U.S. NATIONAL SCIENCE FOUNDATION
2415 EISENHOWER AVENUE
ALEXANDRIA, VIRGINIA 22314

NSF 24-124

Division of Ocean Sciences (OCE) Sample and Data Policy

This document replaces NSF 17-037, Division of Ocean Sciences Sample and Data Policy, December 15, 2016.

September 4, 2024

All NSF proposals must include a Data Management and Sharing Plan that describes what data/samples will be collected, what analyses will be done, and how the project will provide open and timely access to meta data, data, preserved or archived samples, derived products (e.g., models and model output), and other information on the project during and after the project's completion. The Data Management and Sharing Plan also must specifically discuss how the investigators will achieve the specific OCE data archiving and reporting requirements described below in this document. If the project is not expected to generate new data, samples or derived data products, the Data Management and Sharing Plan can include a statement that no detailed plan is needed, accompanied by a clear justification. See the NSF Proposal & Award Policies & Procedures Guide (PAPPG), Chapter II.D.2 for additional information.

NSF Principal Investigators (PIs) are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections, meta data, and other supporting materials created or gathered in the course of work under NSF grants.

DATA AND SAMPLE ARCHIVING REQUIREMENTS

The Division of Ocean Sciences requires that meta data files, full data sets, derived products and physical collections must be made publicly accessible upon publication, or within two (2) years of collection, whichever comes first. This includes software and derived data products (e.g., model results, output, and workflows). A brief description of preferred data and physical collection archives and centers and their criteria for submission can be found on the OCE

[website](#) or through contact with the cognizant Program Officer of the given award. Data should be in a format that is easily accessible; data output in a format that is not readily accessible should be accompanied by a readme file and a script to allow reading it. Any limit on access to original data, samples, or other information beyond the two-year moratorium period must be based on compelling justification, documented in the Data Management and Sharing Plan of the proposal, or approved by the cognizant Program Officer. Such exceptions have been, and are likely to remain, rare.

Derived products related to numerical modeling include model output as well as newly developed software code (programming code, model configuration code, and analysis scripts). It is expected that raw output from high resolution modeling or large ensemble simulations will be very large, while processed output or model output from finalized model versions may be smaller and more suitable for public archiving. The Data Management and Sharing Plan should anticipate these different output types and describe a plan for their management. Model output used in peer-reviewed publications should be made publicly accessible at the time of publication. Code and scripts should also be archived at the time of publication in a research code repository. If the newly developed code improves an existing numerical model, an accompanying readme file should point to the original code, which should be archived and publicly available already, and describe how the new code fits into the original code.

Where no disciplinary, institutional, or generalist data or sample repository or archive exists for collected original data and sample types, the PI is required to identify a preservation plan in the Data Management and Sharing Plan that complies with the general philosophy of sharing research products and data within two years of collection as described above.

For proposals involving biological specimens, the Data Management and Sharing Plan must include a description of how the specimens and associated data will be maintained in the PI's laboratory and accessioned into and maintained in a biological collection.

REPORTING REQUIREMENTS

PIs are required to provide updates on the status of meta data and data archiving in Annual and Final Annual Project Reports according to their Data Management and Sharing Plan, which must be compliant with this OCE sample and data policy. If not deposited in an approved federally or NSF-funded repository, a statement of justification for the use of alternate data sharing methods, with associated Persistent Identifiers (e.g., DOI's) for published data and meta data should be included.

URLs for archived meta data and data should be included in each report in the section entitled "Products-Websites."

If the Final Annual Project Report is due before the required date of sample or data

submission, the PI must report submission of meta data and plans for final data/sample submission. The PI should notify the cognizant Program Officer by e-mail after final data and/or sample submission has occurred, even if this is after the expiration date of the award. The ultimate disposition of data and samples must be described in the "Results from Prior NSF Support" section of future proposals submitted by the PI, as per guidelines in PAPPG Chapter II.D.2.

SPECIAL PROGRAM GUIDANCE FOR OCE

The Rolling Deck to Repository (R2R) program ensures that routine underway data from the fleet of academic research vessels are preserved and accessible, including submission to the National Oceanic and Atmospheric Administration National Centers for Environmental Information (NOAA NCEI) where appropriate. Routine underway data refer to instrumentation that is usually permanently installed on the vessel and managed by operators. R2R receives these data directly from the operators; PIs do not need to submit them to R2R or to NCEI, but do need to respond to R2R requests to release these data in a timely way. Data collected by instruments that are not part of the routine underway instrumentation, or processed versions of data made from the underway data, are not managed by R2R and are the responsibility of the PI/science party to submit to an appropriate repository.

Visual imagery data collected by assets of the National Deep Submergence Facility (NDSF which at the present time includes the submersible Alvin, ROV Jason and AUV Sentry) are archived at the Woods Hole Oceanographic Institution data repositories.

The Marine Geology and Geophysics (MGG) Program encompasses a wide range of data types and data formats. If you are unsure which repository is most appropriate for your digital data and/or data product, please check with a MGG Program Officer during development of your Data Management and Sharing Plan. MGG anticipates that most geological samples will be archived at NSF-approved repositories (https://www.nsf.gov/geo/oce/oce-data-sample-repository-list.jsp). Any analyses tied to a physical sample must follow the data center/repository guidelines for archiving (https://www.nsf.gov/geo/oce/oce-data-sample-repository-list.jsp).

Several Climate Variability and Predictability Program (CLIVAR) activities, like the Carbon/Global Hydrographic Survey, require PIs to submit data collected to the CLIVAR and Carbon Hydrographic Data Office (CCHDO) within six (6) months of collection. CTD and bottle data collected under the GO-SHIP program must also be submitted to CCHDO on the schedule defined by GO-SHIP for Level I and Level II parameters.

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. These programs recommend that PIs use the BCO-DMO-developed template (accessible through the BCO-

DMO DMP Guidance Page) to create data management and sharing plans for their NSF proposals. Upon starting Biological Oceanography or Chemical Oceanography awards, PIs should immediately register their project by submitting project meta data to BCO-DMO. BCO-DMO accepts a wide array of data types from in situ sampling, laboratory experiments, satellite images, derived parameters and model output, and synthesis products from data integration efforts. BCO-DMO can advise on appropriate fit for the repository; if BCO-DMO and the PIs determine that some project outputs would be more appropriately served by other discipline-specific data repositories, meta data should still be deposited in BCO-DMO with links to the data residing in other repositories.

PIs who employ genomic techniques should articulate a strategy in the project Data Management and Sharing Plan for providing timely community access to the data collected and for establishing links between genomic and environmental data. Sequence data should be submitted to a publicly accessible data repository (e.g., National Center for Biotechnology Information). Persistent Identifiers corresponding to sequence data residing in an appropriate *omics repository (e.g., NCBI accession numbers) should be shared with relevant environmental repositories (e.g., BCO-DMO) to establish links between sequence and environmental observations.