and up-to-date information through the life of the data submission.

Add comments as needed

Contacts

Persons included in all communications regarding the data submission.

Provider Contacts

Point of Contact, data manager, general POC

Danie Kinkade

BCO-DMO > Biological and Chemical Oceanography Data Management Office

Data Manager

1 508 289 2291

dkinkade@whoi.edu

Archive Contacts

Data Acquisition, project manager, subject matter expert, general POC	Technical Steward, Programmer
Scott Cross	John Relph
DOC/NOAA/NCEI	DOC/NOAA/NCEI
Project Manager	Programmer
843-460-9646	301-713-4914
Scott.Cross@noaa.gov	John.Relph@noaa.gov

science steward, data content manager

Jonathan Jackson

DOC/NOAA/NCEI

DCM

228-688-1261

Jonathan.Jackson@noaa.gov

Data Overview

This submission agreement covers oceanographic data collected from research projects funded by the Biological and Chemical Oceanography Sections, the Division of Polar Programs Arctic Sciences and Antarctic Organisms & Ecosystems Program at the U.S. National Science Foundation. These data are submitted by staff at BCO-DMO WHOI by placing the datasets onto a FTP site (https://www.bco-dmo.org/files/NCEI/) that will be harvested daily by NCEI through a semi-automated process. The data content of these submissions will be monitored by the DCM.

This work was previously at production stage until the submitter began efforts to convert their metadata from FGDC metadata to ISO 19115-2 metadata. As the conversion to ISO is nearing completion, NCEI and BCO-DMO staff have began to retool the semi-automated process used to archive of these datasets.

Applicable and Reference Documents

Documents applicable to or referenced from this agreement.

1. Citation for applicable/reference document, e.g,. an archive appraisal or project requirements document

Submission Scope

Active Submission Period

2011-01-01 - 9999

Data Types

Below is a summary of the data sizing and submission schedule by data type group. Enter information on at least one data type.

Data Type Name	Data Sizing	Submission Schedule
comma/tab separated, various	typically <10 GB	as prepared

Reviews and Testing

NCEI staff have worked with BCO-DMO staff to setup a semi-automated process to archive datasets held at BCO-DMO. Multiple test accessions have been reviewed by NCEI staff using this process and feed back has been given to the submitter.

Providing System

Identification of the system providing the data to NCEI.

System Name: WHOI FTP site
System Owner: WHOI
Physical Location: Woods Hole, MA
Additional Information: Add comments as needed on applicable data types, etc.

Transfer Interface

BCO-DMO places datasets onto the designated FTP site, https://www.bco-dmo.org/files/NCEI/. Then the automated process downloads the .sha384 files from the FTP site daily, checks to see if the files have been previously downloaded, and if not, downloads the .tar file and .sha384_summary file.

Submission File Inventory

Information on each submitted file type from the Provider. Information on multiple file types can be added below.

File Type Name: Dataname

File Name Pattern:

Data.csv

File Name Field Definitions:

The filename will be meaningful to the data type/project

Example File Name:

US_GLOBEC_GB_chloro_bottle.csv

File Format: CSV

File Compression: None

File Size Average: 34KB

File Count (Rate): variable

Data Volume (Rate): variable

Submission Schedule: as prepared

Additional Information:

Descriptive Information Attributes:

None: date/time fields are usually used as descriptive attributes for this file.

File Type Name: metadata

File Name Pattern:

ISO1911502.xml

File Name Field Definitions:

Contains additional metadata necessary for understanding the data provided
  in the comma- or tab-separated value file(s) above. The file conforms to the
  ISO 19115-2 content standard, and contains attributes corresponding to the
  column headers, describing definitions and units for these attributes.

Example File Name:

ISO1911502.xml

File Format: XML

File Compression: None

File Size Average: 20KB

File Count (Rate): 1 per submission

Data Volume (Rate): one per package

Submission Schedule: as prepared

Additional Information:

Descriptive Information Attributes:

None: date/time fields are usually used as descriptive attributes for this file.

