

NSF-Ridge 2000 Data Policy

Introduction

The data management strategy for the NSF-Ridge 2000 Program is designed to address the needs of the program, individual Ridge 2000 (R2K) investigators, and the larger scientific community. Central to this strategy is timely submission and sharing of all metadata and data collected in both Integrated Study Site (ISS) field programs and Time Critical Study (TCS) rapid response cruises as well as sharing of all relevant historical data. Rapid dissemination of the metadata and data will maximize information transfer across the program, facilitate proposal preparation by investigators new to the program, and encourage integration of science, coordination of research, and the construction and testing of hypotheses. In keeping with this philosophy, all data used in R2K proposals should be in the public domain, or at least metadata identifying the location, data types, and contact person should be in the public domain at least 30 days before a grant proposal is submitted. R2K is a time limited program, thus all data collected should be rapidly released for maximum benefit to all. A strong commitment to data management is required of each participating PI. In requesting and accepting NSF support within the NSF-R2K program, each PI is obligated to meet the data management and disclosure requirements as an integral aspect of their participation in the program.

The Ridge 2000 Data Policy fulfills a community mandate from the R2K planning workshops and provides a mechanism for investigators to fulfill the NSF Division of Ocean Sciences (OCE) policy on the release of marine environmental data to the public domain. This policy is described in NSF circular 94-126 (<http://www.nsf.gov/search97cgi/vtopic>) and is a standard term and condition for all NSF OCE grants. Recipients of federal funding for collection of marine environmental data must release these data to an appropriate repository within two years of the date of collection. The OCE policy also requires that post cruise inventory information, currently in the form of a ROSCOP form, be provided within 60 days of the end of the cruise.

To facilitate data management, a data management system (DMS) will be implemented, maintained and operated by a data management office (DMO). The field data from R2K Time Critical Studies, which are currently limited to the Northeast Pacific, will be included in the Endeavour ISS database. The mission of the DMO will be to ensure that all R2K data sets are readily accessible by all R2K investigators on a common time base and within a common spatial framework. Ultimately, all metadata and data collected under R2K will be permanently archived as required by the OCE policy.

While recognizing the legitimate rights of data originators and collaborating PIs to the first use of the data they collect, the NSF-R2K Program encourages the oceanographic community to use data collected by the program, and in particular, believes that data availability should be restricted only in exceptional cases. As per NSF policies, data normally become publicly available for use without restriction two years after origination.

NSF-R2K Data Policy

The NSF-R2K Data Management Policy is predicated on guidelines that encourage openness and sharing of data for the mutual benefit of the scientific community. This policy sets

responsibilities for release of data with the understanding that some measurements will require long analytical or data reduction procedures that prevent early release after collection.

All data sets must contain a uniform suite of mandatory metadata that conforms with the policies to be developed for the R2K DMS. It is likely that the minimum requirements for each station or observation will include: cruise ID, time and date (UTC), position (lat/long and if available, xy coordinates with system origin), and event/operation number. For sub-samples from a bottle or other bulk sample, each data record must contain: cruise ID, event, dive or cast number, and sample or bottle number. For each data set, the metadata should include: descriptions of standards used for measurement of time and position, shipboard sampling procedures, sample treatment and preparation, analytical procedures, equipment calibrations, data reduction techniques, computation algorithms, analyses of standards or other data suitable for quality control and inter-laboratory comparison, citations, and any other useful information.

It is essential that PIs use standard digital forms to submit metadata to the DMO at the conclusion of each field program to facilitate effective and efficient use. Several levels of metadata exist, each defining a particular stage in the data acquisition to publication process. The levels are defined as follows:

(Level 1) Basic description of the field program including: cruise ID and dates, participating scientists, operation logs, navigation files and corrections, data types, and available underway data.

(Level 2) A final cruise report with complete data inventory in R2K standard format.

(Level 3) Data access information including: data formats, data quality assessments, details of processing procedures, and information on ongoing data processing and experimental studies.

(Level 4) Models and publications derived from the data.

