ADOPTING RDA OUTPUTS FOR...
OCEANOGRAPHY

RDA

RESEARCH **DATA** ALLIANCE

BCO-DMO adopts four RDA recommendations for marine biogeochemical and ecological research

BCO-DMO is the Biological and Chemical Oceanography Data Management Office. BCO-DMO serves as a facility where marine biogeochemical and ecological data and information developed in the course of scientific research can easily be disseminated, protected, and stored on short and intermediate time-frames. The Data Management Office also provides research scientists and others with the tools and systems necessary to work with marine biogeochemical and ecological data from heterogeneous sources with increased efficacy. The BCO-DMO data system can accommodate many different types of data including biological, chemical, and physical measurements and results.

The system provides access to research-quality data (numbers, images, and/or documents) with sufficient metadata, so that others can make full use of these data for their own purposes. The existence of sufficient metadata enables the discovery and accurate reuse of data by more than just the initial investigators who collect and process the data.

Adopting RDA's recommendations helped BCO-DMO to streamline their process efficiency in their data pipelines and reduce duplication of effort or making assumptions (without having a PID record) about what a particular resource in their system was at any point during the data life cycle.

The Challenge

"BCO-DMO has been curating research data for over a decade and is currently operating on a 5-year grant. As the landscape of data management best practices changes within a funding cycle, it can often be hard to make major systemic changes to workflows and infrastructure."

Says Adam Shepherd, Technical Director at BCO-DMO, the adopting organisation.

BCO-DMO as an intermediary data facility was structured to operate in the capacity of transporting data from researchers to the archives ensuring quality metadata for reuse. As data publication and citation needs changed, BCO-DMO realized that it needed to support these efforts as part of its service to the research community. BCO-DMO's existing system allowed researchers to cite data and subsets of the data, but did not have a mechanism for managing those citations as new versions of the dataset became available.

RDA RECOMMENDATIONS ADOPTED

Dynamic Data Citation: supports accurate citation of data subjected to change, for the efficient processing of data and linking from publications.

Data Foundation & Terminology: ensures researchers apply a common core data model when organising their data, thus making data accessible and re-usable.

Data Type Registries: ensures data producers classify their data sets in standard data types, allowing data users to automatically identify instruments to process and visualise the data.

PID Information Types: defines standard core PID information types to enable simplified verification of data identity and integrity.

WHY RDA

We improved data versioning and identification techniques for all data resources with capabilities for supporting dynamic data resources in the future.

We codified the terms and definitions we used to describe the data and information we manage, and publish this as an application ontology that provides the backbone for BCO-DMO data pipelines.

We built ontology models for describing a data type and its required measurements. We publish PID records as Linked Data for all resources in the repository. Dataset records include data citation information, versioning histories, and any valid data type designations. These PID records use terms from our ontology and other vocabularies.

ANSWERING COMMUNITY NEEDS

Without making these adoptions it was possible for citations to be made that didn't specify the version of the data being cited. This could potentially pose problems for datasets where the responsible researcher submits updates or corrections to a dataset after a citation has been generated.

Find out more

Visit RDA @ rd-alliance.org
Email: enquiries@rd-alliance.org

The Adoption

BCO-DMO has been building pipelines to improve data transport workflows from researcher contributions to the repository and from repository to long-term archives and aggregation services. While already employing Linked Data best practices for identification of resources in the repository, BCO-DMO modified their workflows and infrastructure to improve data identification throughout transport life cycles. This included adopting outputs from PID Information Types, Data Foundation and Terminology, Data Type Registries and Dynamic Data Citation Working Groups to add data versioning support for improved accuracy of citations and data immutability as data move through these pipelines. Adopting these recommendations qualified BCO-DMO to more easily adopt the APIs set forth by DataONE for inclusion as a Member Node of their federated data network.

Our adoption effort helped us improve how we model persistent identifiers, the multiple versions of a dataset, and the immutable files and information associated with those specific versions, or snapshots in time. With the concept map and ontology in place, we started to alter/build software to improve the workflows of our office's data managers, streamlining the pipeline from BCO-DMO to long-term archives and aggregation services. The software developed lets data managers decide in concert with the submitters, when a data product is ready to be labelled and final and complete with respect to its format and metadata records. At this point, the new software submits the data and related information resources to an institutional repository, the Woods Hole Open Access Server at Woods Hole Oceanographic Institution, where it receives a DOI. Once the DOI is minted and resolves, the software can then send an updated data package including its DOI, to a long-term archive and also to aggregation services like DataONE. Aside from the BCO-DMO website, researchers can discover the data at both of these locations and cite the data appropriately.

Lesson Learnt

Being open-minded when we approached the recommendations helped us determine how best to adopt the recommendations within the scope and mission of our project. We adopted as many of the recommendations as made sense to our service in ways that meshed with our existing efforts in data transport and linked data publishing.

The Biological and Chemical Oceanography Data Management Office

BCO-DMO is a thematic, domain-specific repository created to serve principal investigators funded by the National Science Foundation Biological and Chemical Oceanography Sections and Division of Polar Programs Antarctic Organisms & Ecosystems Program.

Name: Adam
Surname: Shepherd
Qualification: Technical Director
Country: USA
Email: ashepherd@whoi.edu