Assessment Information

CoreTrustSeal Requirements 2017–2019

Repository:	Biological and Chemical Oceanography Data Management Office
Website:	https://bco-dmo.org
Certification Date:	13 July 2020

This repository is owned by: **Woods Hole Oceanographic Institution**

Biological and Chemical Oceanography Data Management Office

Notes Before Completing the Application

We have read and understood the notes concerning our application submission.

True

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments:

CORE TRUSTWORTHY DATA REPOSITORIES REQUIREMENTS

Background & General Guidance

Glossary of Terms

BACKGROUND INFORMATION

Context

R0. Please provide context for your repository.

Repository Type. Select all relevant types from:

Domain or subject-based repository, Research project repository

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Comments

BCO-DMO is a domain-specific repository providing data management services to researchers funded by the National Science Foundation (NSF) Biological and Chemical Oceanography Sections and Office of Polar Programs (NSF, award #1435578, https://nsf.gov/awardsearch/showAward?AWD_ID=1435578). BCO-DMO data managers curate marine ecosystem research data and information at no cost for investigators funded by the same NSF Division that fund BCO-DMO. The office works with researchers throughout their data's lifecycle from proposal to preservation.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Brief Description of the Repository's Designated Community.

BCO-DMO is considered a domain repository serving the broader ocean ecosystem research community. Ocean ecosystem research is, however, highly interdisciplinary and encompasses a number of subdomains. The repository has worked with over 2,600 data contributors from physical, chemical, biological and/or ecological and biogeochemical sub-domains on data from marine, limnological, interstitial, benthic, and atmospheric and aerosol studies. BCO-DMO provides data management services free of charge to those investigators funded through the NSF's Geoscience Directorate, Biological and Chemical Oceanography Sections and Office of Polar Programs, and collaborates with investigators and funders from other organizations on a fee-for-service basis. BCO-DMO's primary stakeholder researchers are academic, however the repository also supports US state and locally funded researchers as well.

The NSF Biological and Chemical OCE Sections, and the Division of Polar Programs created BCO-DMO in 2006 to work closely with oceanographic researchers from these NSF sections throughout their project data's life cycle, helping them fulfill the data sharing requirements of the NSF OCE Sample and Data Policy (NSF17-037, 2017 https://www.nsf.gov/pubs/2017/nsf17037/nsf17037.jsp), and educating them in basic data skills. Since its inception, the

office has cultivated an effective relationship with its research community, and continues to serve it, remaining responsive to their data curation needs by leveraging technological advancements in information management.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Level of Curation Performed. Select all relevant types from:

D. Data-level curation – as in C above; but with additional editing of deposited data for accuracy

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Comments

BCO-DMO's core activities support the entire research data lifecycle and require highly-trained personnel. Staff work closely with Principal Investigators (PIs) to provide data management services as early as the proposal stage and continue through data acquisition, submission, publication, and ultimately preservation. Office staff possess both formal domain science training through advanced degrees, and hands-on experience from fieldwork, making them familiar with many of the data types being curated. The office supports approximately seven full and part time data managers all possessing advanced degrees in an oceanographic or related discipline, in addition to data management experience.

BCO-DMO data managers work with data submitters to process and publish their data, metadata and information online. This is a collaborative and sometimes iterative effort with the data originator, to generate a product that facilitates the understanding necessary for reuse and reanalysis of project output. Work performed by a data manager may include merging, extraction, reformatting, and transformation of data to tabular, non-proprietary formats. Gross quality control is performed and observational values may be edited at the PIs request. In addition, structured metadata forms, scholarly publications, and direct communications are used to develop detailed metadata necessary to interpret and re-use the data. Related data and publications are linked to dataset metadata. Once published, BCO-DMO continues to update and version data as necessary, and submits the data to both an Institutional Library (WHOI-MBL Library) for versioning and DOI assignment, and the National Archive (National Centers for Environmental Information) for long-term preservation.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Outsource Partners. If applicable, please list them.

- DataONE is an interdisciplinary data aggregator that harvests and catalogs BCO-DMO data for increased discovery. (https://www.dataone.org/)

- The MBLWHOI Institutional Library's Data Library and Archives (WHOAS) is used to manage dataset publication activity, e.g.,DOI assignment and management of versioned datasets (https://darchive.mblwhoilibrary.org/)

- Woods Hole Oceanographic Institution Information Services (https://whoi-it.whoi.edu/) provides technical assistance for server management and security of data store.

- Second Creek Consulting: Charlton Galvarino maintains the existing geospatial search interface on the BCO-DMO website. (http://2creek.com/)

- Offsite backup company, Markley Group (https://www.markleygroup.com/data-center) provides offsite backup services for data store and website.

- The National Centers for Environmental Information (NCEI) partners with BCO-DMO (via MOU and Submission Agreement, https://www.ncei.noaa.gov/) to ensure long-term archive of data curated by the office. Data are pulled from BCO-DMO once they are validated and published with a DOI and deposited in WHOAS.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

Other Relevant Information.

Use and impact of repository holdings, National, regional, global role the repository plays:
BCO-DMO is a well-known data repository within the oceanographic community; accessed by over 260 countries world-wide, and is the recommended domain repository for the U.S. National Science Foundation's Biological and

Chemical Oceanography Programs and Office of Polar Programs, as identified in NSF OCE Sample and Data Policy (https://www.nsf.gov/pubs/2017/nsf17037/nsf17037.jsp). Researchers funded through these programs are required to deposit their project data with BCO-DMO in order to comply with the policy. As a result, the repository has worked with over 2,600 data contributors to date from physical, chemical, biological and/or ecological and biogeochemical sub-domains on data from marine, limnological, interstitial, and aerosol studies, representing 1,000 funded projects. The office curates the data resulting from this research, and in turn, provides public access to over 9000 datasets for use in national and international research (e.g., Amundsen Sea Polynya International Research Expedition [ASPIRE] Project), academic and teaching environments (e.g., University of Rhode Island's Graduate School of Oceanography), and by the general public.

BCO-DMO is a Member Node of the Data Observation Network for Earth (DataONE; https://www.dataone.org/), contributing its holdings to the DataONE catalog for discovery and use in inter- and cross-disciplinary research. The BCO-DMO catalog is also used by the NSF projects: The Center for Dark Energy and Biosphere Investigations (CDEBI, https://www.darkenergybiosphere.org/) through display of BCO-DMO datasets relevant to the CDEBI community; and PlanetMicrobe (http://www.planetmicrobe.org/) and the Ocean Protein Portal (http://) through harvest of BCO-DMO's marine ecosystem data necessary for interpretation of microbial research results.