A suite of basic environmental data is essential to enable interpretation of many data sets in the context of the ISS. Basic field data include tide data, pressure sensor data, current meter data, bathymetry, vent field maps, and CTD or comparable data on water column temperature and chemistry. All basic environmental data and metadata should be submitted to the DMO for inclusion in the DMS within 6 months of collection. PIs may place reasonable, time-limited restrictions on data use (less than two years). In some cases, it may be appropriate to provide metadata that describe derived data or analyses that are currently in progress. It is essential that all investigators using data from the DMS cite the originators of the data, even if no restrictions apply to its use.

All other data should be submitted to the DMO for inclusion in the DMS within 12 months of data acquisition. Data sets and collections that require lengthy analytical and/or processing procedures should be submitted as they are completed. In these cases metadata describing the work in progress are expected to be included in the DMS. For laboratory or theoretical studies, (meta)data to be submitted to the DMS include procedures, techniques, model parameters and computer codes. Historical data that would increase the value of the DMS should also be submitted promptly.

Responsibilities of Principal Investigators and Chief Scientists

The principles outlined above impose a series of responsibilities on principal investigators, chief scientists and the data management office. Chief scientists, in particular, have an ongoing responsibility to ensure that data are submitted and updated in timely and user-friendly fashion.

1. The Chief Scientist of each NSF-R2K cruise must submit Level 1 metadata as soon as the field program is complete. The Chief Scientist should ensure that a uniform, detailed operations log records at least the following information for every sampling operation: dive/operation number, station number, date, time, position, sampling device, and other comments. Standard digital forms will be provided by the DMO.

2. Digital cruise reports in DMO-standardized format, including the detailed operations log and cross-referenced detailed sample inventories, will be submitted to the DMO within 60 days of the end of the cruise.

3. Consistent with shipboard processing capabilities, basic data (*e.g.* bathymetric maps) should be available in preliminary form at the end of each cruise. The Chief Scientists should distribute these data, labeled as preliminary, to the DMO at the end of the cruise.

4. Final versions of basic environmental data should be submitted to the DMO for inclusion in the DMS as soon as possible and no later than 6 months after sample collection or instrument retrieval. Where this is not possible due to the nature of the analytical or data reduction process, Level 3 metadata indicating the existence and status of data-in-progress must be submitted and updated every 6 months. Data being used for Masters or PhD theses should be identified within the metadata and investigators wishing to use such data should first discuss their use with the PI.

5. Within one year of each cruise, PIs must submit all available data to the DMO, accompanied by Level 3 metadata. For data-in-progress, metadata indicating the existence and status of data-in-progress must be submitted and subsequently updated every 6 months. PIs making delayed measurements should strive to meet a timely release date. Unless authorized for early release by the responsible PI, all data will be on “restricted release” until 2 years post-cruise, after which time they will be freely available. Requests for data on restricted release will be referred by the DMO to the responsible PI.

6. Principal Investigators are responsible for the quality and correctness of data submitted to the DMS and should interact with the DMO to ensure that: (1) data comply with R2K DMS standards; (2) data subject to revision are updated promptly in the DMS; and (3) queries and criticisms from other users are promptly resolved.

Responsibilities of the Data Management Office

1. The DMO will provide a secure, web-based data retrieval system. The DMO will catalog submitted data and documentation such that they can be retrieved using criteria such as time, location, keyword, and/or sample identifier. Moreover, with input from the community, the DMO will define the data formats to be used for all types of data and provide PIs with digital forms on which to record their Level 1 and 2 metadata.

2. While PIs have primary responsibility for data quality, the DMO will provide basic assessment of all data for compliance with R2K DMS standards. The DMO will notify investigators of problems identified in their data sets by the DMO or by other users and work

with investigators to resolve such problems. The DMS will be a circular system that responds to feedback from users and providers of data and metadata.

3. The DMO will ensure that Level 1 and 2 data and metadata are compiled and submitted to appropriate national data repositories in a timely fashion following a cruise.

4. The DMO will release all data to the public domain two years after sample collection or instrument retrieval. Where appropriate, the DMO will ensure that R2K metadata and data sets are transferred to NODC and NGDC or other national databases. This release/submission will fulfill the obligation of the PIs as defined in the OCE data policy, but will not shift responsibility from the PI.

5. The DMO will liaise with PIs, the ISS coordinators, the R2K database Working Group and the R2K Office to encourage and evaluate community feedback, to ensure that community needs are being met and to ensure that all levels of metadata are available in the appropriate time frame.