

Data Sharing Requirements for Awards

The ocean science community has enjoyed an increase in funding for fundamental research that addresses responses to climate and environmental change. For many proposals the strongest broader impacts of the project may be the legacy of high quality observations that would be made available for use by others beyond the immediate investigator or project.

NSF has a general policy that data and research products from awards must be openly shared. The Division of Ocean Sciences also has a specific data sharing policy that is sent to all awardees when a project is recommended for funding. This document is provided as an attachment. The Biological Oceanography Program is required to provide oversight for data compliance for projects we support. The Program will not recommend funding for projects that will be collecting significant amounts of environmental or biological data and lack a robust plan to share and archive these data.

Compliance with data sharing policies needs to be a priority consideration from the beginning of the project because when compliance is deferred to the end of project, the results are often inadequate. As part of our Program review for a possible funding recommendation we are asking for a Data Management Plan that will become part of the award jacket should our recommendations result in an award.

The following statements reiterate important points covered in the Division of Ocean Science Data and Sample Policy (<http://www.nsf.gov/pubs/2004/nsf04004/nsf04004.pdf>) that we feel need emphasis and clarification:

- You have an obligation to provide both inventory metadata and primary data within specified time limits (section III, A). The data policy now specifies marine environmental data, but with increasing use of genomic based approaches (see section VI, D) there is a much wider range of data that should be reported and shared. Clearly, there are measurements and observations that are neither practical nor useful to regard as data to be openly shared. However, we do expect an adherence to the overall data sharing philosophy (section II) and a pragmatic approach to achieving these objectives.
- We expect you to address data sharing issues in annual and final reports (Section IV), and we will return reports not addressing progress with data compliance.
- The Biological and Chemical Oceanography Data Management Office (BCO-DMO; <http://bco-dmo.org/>) was created to serve PIs funded by the NSF Biological and Chemical Oceanography Sections as a location where marine biogeochemical, ecological and oceanographic data and information developed in the course of scientific research can easily be disseminated, protected, and stored on short and intermediate time-frames. We encourage use of this supported facility (this updates information in Appendix II, B and C in NSF 04-004).

The following are examples of the kinds of data you might be collecting as part of your project and some suggestions about how the data sharing requirements can be fulfilled.

1) Hydrographic and other oceanographic data from cruises and/or observatories. Metadata will be reported in UNOLS cruise report and posted on PI web site; final data will be supplied to NODC. Other standard oceanographic data such as phytoplankton primary productivity, HPLC pigments, etc can also be reported and submitted in a similar fashion. BCO-DMO should be contacted for these data.

BCO-DMO should be used for any ship-based projects and for any projects where an environmental data base will be created. This facility may also be a convenient repository for metadata for the project even if the primary data are more appropriately deposited elsewhere (see 2 and 3 below).

2) Genomic data should be deposited in NCBI data bases or other data repositories as appropriate. Program/project/cruise metadata tied to these databases should be provided to BCO-DMO so that there is a tie between the primary genomic data and the sampling program and/or any environmental data.

3) Some other primary data types may not be appropriate for BCO-DMO, but this facility may be appropriate for meta data about the project.

4) If there is no appropriate data management office or repository, you should describe plans for data sharing using an institutional web site or other appropriate venue.

5) Some experimental data and observations may not be appropriate for sharing and may be of no practical use for other beyond providing supporting data for information that will be part of publications. If you do not think your study will produce any data appropriate for sharing or archiving, please explain.

We know that the data support requirements within ocean sciences are changing rapidly and we expect to expand support for data base activities as needed. Asking for this formal data management plan is one way we have of being able to assess this changing need. We thank you for taking the time to provide this information.

Your Data Management Plan – At the end of this message, please describe the data to be collected as a part of your project and your plans for sharing and/or eventually archiving the data. If you have provided a data management plan in your proposal, please copy the existing text into this document and update or augment as needed.

Data Management Plan

Our proposed research outlines a series of quantitative statistical and modeling analyses that will be conducted on data from a range of fishery ecosystems along the northeast Atlantic coast of North America. Specifically, the proposal calls for a series of nested analyses to be conducted on the data partitioned at different spatial scales. We anticipate these analyses will rely exclusively on extant data and that no data will be “reconstructed” from archived records. **As such the project will not develop any new oceanographic data, or genomic data that can be archived in any of the national data centers or in the BCO-DMO.**

Here we summarize the overall data management plan for the project and for each of the core objectives.

A) Overall Data Management Responsibilities

All PIs are committed to an open data access policy. Each PI maintains an active web site that already provides colleagues and stakeholders access to information on former and current research projects. However, for this project we will develop a new portal that will provide a single point of access to results from all analyses conducted. PI Miller will have overall responsibility for development and support of open data access for the project. The results of individual analyses may be hosted on the web sites of the appropriate PI, but in all cases the main portal will provide hyperlinks to the hosting institution. However, we note that the nature of the products that will result from our research are not static as in the case of empirical data, we anticipate that the research products provided over the projects’ web portal will also be dynamic. We are committed to providing access to such information in a reasonable time frame.

B) Model development and results presentation

Development of the model code is a central objective of the work plan in the first year. Following this development period, we will make the finalized model code for both the statistical analyses and the population modeling available over the web to the general public as we have done in previous projects (e.g., geostatistical analysis of blue crab distributions - <http://hjort.cbl.umces.edu/crabs/LCPkrige.html>, and population dynamic of blue crab - <http://hjort.cbl.umces.edu/crabs/Assessment05.html>). The finalized versions of each analysis will be fully documented. Results of principal model runs that use the final model code will be documented and stored in a standard format. For the population models we will also provide sample input files that users can download and modify for their own applications.