The office is a charter member and executive committee member of the Council of Data Facilities (https://www.earthcube.org/group/council-data-facilities), and is a signatory of the Coalition for Publishing Data in the Earth and Space Sciences (COPDESS) Statement of Commitment (http://www.copdess.org/home/about-copdess/), and the Enabling FAIR Data Commitment Statement in the Earth, Space, and Environmental Sciences (http://www.copdess.org/enabling-fair-data-project/commitment-to-enabling-fair-data-in-the-earth-space-and-environmental-sciences/). BCO-DMO is an active member and/or participant of the Earth Science Information Partners (ESIP, https://www.esipfed.org/), the Research Data Alliance (https://www.rd-alliance.org/), the American Geophysical Union (https://www.agu.org/), the U.S. National Science Foundation's EarthCube Project (https://www.earthcube.org), and is listed in the Registry of Research Data Repositories (https://www.re3data.org/repository/r3d100000012). Actively engaging in these informatics communities of practice allows BCO-DMO to both stay apprised of novel technologies and best practices, and help achieve broader goals of these organizations that better the repository community.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

ORGANIZATIONAL INFRASTRUCTURE

I. Mission/Scope

R1. The repository has an explicit mission to provide access to and preserve data in its domain.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

BCO-DMO's mission is to work with principal investigators to curate and serve data and information online from research projects funded by the National Science Foundation (NSF) Biological and Chemical Oceanography Sections and Division of Polar Programs Antarctic Organisms and Ecosystems. This mission is a direct result of the office's funded NSF project award (see https://www.bco-dmo.org/).

This role has been agreed upon and contracted by our funders, and is evidenced by BCO-DMO's project award and by documentation provided (by the funder) to BCO-DMO's stakeholder community as described in the following NSF documents:

NSF OCE Division Sample and Data Policy (NSF 17-037): https://www.nsf.gov/pubs/2017/nsf17037/nsf17037.jsp:
NSF OCE guidance webpage listing approved data repositories:
https://www.nsf.gov/geo/oce/oce-data-sample-repository-list.jsp
BCO-DMO award pages:
https://www.nsf.gov/awardsearch/showAward?AWD_ID=1435578
https://www.nsf.gov/awardsearch/showAward?AWD_ID=1924618

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

II. Licenses

R2. The repository maintains all applicable licenses covering data access and use and monitors compliance.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

BCO-DMO's website, data, and supporting documentation are freely available for use, providing users follow the Terms of Use (https://www.bco-dmo.org/terms-use), including the understanding that any such use will properly acknowledge the originating Investigator (see full terms on Terms of Use page).

"Submitters warrant that they own the rights to the content they are submitting and are authorized to do so under original copyright. Items in BCO-DMO are protected under original copyright with all rights reserved, where applicable. By submitting data and information, the submitter grants BCO-DMO the rights needed to copy, store, redistribute, and share data, metadata, and any other content. By validating and making their data publicly available, submitters grant BCO-DMO and any other users the right to reuse their data according to the terms of the CC BY 4 license:

By using this Website you are agreeing to be bound by these Terms and Conditions. [Creative Commons License logo] All datasets and associated information are licensed under a Creative Commons Attribution 4.0 International License. Per the CC BY 4 license it is understood that any use of the data set will properly acknowledge the PI.

BCO-DMO recommends that you contact the original principal investigator (PI) should you require additional information about the data. In the event that the data originator is not available, please contact BCO-DMO (info@bco-dmo.org) for guidance. "

Each BCO-DMO dataset metadata page provides the licensing and citation information (e.g. see "Cite this Dataset" button on dataset https://www.bco-dmo.org/dataset/772645)

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

III. Continuity of access

R3. The repository has a continuity plan to ensure ongoing access to and preservation of its holdings.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

BCO-DMO assumes responsibility for mid-term curation and long-term preservation of data applicable to its funding purview. Within mid-range time scales, data are processed, curated and made publicly accessible through the BCO-DMO website. Risk of inaccessibility or data loss is mitigated via multiple strategies: through the harvest and replication of data by DataONE, an NSF-funded data aggregation entity; and through deposition of data to BCO-DMO's institutional data library (the Woods Hole Open Access Server (WHOAS), a repository of the Marine Biological Laboratory-Woods Hole Oceanographic Institution Library, which is also Core Trust Seal certified; https://darchive.mblwhoilibrary.org/). The library assigns DOIs and maintains an immutable copy of each version of data for which a DOI is obtained.

This strategy is augmented for long-term preservation via relationship with an appropriate National Archive (the National Centers for Environmental Information, NCEI, https://www.ncei.noaa.gov/). Once data files have been processed, published via the BCO-DMO website, deposited in the Institutional library, and harvested by DataONE, they are then deposited with NCEI, a digital data archive of the National Oceanographic and Atmospheric Administration (NOAA), serving oceanography, geography and climate communities. It should be noted that BCO-DMO adopted a redundant

archive strategy, in part, to accommodate data types not accepted by NCEI (e.g., experimental and model data, genetic and other cellular-level data) and mitigate challenges with archiving legacy data (i.e., this activity is in process). However, we believe this redundant approach assures that BCO-DMO data will persist through unplanned lapses, or cessation of the repository project.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

IV. Confidentiality/Ethics

R4. The repository ensures, to the extent possible, that data are created, curated, accessed, and used in compliance with disciplinary and ethical norms.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

BCO-DMO's mission is to curate data from research awardees funded by the National Science Foundation (NSF) Biological and Chemical Oceanography Sections and Office of Polar Programs. As a function of receiving their awards, investigators agree to comply with all appropriate NSF data collection rules, as outlined in the NSF Proposal and Award Policies and Procedures Guide (PAPPG, https://www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf19001, and relevant subsections within sections B, D, E, G, H, K, and M of Chapter XI, https://www.nsf.gov/pubs/policydocs/pappg19_1/pappg_11.jsp#XID). These policies and procedures include detailed criteria for sensitive data collection and disclosure requirements that are complied with on behalf of the awardee prior to

BCO-DMO receiving data. BCO-DMO operates under the understanding that awardees have complied with the appropriate NSF award criteria and rules by virtue of successfully obtaining their award and remaining in good standing with their funders and institutions. BCO-DMO does not serve nor intend to serve data associated with disclosure risk. Should there be a question about the data collection or creation BCO-DMO would resolve with NSF funders and PIs accordingly.