Submission Manifest

A submission manifest file with a 32-character MD5 checksum value is required for each submitted file in order to ensure the integrity of the submitted data.

File Content Specification:

sha384_summary manifest for the files contained in the .tar file

File Transmission:

variable

File Name Pattern:

The sha384 files will have the same name as the tar file except for the file extensions

File Name Definitions:

Definitions of the fields in the submission manifest file name pattern

Example File Name:

McMurdo_Predator_Prey_McMurdo_Sound_sea_ice_thickness.sha384

Archive Ingest

Ingest processing steps at the Archive and communication with the Provider.

Receipt Verification:

The automated process downloads the .sha384 files daily from the FTP, checks to see if the files have been downloaded previously, and if not, downloads the .tar file and .sha384_summary file. Then calculates the SHA-384 checksums of the .tar and .sha384_summary files, if these files match the values in the .sha384 file then the tar file is

unpacked. The SHA-384 checksums of all of the files extracted from the .tar file is calculated and compared against the values in the .sha384_summary file. If those match and the "ISO1911502.xml" metadata file is included the SIP is considered valid and is copied to the /nodc/data/BCO-DMO/SIP directory for ingest processing.

Error Reconciliation:

The Archive will report any problems or errors with file integrity, file name, checksum validation, or other errors that inhibit the data ingest and archive to the Provider. A new corresponding submission manifest will be required for files re-submitted by the Provider.

Receipt Confirmation:

The Archive will provide an inventory of the data ingested once it is completed or as requested by the Provider.

Quality Assurance:

No quality checks on the submitted data are planned.

Archive File Packaging:

The packages are submitted as a tar file with an associated checksum. NCEI will untar the file to reveal the contents of data.csv, data.xml, and any other supplemental files in a supplemental directory. The tar file and checksum will be discarded after extraction and validation. An automated process extracts various metadata from the ISO 19115-2 XML file included in the package. The gmd:fileIdentifier from the ISO is used to determine the BCO-DMO package ID. Each BCO-DMO package ID corresponds to one NCEI accession. Title, Abstract, and other fields from the ISO metadata are mapped to NCEI accession metadata. If required ISO metadata fields are not present, the package is considered invalid.

Archive Storage

Archive attributes of each archived file type.

<table>
<tr><td colspan="2">Archive File Type Name: Descriptive name for this archive file type
Archive File Attributes/IDs:</td></tr>
<tr><td>Attribute/ID Type</td><td>Value</td></tr>
<tr><td>Metadata ID</td><td>Attribute/ID value</td></tr>
</table>

Archive Updates

Data submissions intended to update an existing archive record require adequate notification and justification. Updates can supersede previous data submissions as a newer or improved version, however any previously submitted data will not be removed from the archive for the purpose maintaining version control and traceability in the archive.

Retention Schedule

The data will be retained in the Archive for long-term preservation in accordance with NOAA data management standards. Information on data usage and archive value may be used for making decisions on continuing the duration of the archive.

(Notional) Disposition: Unknown/TBD

Constraints

No constraints apply or will apply to the archived data.

User Community
Oceanographers and other marine oriented researchers

User Documentation and Metadata
The Provider will supply information to the Archive for writing and maintaining standard archive metadata, which includes data discovery information, references and data archive access links for users. The following published documents and archived items will be referenced from the metadata and made available to users.

Representation Information Items
For data to be useful to users, present and future, its format specification and characteristics must be documented and preserved with the data. Representation Information provides users with syntax (structure) and/or semantics (meaning) to decode the encoded data.

Item	Description
Item name or citation	Item description or intended use

Preservation Descriptive Information Items
Preservation Descriptive Information items contain context, provenance, and/or quality information for the data.

Item	Description
Item name or citation	Item description or intended use

Access and Dissemination
The Archive will provide access services for the data and supporting information to the designated user community.

Additional Terms
None.