

Appendix II: Program Specific Requirements

NOTE: The addresses provided (as of April 2003) may change. Please check with relevant Program Officers of the NSF Division of Ocean Sciences if necessary.

A. U.S. CLIVAR

All CLIVAR data shall be made available no later than *two (2) years* after collection, unless specifically waived by the international CLIVAR Scientific Steering Group (SSG). However, several CLIVAR activities, like the Global Hydrographic Survey, require Principal Investigators to submit data collected to a Data Assembly Center (DAC) for the purposes of quality control and data synthesis within shorter time periods. In general, the CLIVAR program requirements for data submission are similar to those found in WOCE Report No.104/93, WOCE Data Management. For more information contact:

Dr. David Legler
U.S. CLIVAR Office
1717 Pennsylvania Avenue, NW
Suite 250
Washington, DC 20006
Phone: (202) 223-6262
Fax: (202) 232-3065
<http://www.usclivar.org/>

B. U.S. GLOBEC

In addition to the data submission requirements mentioned in this document, the U.S. GLOBEC Scientific Steering Committee (SSC) requires all Principal Investigators to submit plans for the collection of data to the U.S. GLOBEC Data Management Office (DMO) at least *three (3) months* prior to execution of a sampling program. Specifics to be included in the data collection plan are detailed in U.S. GLOBEC Data Policy, Report Number 10, February 1994, available from:

U.S. GLOBEC National Coordinating Office
P.O. Box 1459
Leonardtown, MD 20650
Phone: (301) 997-0853
Fax: (301) 997-0854
<http://www.usglobec.org/>

Principal Investigators are responsible for documenting measurement and analysis techniques used to produce data sets and estimating accuracy and precision of these measurements. Specific physical measurements must be acquired along with all biological measurements and must meet pre-defined standards (see Report No. 10). In addition, the report specifies requirements for preservation of biological samples, including for the purpose of subsequent genetic analysis.

Data from measurements which do not involve manual analysis and which would be useful to the scientific community must be submitted by the principal investigator to the DMO within *six (6) months* after collection. All other measurements and any standard analyses of these measurements must be available to the community within *one (1) year* after collection. PIs will submit data either directly to the DMO or by placing it on-line as a U.S. GLOBEC distributed database. Format standards for submission of data and development of the database will be specified by the DMO. The DMO will serve as an intermediate archival location and data source and will transfer data to the NODC and prepare necessary documentation for data collected in foreign waters.

C. U.S. JGOFS

U.S. JGOFS chief scientists are required to submit all data to the Data Management Office (DMO) within *one (1) year* after the sampling date. However, data derived from long analytical procedures (e.g. ^{228}Ra) which prevent the researcher from being able to readily analyze/publish can be exempted from this *one (1) year* requirement. In addition, final versions of Basic Core Measurements (i.e. temperature, salinity, dissolved oxygen) must be received by the DMO within *six (6) months* after the sampling date. Again, some exceptions can be made for data requiring extensive analyses. However, all principal investigators making core measurements are urged to make their data available as quickly as possible. When submitting data to the JGOFS Distributed Data Management System (DDMS), Program principal investigators have two options: 1) to store data locally serving as a host node on the DDMS, or 2) to submit data to the U.S. JGOFS Data Management Office (DMO) and they will serve the data. CO_2 measurements should be submitted to the WOCE World Hydrographic Programme (WHP). More detailed information on the U.S. JGOFS requirements for data submission is available from:

U.S. JGOFS Planning and Data Management Office
GEOSECS Building
MS 43
Woods Hole Oceanographic Institution
Woods Hole, MA 02543-1535
Phone: (508) 289-2497
Fax: (508) 457-2161
<http://www1.whoi.edu/jgofs.html>

D. Ocean Drilling Program

The Ocean Drilling Program supports regional geological and geophysical field studies which can be used to develop mature drilling proposals in the Joint Oceanographic Institutions for Deep Earth Sampling (JOIDES) system. The geological and geophysical data from these projects are a primary source of information in planning drilling and should be available for review by the Site Survey and Pollution Prevention and Safety panels of JOIDES. Site survey data requirements for mature drilling proposals are identified in the JOIDES Journal issue titled, "Guide to the Ocean Drilling Program." Additionally, such data can be important in interpreting the results of a drilling leg and should be available to cruise participants.