Likewise, the output of NSF funded research is required to be open and public within 2 years of collection. See NSF OCE Division Sample and Data Policy (NSF 17-037): https://www.nsf.gov/pubs/2017/nsf17037/nsf17037.jsp. BCO-DMO treats all data submitted to the repository to be in the public domain as per the data policy referenced above, unless explicitly noted by the submitter (i.e., request for embargo). If a PI requests an embargo period beyond 2 years (e.g., for extended publication times), BCO-DMO directs the PI to their funding Program Manager to negotiate this request. BCO-DMO then works with the PI and NSF to extend any embargo period as mutually agreed upon by both parties. This process is revisited as necessary.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

V. Organizational infrastructure

R5. The repository has adequate funding and sufficient numbers of qualified staff managed through a clear system of governance to effectively carry out the mission.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

BCO-DMO Facility Support:
The NSF Biological and Chemical Oceanography Sections, and the Division of Polar Programs funded the creation of the BCO-DMO in 2006 and is on its fourth award (NSF award OCE-1924618: https://nsf.gov/awardsearch/showAward?AWD_ID=1435578, five years).

The BCO-DMO repository headquarters physically sits within the Woods Hole Oceanographic Institution's (WHOI's, https://www.whoi.edu/) campus infrastructure. The Institution is a world renowned oceanographic research facility. This location facilitates repository operations through access to domain expertise, proximity to several community data submitters and users, access to WHOI's technical infrastructure and support, and it's Institutional digital library and archive. WHOI personnel have offices located at the Woods Hole village or Quissett Campus locations of WHOI. BCO-DMO has a conference room for local group meetings that is custom built to support informatics design sessions, including support for remote participation.

The WHOI Information Services (IS) Department provides a virtual machine environment and large research data storage, including local tape backup service. This data store serves to store the BCO-DMO datasets. A dedicated, network-accessible storage unit is located in Boston, MA to provide secure and off-site backup of all the data and metadata stored on the BCO-DMO servers at WHOI. This backup is in addition to local system backups that are maintained and updated daily at WHOI's IS Department. Due to increasing costs, we plan to migrate this offsite backup to the commercial cloud. All local machines are connected to the Institution's TCP/IP based network. The Institution is connected to the Internet at Internet 2 and T1 (as backup) speeds.

Additionally, the WHOI Graphics Department assists with website redesign and implementation, preparation of documents for distribution and figures for presentations at workshops and meetings, and for publication. The IS and Communications groups at WHOI offer Web application development support and provide support as needed for code development and system administration. Departmental administrative support personnel assist with annual reporting requirements, manuscript preparation and meeting logistics support.

The WHOI Data Library and Archives (https://darchive.mblwhoilibrary.org/) is part of the larger Marine Biological Laboratory and Woods Hole Oceanographic Institution (MBLWHOI) Library system located in Woods Hole, MA. BCO-DMO collaborates with library staff who contribute expertise specifically in the area of data publication and citation, and who coordinate DOI assignment of BCO-DMO published datasets.

BCO-DMO forms key collaborations with science and informatics communities to achieve specific goals and objectives (see section: Background Information, Other Relevant Information above). For example, partnering with the NOAA National Centers for Environmental Information (NCEI, https://ncei.noaa.gov/) allows BCO-DMO to archive its data holdings for long-term viability and persistence. Collaborations with science communities help to solve targeted community data challenges such as increasing data interoperability and improved data discovery. Partnering with informatics communities of practice, such as the Earth Science Information Partners (ESIP, https://www.esipfed.org/) affords BCO-DMO access to emerging technology and best practices that benefit the science community.

Data Management Staff:
Fulfillment of BCO-DMO's mission relies heavily on its data management personnel, and significant investment is made to obtain staff possessing oceanographic domain expertise. BCO-DMO is fortunate to have seven excellent data managers in its group who have been selected for their academic background and field experience with oceanographic instruments and data types, as well as technical skills of data processing tools and software. This strategy ensures our data managers understand oceanographic concepts, research themes, scientific publications, and the data produced through stakeholder projects. All staff are encouraged to attend professional meetings in both oceanographic and informatics domains to represent the office, explore collaborations, share successful activities and improve skills. BCO-DMO allows its staff to travel to local and international meetings alike and cycle personnel to regularly attended meetings to ensure all staff participate at regular intervals. BCO-DMO administration also supports staff in personal development and trainings such as Software Carpentries and bootcamps as well as applications for internal staff training awards.

Affiliations (mentioned previously in "Background Information" section):
BCO-DMO is a Member Node of the Data Observation Network for Earth (DataONE; https://www.dataone.org/), contributing its holdings to the DataONE catalog for discovery and use in inter- and cross-disciplinary research. The BCO-DMO catalog is also used by the NSF projects: The Center for Dark Energy and Biosphere Investigations (CDEBI, https://www.darkenergybiosphere.org/) through display of BCO-DMO datasets relevant to the CDEBI community; and PlanetMicrobe (http://www.planetmicrobe.org/) and the Ocean Protein Portal (https://proteinportal.whoi.edu/) through harvest of BCO-DMO's marine ecosystem data necessary for interpretation of microbial research results.

The office is a charter member and executive committee member of the Council of Data Facilities (https://www.earthcube.org/group/council-data-facilities), and is a signatory of the Coalition for Publishing Data in the Earth and Space Sciences (COPDESS) Statement of Commitment (http://www.copdess.org/home/about-copdess/), and the Enabling FAIR Data Commitment Statement in the Earth, Space, and Environmental Sciences (http://www.copdess.org/enabling-fair-data-project/commitment-to-enabling-fair-data-in-the-earth-space-and-environmental-sciences/).
BCO-DMO is an active member and/or participant of the Earth Science Information Partners (ESIP, https://www.esipfed.org/), the Research Data Alliance (https://www.rd-alliance.org/), the American Geophysical Union (https://www.agu.org/), the U.S. National Science Foundation's EarthCube Project (https://www.earthcube.org), and is listed in the Registry of Research Data Repositories (https://www.re3data.org/repository/r3d100000012). Actively engaging in these informatics communities of practice allows BCO-DMO to access domain science and information science expertise, while maintaining the awareness of novel technologies and best practices.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:

Accept

VI. Expert guidance

R6. The repository adopts mechanism(s) to secure ongoing expert guidance and feedback (either inhouse or external, including scientific guidance, if relevant).

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

Repository management make-up is split between science and information management domains, with science co-PIs providing valuable connections and insights into the research community. BCO-DMO strives to ensure consistent community feedback and guidance with respect to evolving community needs. Strategies include creation and constitution of a Strategic Planning Committee, partnering with domain researchers for opportunities to gather relevant feedback, ad-hoc focus groups, participation in organizations of practice in data science, attendance at both data science and oceanographic science meetings, establishing relationships (MOUs) with necessary organizations/collaborators.

The BCO-DMO repository headquarters physically sits within the Woods Hole Oceanographic Institution's (WHOI's, https://www.whoi.edu/) campus infrastructure. The Institution is a world renowned oceanographic research facility. This location facilitates repository operations through access to domain expertise. BCO-DMO staff frequently leverages the domain knowledge of WHOI's research community in addition to forming strategic partnerships with oceanographic scientists to meet their specific data management needs (e.g., the ocean microbial communities with Planet Microbe and the Ocean Protein Portal, two projects that leverage BCO-DMO's environmental data to advance their own microbial analyses: https://www.planetmicrobe.org/ and https://proteinportal.whoi.edu/). These partnerships help BCO-DMO better understand the needs of emerging communities, and the data types, instruments, and analysis/visualization tools as well.

The Strategic Planning Committee (SPC) serves to provide consistent access to BCO-DMO's domain science community

for feedback and guidance on strategy development. The committee was constituted with the help and recommendations of BCO-DMO's funders and consists of seven members representing a range of oceanographic subdomains and research data. The committee meets 3-4 times per year (virtually 2-3 times, and once in-person), for two years to review BCO-DMO current data management practices, explore new ideas for meeting community needs, and provide feedback on proposed strategies and prototypes..

BCO-DMO leverages professional meetings of opportunity to engage and solicit feedback from its community. For example, during the 2018 Ocean Sciences Meeting, BCO-DMO conducted brief interviews with oceanographic scientists on their data search habits and preferences and used this output to guide modifications in its own infrastructure. We continued this user-engagement and feedback solicitation with an interactive poster session on researcher preferences for file formats, tools and languages, search criteria,

BCO-DMO obtains access to data science expertise in a similar manner, by attending meetings, presenting work, and forming collaborations for strategic enhancements or problem solving. BCO-DMO is a member of several community of practice organizations such as the Earth Science Information Partners (ESIP, https://www.esipfed.org/), the Research Data Alliance (https://www.rd-alliance.org/), and the American Geophysical Union (https://www.agu.org/). More locally, BCO-DMO is a member of the WHOI Ocean Informatics Advisory Group (http://www.whoi.edu/DoR/special-projects/ocean-informatics-working-group), charged with keeping the Institution apprised of external data science activities and ensuring the data science capacity development of Institutional students and staff. Within this organization, BCO-DMO has access to library science and computer science expertise. In addition, BCO-DMO Staff sit on the internal Information Services Advisory Team, which has a similar mission, however more focused on computer service and infrastructure for efficient operations.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

DIGITAL OBJECT MANAGEMENT

VII. Data integrity and authenticity

R7. The repository guarantees the integrity and authenticity of the data.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

BCO-DMO data management processes use multiple systems to track workflow and provenance:
Small data submissions are primarily done via email attachment sent to the office's general informational address (info@bco-dmo.org). Large file transfers are negotiated based upon submitters' capabilities (e.g., private Dropbox, mailing of disc drives, etc.) with correspondence still occurring via email. The "info" email address is monitored by all data managers, but has a dedicated and rotating data manager whose responsibility it is to accept, perform initial assessment, and respond to submitters. Submissions are then placed in a ticketing system (Redmine, https://www.redmine.org/) including all original emails, minutes of phone conversations, and attachments, and is categorized as to nature (i.e., inquiry, submission of data, metadata, project information, requested edits, etc.). Submissions are dated and processed in first-received order. Data submission tickets are linked internally to each publicly served dataset metadata page. All dataset edits are also tracked within each dataset ticket.

Submitters are informed of the full collaborative data management process on our website and are directed to our resources page for instructions, and an overview of working with office (see BCO-DMO Resources webpage: https://www.bco-dmo.org/how-get-started and the BCO-DMO Quick Start Guide: https://www.bco-dmo.org/files/bcodmo/BCO-DMO_Quick_Start_Guide.pdf).

As data are processed and metadata are assembled, detailed notes are tracked in the dataset-specific ticket and in README files captured in metadata, including differences between versions. Software processing steps generate checksums and capture provenance in a structured (.yaml) file accompanying each data package. Any change to a file will trigger an event that will record provenance, thereby alerting staff to unexpected changes. All edits are approved by submitters via a request for validation confirming completeness and correctness. As datasets are published online, dataset versions are tracked within a structured metadata field. Strategies determining need for major and minor revisions are documented in internal Standard Operating Procedures (SOPs) and communicated to submitters during processing. Each version is published with a DOI and deposited in both BCO-DMO's Institutional data library and with the appropriate national archive (described in section III, Continuity of Access, R3.).

The identity of data submitters is confirmed via a number of processes. It is first matched against the NSF award database using an API to validate award-related metadata information. Open Researcher and Contributor Identification (ORCiD) numbers are also requested upon submission (in metadata forms) and are validated against the ORCiD database and

corrected as needed. Roles are then assigned (using standardized vocabularies) to all persons identified on the metadata forms and are included in published metadata and DOIs.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

VIII. Appraisal

R8. The repository accepts data and metadata based on defined criteria to ensure relevance and understandability for data users.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

The BCO-DMO collection is guided by its mission and project objectives to curate only those data that fall under the research direction of the U.S. National Science Foundation Biological and Chemical Oceanography sections, and Office of Polar Programs (i.e., data from funded investigators within the same NSF sections that fund BCO-DMO). Investigators from these funding sections are prescribed to use BCO-DMO as a result of award compliance criteria and are encouraged to begin contributing information to the repository as early as being recommended for funding (see NSF OCE Sample and Data Policy (NSF17-037, 2017), at https://www.nsf.gov/pubs/2017/nsf17037/nsf17037.jsp):

“The Biological and Chemical Oceanography Data Management Office (BCO-DMO) is the primary data management archive for the Biological Oceanography and Chemical Oceanography programs, as well as several associated special programs. When awards are initialized, investigators should immediately contact BCO-DMO and register their projects by

submitting project metadata. For projects where data cannot be served by BCO-DMO, or where they are more appropriately served by other community data repositories, metadata should still be deposited in BCO-DMO with links to the other data repositories."

For investigators wishing to deposit data into the BCO-DMO catalog who are not funded by the NSF OCE sections above, BCO-DMO assesses each request to determine if the project and resultant data are within the same domains and relevant to BCO-DMO's ocean ecosystem community. Those investigators whose data are better served through an alternate repository are directed to appropriate sources of information, including Other Suggested Data Management Centers on the BCO-DMO Resources page, and the Registry of Research Repositories (re3data.org).