Successful applicants are expected to deposit data from their cruises in the Ocean Drilling Program Site Survey Data Bank at Lamont-Doherty Earth Observatory, in addition to other data archiving requirements described in this document (see Appendix I.C.). The address is the following:

ODP Site Survey Data Bank
Lamont-Doherty Earth Observatory
Palisades, New York 10964
Phone: (845) 365-8542
Fax: (845) 365-8159
Email: odp@ldeo.columbia.edu
<http://www.ldeo.columbia.edu/databank/>

At the earliest possible date, the chairperson of the JOIDES Site Survey Panel, the manager of the Data Bank, and the representative of the appropriate national data center should be notified of the data types and schedule for submission.

The Ocean Drilling Program also supports more limited data collection activities through the U.S. Science Support Program administered by the Joint Oceanographic Institutions (JOI). Data reporting requirements under this program are the same as those identified above.

E. MARGINS

An important element of the MARGINS Program is that all data and results be rapidly shared in order to encourage integration of science, coordination of research, and the construction and testing of hypotheses. All data collected with MARGINS funding must be archived as soon as practically possible, along with all relevant metadata, in the institutional archives that are standard for a particular discipline (e.g., Incorporated Research Institutions for Seismology (IRIS) for seismological data, University NAVSTAR Consortium (UNAVCO) for Global Positioning System (GPS) data, Core repositories for marine geological samples, and NGDC for marine geological and geophysical data per Appendix I.C). Data for which no standard archive exists (e.g., Multi Channel Seismic (MCS) data, swath data and land geological samples) must be archived by the Principal Investigator and made available (with the cost of copying paid by the recipient) to researchers upon request.

Basic metadata (e.g., data types, sample locations, cruise tracklines, etc.) must be provided to the MARGINS Office within *sixty (60) days* of ending a field program. In due course and in collaboration with ongoing efforts in the Marine Geology and Geophysics Program, the MARGINS Office is currently developing tools for preparing and formatting these metadata files. It is the responsibility of the Principal Investigator to provide to the MARGINS Office, for publication on the MARGINS Office web site, details of and links to all datasets acquired or generated with MARGINS funding. Contact information for the MARGINS office is provided below.

MARGINS Office
Lamont-Doherty Earth Observatory
P.O. Box 1000, 61 Route 9W
Palisades, New York, 10964 USA

Phone: (845) 365-8665

Fax: (845) 365-8156

<http://www.ldeo.columbia.edu/margins/Home.html>

All raw data must be made freely available *two (2) years* after ending a field program, consistent with the data release policies of the NSF Division of Ocean Sciences and other national and program centers (IRIS, Program for Array Seismic Studies of Continental Lithosphere (PASSCAL), and US National Ocean Bottom Seismography Instrument Pool (OBSIP)). In the case of datasets that are not available to the investigators at completion of the field-season/cruise, for example, because they are assembled by the relevant data-center before distribution, the two year moratorium period begins on the date that the complete dataset is made available to investigators. However, Principal Investigators are encouraged to release data to other Focus Site investigators as soon as possible following the end of a field season or completion of dataset processing.

Processed, derived and interpreted datasets must be made publicly available as soon as possible, certainly within the lifetime of the grant. This policy applies even to those data and results that Principal Investigators have traditionally not been required to make publicly available (e.g. stacked and migrated seismic sections, geochemical analyses, Digital Elevation Models (DEMs) and other rasters, geological samples and geochemical analyses).

F. Ridge 2000

1. Introduction

The data management strategy for the Ridge 2000 (R2K) Program is designed to address the needs of the program, individual 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.

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.

While recognizing the legitimate rights of data originators and collaborating PIs to the first use of the data they collect, the 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.

2. Data Policy

The 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 (conductivity, temperature, depth) 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.

3. 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 (DMO). Chief Scientists, in particular, have an ongoing responsibility to ensure that data are submitted and updated in timely and user-friendly fashion.

- a. The Chief Scientist of each 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.
- b. 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.
- c. 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.
- d. 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.
- e. 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.
- f. 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.

4. Responsibilities of the Data Management Office

- a. 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.

- b. 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.
- c. 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.
- d. 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.
- e. 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.