Each submission undergoes an initial assessment by BCO-DMO staff of package components upon receipt before being transferred to the internal ticketing system for tracking. Submissions are then accepted by data managers for processing, and data and metadata files are then further reviewed for completeness by the BCO-DMO data manager assigned to the submission. If required fields are missing in metadata and/or data, the data manager will request additional content or return the submission for editing/completing. This appraisal is on a submission-by-submission basis and frequently depends upon the degree of incompleteness.

Data holdings are highly heterogeneous and encompass the 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. Data are from marine, limnological, interstitial, aerosol studies; and all sub-domains of oceanography with methodologies varying among subdomains and even laboratories. Therefore, assessment of data file completeness depends highly upon the knowledge of the data manager and their close interaction with the submitting researcher to confirm all the data that are intended to be shared are complete and well documented. This is a highly collaborative process and one of BCO-DMO's strengths (and reason for hiring domain-knowledgeable data managers). Data are run through internal processing software, where high level routine quality checks are performed (e.g., ill-formed tabular formatted files, missing values, text present in numeric fields, etc.), in addition to providing statistical summaries that allow data managers to check for out of range values, and/or erroneous precisions, etc. Routines run are documented in provenance records included in the data package, as well as within a README file and documented in dataset metadata.

BCO-DMO recognizes the variability in metadata needs across the diverse set of research communities it serves under the broad umbrella of Biological and Chemical Oceanography. In an effort to remain fluid to those needs, but ensure sufficient metadata is captured, BCO-DMO collects structured metadata, informed by existing metadata standards and domain practices to ensure that each discipline being served has well-curated, thorough metadata for findable, interoperable and reusable data. (see forms linked to BCO-DMO Resources webpage: https://www.bco-dmo.org/how-get-started). If fields within the forms are missing or sparse, the data manager will request additional information, and/or use submitted publications to extract relevant methodology sections for acquisition and analyses.

BCO-DMO's metadata model joins concepts from Dublin Core, Dataset Catalog Vocabulary (DCAT), Friend of a Friend

(FOAF), Provenance Data Model (PROV-DM), OGC GeoSPARQL, Simple Knowledge Organization System (SKOS), Vocabulary of Interlining Datasets (VoID), OWL-Time Ontology, Schema.org, Bibliographic Ontology (BIBO) and the Ocean Data Ontology (ODO, available from https://github.com/BCODMO/Ocean-Data-Ontology/blob/master/README.md). Together, these vocabularies are used to generate RDF resources which are then transformed into metadata standards specific to certain use cases. Currently, the standards generated at BCO-DMO include ISO19115-2, Dublin Core, DCAT, Schema.org, Crossref Metadata Kernel (DOIs), Frictionlessdata Data Package, and Directory Interchange Format (DIF). When new metadata uses cases arise, we align that metadata model to our internal model, and develop any identified gaps within the Ocean Data Ontology. From there, our internal metadata can then be translated into the new metadata standard.

The office also produces a data management best practices guide for community use that includes domain-specific guidance for pre-data collection (https://www.bco-dmo.org/data-management-best-practices-guide). The office website provides guidance on data formats and commonly submitted file formats (https://www.bco-dmo.org/how-get-started). Data are accepted in any format, but wherever possible endeavors to serve data in a non-proprietary format (i.e., comma- and tab-separated, .csv, .tsv). Non-tabular data (i.e., image, NetCDF, sequence data, etc.) may be served in its native format, but when possible displayed as a tabular (.tsv) list of downloadable file names along with associated parameters. These data are accepted in their native format, and made accessible through the existing download capabilities of the current website in addition to the newly implemented ERDDAP server (described in section XIV. Data Reuse). When provided, descriptions of relevant format-specific tools (for visualizing and analyzing the data) are included in the dataset metadata. Data that are submitted in proprietary formats (e.g., Excel,Matlab, etc.) are extracted or exported and served as non-proprietary .tsv format for online publication. When requested, original file formats are retained and served alongside non-proprietary formats.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

IX. Documented storage procedures

R9. The repository applies documented processes and procedures in managing archival storage of the data.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

Data Management processes (all activities performed on data and metadata) are documented in internal Standard Operating Procedure (SOP) documents accessible to all staff and used for training. These are held in an internal WHOI Enterprise Google Folder and are revised as necessary to accommodate infrastructure changes, as well as funder, submitter and user needs. SOP documents are broken up into large sections covering key concepts, data processing, data ticketing, metadata entry, data access tool, internal FAQ's, and office processes. Individual topics range from repository philosophy, background, mission and vision, community and funding partners, and data management planning basics; to instructions for staff interactions with data submitters and other stakeholders (including email templates for routine communications), and how to triage a data set and add it to the data processing ticketing system; to instructions on using internal data processing tools and inputting metadata into the system; to loading completed datasets into the data access tool; to how our office conducts routine activities such as meetings, and Google Calendar and Drive use. (public facing documentation pages are in the process of being migrated to Gitbooks for ease of publishing but samples may be made public upon request)

All data managed by BCO-DMO are public after a 2 year optional embargo period as requested by the author. Any embargo period beyond the two year allowable embargo must be negotiated between the investigator and their NSF Program Manager. During a requested embargo period the office maintains public metadata for discovery.

The WHOI Information Services (IS) provides a virtual machine environment and large research data storage, including local tape backup service (https://whoi-it.whoi.edu/virtual-servers/ and https://whoi-it.whoi.edu/central-data-storage/). This data store serves to store the BCO-DMO datasets and related content. A dedicated, network-accessible storage unit is located in Boston, MA to provide secure and off-site backup of all the data and metadata stored on the BCO-DMO servers at WHOI (see Markley Group in Outsource Partners section). This backup is in addition to local system backups that are maintained and updated daily at WHOI's IS Department. Risk management assessment of storage and backups is performed routinely within funding cycles and strategies are adjusted accordingly (e.g., due to increasing costs, we plan to migrate this offsite backup to the commercial cloud).

The IS Security team has configured all machines with monitoring and mitigation tools for detecting hardware failures and security intrusions to ensure that all WHOI systems are protected by a multi-layer defense strategy, while safely allowing outside access to public services. IS ensures compliance with Defense Federal Acquisition Regulation (DFAR) and National Institute of Standards and Technology (NIST) guidelines. The IS department also monitors unusual activity and server access.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

X. Preservation plan

R10. The repository assumes responsibility for long-term preservation and manages this function in a planned and documented way.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

BCO-DMO is an intermediate data facility supported through a project awarded by the US NSF, and as such may not serve as a true archive, providing guaranteed long-term access to its data catalog. To ensure its data catalog persists, BCO-DMO employs a redundant preservation strategy that includes its Institutional Library's Data Archive (WHOAS) and the US National Oceanographic and Atmospheric Administration (NOAA's) National Centers for Environmental Information (NCEI); see Section III Continuity of Access. Data packages, once complete, quality checked, and validated (approved) by the submitter, are submitted to WHOAS and NCEI for preservation. BCO-DMO has a Submission Agreement in place with NCEI that outlines its relationship and process.

By virtue of its NSF award and resulting mission, the repository is expected to assume custody, copy, transform, publish, store and submit for archive, all digital project output from NSF-funded projects awarded through the relevant NSF Oceanography Divisions under the repository's purview. BCO-DMO performs these actions under the consent of the project awardees who are required to deposit data with BCO-DMO as a result of their funding criteria outlined in the NSF OCE Sample and Data Policy (https://www.nsf.gov/pubs/2017/nsf17037/nsf17037.jsp) and the NSF OCE guidance

webpage listing approved data repositories (https://www.nsf.gov/geo/oce/oce-data-sample-repository-list.jsp). All digital data objects receive the same level of curation. Data submitters are directed to the repository's Terms of Use page (https://www.bco-dmo.org/terms-use) outlining expectations of data submitters to grant BCO-DMO the right to copy, transform, store, and share the submitted data and information.

Data submitters are directed to BCO-DMO's webpage for instructions and expectations of how to engage with the repository (https://www.bco-dmo.org/how-get-started) this outlines instructions for submission and contains general information on the process of collaborating with the repository for data sharing.

Data are contributed to the repository for the purposes of publishing and contributing to the National Archive (NCEI) for long-term preservation. BCO-DMO data managers work closely with each submitter to arrive at a final data representation that satisfies their sharing criteria and best suits the submitters specific research community. It is communicated and understood through this collaboration that data may undergo transforms and quality control activities during this process. BCO-DMO directs submitters to its Quick Start Guide (https://www.bco-dmo.org/files/bcodmo/BCO-DMO_Quick_Start_Guide.pdf) that includes very detailed instructions on data submission requirements and processes (such as preparing metadata, what will happen once data are submitted, expectations on behalf of the submitter, what is meant by a "dataset" and what transforms may be undertaken to arrive at a publishable dataset within the BCO-DMO repository).

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

XI. Data quality

R11. The repository has appropriate expertise to address technical data and metadata quality and ensures that sufficient information is available for end users to make quality-related evaluations.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

BCO-DMO is a domain-specific repository, which relies on highly trained staff who possess both formal domain science training (e.g., Master's degrees in physical, biological and chemical oceanography), hands-on experience from fieldwork, and data management training, making them familiar with many of the data types being curated. Staff work collaboratively with submitters to complete all necessary metadata fields required for sufficient discovery and reuse (see metadata forms available at: https://www.bco-dmo.org/how-get-started). In addition, related publications and supplemental documentation are assembled and linked to primary dataset metadata to provide additional context for end-users. Related datasets within the catalog are also linked to facilitate discovery of complementary content.

Software used to serve data files online contain automated error checking of tabular data (e.g., ill-formed tabular formatted files, missing values, text present in numeric fields, etc.), in addition to providing statistical summaries that allow data managers to check for out of range values, and/or erroneous precisions, etc. Errors and issues are resolved collaboratively and changed with permission of the submitter. All edits are documented in provenance records, README files and incorporated into dataset metadata. Known errors or omissions in contributed datasets that are identified by submitters are captured in methodology metadata sections (i.e., known sensor failures, instrument drift, etc.). BCO-DMO honors quality flags when presented by the submitter, but does not apply data quality flags as a matter of course. BCO-DMO does serve to liaise and adjudicate user feedback identifying data errors and will attempt whenever possible to resolve and update files with the original submitter (see versioning in section VII. Data integrity and authenticity).

Once assembled, the data manager requests that the data package be validated (i.e., reviewed and approved) by the submitter to ensure completeness and accuracy. Upon successful validation the dataset is considered ready for publication.

Once a data manager requests publication of a dataset, internal software performs quality assessment of data and metadata, and schema validation for various metadata formats. Any final corrections to package contents are made to obtain a version of the data ready for deposition into the institutional library data archive (WHOAS) and the appropriate national archive (NCEI).

In the event that errors in data are identified by an enduser but the originating submitter is no longer available to provide input on identified errors, an order of operations is set in place that includes: 1) contacting other project-related roles ranked by internal hierarchy (e.g., Co-PI, contact, technician, etc.) and attempt to resolve; 2) in the event of a single PI/submitter role, we then research possible solutions with the individual who identified the error/edits and resolve accordingly, documenting any changes (including contact information of the identifier) and versioning as necessary; 3) in the event that a resolution is not possible (i.e.,errors are known to exist, but we are unable to resolve or correct), we

identify known errors, change to missing values where/when relevant, document all actions taken and version as necessary.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

XII. Workflows

R12. Archiving takes place according to defined workflows from ingest to dissemination.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

Data Management processes (all activities performed on data and metadata) are documented in internal Standard Operating Procedure (SOP) documents accessible to all staff and used for training. These are held in an internal WHOI Enterprise Google Folder and are revised as necessary to accommodate infrastructure changes, as well as funder, submitter and user needs. Changes to workflow SOPs are discussed as an entire group (administration to data managers) and captured in Architectural Decision Records (ADRs) located in project Github repositories and/or alongside SOPs in Google Drive. These SOP documents are in the process of being migrated from Google Drive to Gitbooks in order to be published and linked to our website.

Summary of workflow follows below:

BCO-DMO staff assist investigators as early as their proposal preparation stage to construct descriptive Data Management Plans (DMPs). A DMP template was created by BCO-DMO and NSF to help streamline this process. BCO-DMO endeavors to accept all data types (e.g., tabular, image, sequence, spectral, acoustic, etc.), and file sizes in an effort to accomodate investigator needs. Then, as investigators are recommendation for funding, they are instructed by their NSF Program Manager to reach out to BCO-DMO as per the OCE Sample and Data Policy. Expectations on use are described in BCO-DMO's Terms of Use webpage and clear instructions and description of BCO-DMO repository processes are described on the Resources webpage (https://www.bco-dmo.org/resources), which all new submitters are directed to. This page includes links to instructions on producing data management plans, how to submit data, acceptable formats, preparing data, and frequently asked questions links to data citation instructions and terms of use. This page also directs submitters to alternate data repositories that are endorsed by NSF OCE funders (should they be supported by NSF awards). The office assists those individuals not funded by NSF in determining an appropriate repository through conversations in email or by phone (see section VIII. Appraisal)

Once data are submitted, data managers confirm submissions and enter all files and correspondence into a ticketing tracking system and processing queue (Redmine) on secure, private servers (see sections XVI on Security). BCO-DMO does not serve nor intend to serve data associated with disclosure risk, and submitters are subject to the ethical requirements instituted by his or her institution, funder, and discipline (see section IV Confidentiality/Ethics for determination of sensitive data).

Staff assess each file for completeness, perform gross quality control and reformatting as needed, and work iteratively with data submitters to organize and display data in the most appropriate manner for each data type (see section XI Data Quality). Both commercial and personally scripted software are used to process data (internal processing scripts are managed in Github). Data processing workflows, notes and provenance are captured, and stored using the Provenance Data Model (PROV-DM), incorporated into dataset metadata. Staff collect, assemble, and enter into the database robust metadata necessary to discover, understand, and reuse the data, often reading and extracting contextual information from related publications and reports. In addition, related publications and supplemental documentation are assembled and linked to primary dataset metadata to provide additional context for end-users. Related datasets within the catalog are also linked to facilitate discovery of complementary content.

Once processing is completed, staff ensure all content is validated with the data contributor, link data to metadata, and publish the full data package version online with a usage license (Creative Commons CCBy4) and DOI. To satisfy data management best practices of redundancy and versioning, a copy of the final curated package is stored locally, with BCO-DMO's institutional library, and deposited at an appropriate national archive (e.g., NCEI) for long-term preservation (see section III Continuity of Access).

BCO-DMO assumes data are public unless embargo is requested by submitter to accommodate publication requirements. Public data is open to all users and requires no authentication to access these resources (no personally identifiable information is collected or stored). Embargoed data are denied access to all unless the author(s) have requested it be password protected. We are transitioning password protection from web server authentication of defined

username/password combinations to oAuth access through external identity management providers such as ORCiD and Google. This new capability will be enforced through our data access server (Environmental Research Division Data Access Protocol (ERDDAP), https://coastwatch.pfeg.noaa.gov/erddap/index.html) relieving BCO-DMO of having to manage passwords securely. Per the NSF Sample and Data Policy (NSF17-037, 2017), embargoed data are given 2 years from the time of collection until they are expected to be made publicly available, and NSF retains the right to reject any request for an embargo or embargo extensions.

Data are stored on secure servers located at WHOI behind Institutional firewalls (see section IX Documented Storage Procedures, and XVI. Security for detail).

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

XIII. Data discovery and identification

R13. The repository enables users to discover the data and refer to them in a persistent way through proper citation.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

BCO-DMO provides free access to oceanographic data through a web-based system that includes text and geospatial interfaces with tools and features facilitating assessment of fitness for purpose (see home page at https://bco-dmo.org). The BCO-DMO metadata catalogue is stored in a standards agnostic database, but is exported in a number of

international and community standards: ISO19115-2, Dublin Core, DCAT, Schema.org, Crossref Metadata Kernel (DOIs), Frictionlessdata Data Package, and Directory Interchange Format (DIF). Elasticsearch is used to power all search capabilities on the BCO-DMO website. All datasets are contain citable URI and suggested citations (see below).

BCO-DMO supports Schema.org harvesting via sitemap.xml, the DataONE API, a SPARQL endpoint accessed via the /.well-known/void URL, and can support OAI-PMH as needed (however the repository has never had use cases/requests for it).

BCO-DMO is registered in the Registry of Research Data Repositories (Re3Data.org, https://www.re3data.org/repository/r3d100000012), EarthCube Resource Registry (e.g., https://www.earthcube.org/webapps/geocodes/discovery/ui/textSearch.html), FAIRSharing (https://fairsharing.org/biodbcore/?q=bco-dmo) and COPDESS' Datacite Repository Finder tool (https://repositoryfinder.datacite.org/).

BCO-DMO assigns Digital Object Identifiers to all of its contemporary datasets, and is in the process of assigning DOIs to all of its legacy datasets as well (anticipated completion by EOY 2019). BCO-DMO suggests an appropriate citation (APA format) on each dataset's metadata landing page that includes the dataset's DOI. BCO-DMO's also directs users to review Terms of Use (https://www.bco-dmo.org/terms-use) for licensing and citation information and instructions.

The Marine Biological Laboratory-Woods Hole Oceanographic Institution Library, (also Core Trust Seal certified; https://darchive.mblwhoilibrary.org/) is used for acquisition of Digital Object Identifiers (DOIs). Once data are processed, data and metadata are validated by the submitter, and the package is checked for errors/inconsistencies, an automated system sends the package to the library for DOI minting. The library maintains the immutable copy of each version of data for which a DOI is obtained, and the DOI is referenced in the dataset metadata.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

XIV. Data reuse

R14. The repository enables reuse of the data over time, ensuring that appropriate metadata are available to support the understanding and use of the data.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

BCO-DMO recognizes the variability in metadata needs across the diverse set of research communities it serves under the broad umbrella of BIological and Chemical Oceanography. In an effort to remain fluid to those needs, BCO-DMO collects detailed use-metadata in a standards-agnostic format, specific to its community research needs and informed by existing metadata standards and domain practices. BCO-DMO's metadata model joins concepts from Dublin Core, Dataset Catalog Vocabulary (DCAT), Friend of a Friend (FOAF), Provenance Data Model (PROV-DM), OGC GeoSPARQL, Simple Knowledge Organization System (SKOS), Vocabulary of Interlining Datasets (VoID), OWL-Time Ontology, Schema.org, Bibliographic Ontology (BIBO) and the Ocean Data Ontology (ODO). Together, these vocabularies are used to generate RDF resources which are then transformed into metadata standards specific to certain use cases. Currently, the standards generated at BCO-DMO include ISO19115-2, Dublin Core, DCAT, Schema.org, Crossref Metadata Kernel (DOIs), Frictionlessdata Data Package, and Directory Interchange Format (DIF). When new metadata uses cases arise, we align that metadata model to our internal model, and develop any identified gaps within the Ocean Data Ontology. From there, our internal metadata can then be translated into the new metadata standard.

Initial capture of metadata from the submitter is accomplished through the use of forms for describing oceanographic research Programs, Projects, Datasets, and Deployments (found online: https://www.bco-dmo.org/how-get-started). BCO-DMO does accept synthesis products and can capture provenance via existing metadata infrastructure. See section II Licensing, for description of data reuse rights.

Observational data files are submitted to the repository by investigators in various native formats, many of which are generated by proprietary software (e.g., Excel, Matlab, etc.). Processing that occurs at the repository includes corrections, extraction, and conversion to non-proprietary formats (.tsv, .csv) for serving online. This ensures broad accessibility of data in non-proprietary formats. BCO-DMO currently provides access to download data in a variety of formats (via the Downloads and Other Operations tools available on each dataset view e.g., http://optserv1.whoi.edu/jg/otheropt3//data.bco-dmo.org:80/BCO/GEOTRACES/Arctic/trace_elements_shiller.html1%7Bdir=data.bco-dmo.org/jg/dir/BCO/GEOTRACES/Arctic/,info=data.bco-dmo.org/jg/info/BCO/GEOTRACES/Arctic/Trace_elements%7D?), but is transitioning to a new open source data access tool called the Environmental Research Division Data Access Protocol (ERDDAP, https://coastwatch.pfeg.noaa.gov/erddap/index.html) to direct users to alternate formats for data download. ERDDAP

provides download to several oceanographic domain data formats (matlab, NetCDF, Ocean Data View, etc.) and is in active development with the oceanographic community (see BCO-DMO's implementation here https://erddap.bco-dmo.org/erddap/info/index.html?page=1&itemsPerPage=1000). As new data formats emerge, BCO-DMO will continue to work with the ERDDAP community to adapt its capabilities to meet community needs.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

TECHNOLOGY

XV. Technical infrastructure

R15. The repository functions on well-supported operating systems and other core infrastructural software and is using hardware and software technologies appropriate to the services it provides to its Designated Community.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

Technical Infrastructure

BCO-DMO centers its technical infrastructure around a knowledge graph of Resource Description Framework (RDF) resources that describe data and metadata resources. Sitting on top of the knowledge graph are APIs and queryable

interfaces that power data access mechanisms and website content, and other user interfaces. This strategy helps remove any dependencies on a particular software stack for data and metadata to be collected, and thereby provides BCO-DMO with the flexibility to replace software components in its architecture without impacting the underlying data model. The knowledge graph is stored in an open-source Virtuoso RDF triplestore. Currently, the website is powered by Drupal 7 and access to data is provided by ERDDAP, a data server for oceanographic datasets developed by NOAA (https://coastwatch.pfeg.noaa.gov/erddap/index.html). ERDDAP is currently replacing an historic JGOFS Data Server, the genesis of the OpenDAP protocol. Elasticsearch is used to power all search capabilities on the BCO-DMO website. Redis serves as our cache server for improving API and query performance.

Figure 1. The BCO-DMO cyberinfrastructure: https://drive.google.com/open?id=1rsg7jF3jFKZc5CcI1F-BSCT-glFR5kC-

All BCO-DMO software is deployed using Docker containers, and all custom software development is managed in Github (https://github.com/BCODMO/) some repositories are marked private on Github for security reasons. This overall strategy is the result of periodic assessment and evaluation which happens through funding cycle proposals and mid-project review required by the NSF funding agency, and with the assistance of BCO-DMO's Strategic Planning Committee.

Metadata Standards

BCO-DMO recognizes the variability in metadata needs across the diverse set of research communities it serves under the broad umbrella of BIological and Chemical Oceanography. In an effort to remain fluid to those needs, BCO-DMO collects metadata, informed by some of the existing metadata standards and domain practices to ensure that each discipline being served has well-curated, thorough metadata for findable, interoperable and reusable data. BCO-DMO's metadata model joins concepts from Dublin Core, Dataset Catalog Vocabulary (DCAT), Friend of a Friend (FOAF), Provenance Data Model (PROV-DM), OGC GeoSPARQL, Simple Knowledge Organization System (SKOS), Vocabulary of Interlining Datasets (VoID), OWL-Time Ontology, Schema.org, Bibliographic Ontology (BIBO) and the Ocean Data Ontology (ODO). Together, these vocabularies are used to generate RDF resources which are then transformed into metadata standards specific to certain use cases. Currently, the standards generated at BCO-DMO include ISO19115-2, Dublin Core, DCAT, Schema.org, Crossref Metadata Kernel (DOIs), Frictionlessdata Data Package, and Directory Interchange Format (DIF). When new metadata uses cases arise, we align that metadata model to our internal model, and develop any identified gaps within the Ocean Data Ontology. From there, our internal metadata can then be translated into the new metadata standard.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

XVI. Security

R16. The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users.

Compliance Level:

4 – The guideline has been fully implemented in the repository

Reviewer Entry

Reviewer 1

Comments:
4 – The guideline has been fully implemented in the repository

Reviewer 2

Comments:
4 – The guideline has been fully implemented in the repository

Response:

The BCO-DMO repository sits within the Woods Hole Oceanographic Institution where it is able to take advantage of existing Institutional infrastructure for security and appropriate access controls. Only BCO-DMO staff are allowed access to the data stores and no outside upload of content is allowed at this time. However, BCO-DMO is in the development stages of an end-user submissions system, at which time user-uploaded content will remain in a designated quarantine area accessible by staff during the appraisal stage. Varying levels of systems administration access are granted as per the project Technical Director, who is the designated liaison between the project and the appropriate Institutional Information Services staff. Risk assessment of storage and backups is performed routinely within funding cycles and strategies are adjusted accordingly.

The WHOI Information Services (IS) Department provides a virtual machine environment and large research data storage, including local tape backup service. The Security team at IS has configured all machines with monitoring and mitigation tools for detecting hardware failures and security intrusions to ensure that all WHOI systems are protected by a multi-layer defense strategy, while safely allowing outside access to public services. IS ensures compliance with Defense Federal Acquisition Regulation (DFAR) and National Institute of Standards and Technology (NIST) guidelines. The IS department also monitors unusual activity and server access.

A dedicated, network-accessible storage unit is located in Boston, MA to provide secure and off-site backup of all the data and metadata stored on the BCO-DMO servers at WHOI. This backup is in addition to local system backups that are maintained and updated daily at WHOI's IS Department. Due to increasing costs and greater offered protections, we plan to migrate this offsite backup to the commercial cloud. All local machines are connected to the Institution's TCP/IP based network. The Institution is connected to the Internet at Internet 2 and T1 (as backup) speeds. A Disaster Recovery Plan is in place by IS to speedily replicate the BCO-DMO technical infrastructure on new virtual machines, running the BCO-DMO Docker containers, and mounted data storage.

Reviewer Entry

Reviewer 1

Comments:
Accept

Reviewer 2

Comments:
Accept

APPLICANT FEEDBACK

Comments/feedback

These requirements are not seen as final, and we value your input to improve the core certification procedure. To this end, please leave any comments you wish to make on both the quality of the Catalogue and its relevance to your organization, as well as any other related thoughts.

Response:

Reviewer Entry

Reviewer 1

Comments:

Reviewer 2

Comments: