

Ridge 2000 (R2K) Data Compliance Plan  
March 2, 2009

D. Fornari, B. Govenar, and Ridge 2000 Steering Committee

## 1. Statement of Purpose

The primary goal of the Ridge 2000 Program (R2K) is to achieve an integrated, holistic understanding of global mid-ocean ridge processes. Now in its final phase, the program is focused on the compilation, integration, and synthesis of datasets collected through R2K projects, future R2K-funded projects, and legacy data (R2K-relevant, but not funded using R2K funding) from each of the three Integrated Study Sites (ISS) and Time Critical Studies (TCS). This integration will not be fully achievable until there is complete and seamless sharing of data across disciplinary boundaries with participation from all investigators. These data need to be properly archived and accessible through the Ridge 2000 Data Management Office (DMO), part of the Marine Geophysical Data System (MGDS), at Lamont-Doherty Earth Observatory (LDEO). To date, much of the Level 1 and Level 2 metadata (e.g., sample ID and position, see footnotes) and environmental field data (e.g., raw sensor data) have been submitted to the DMO, but only a small fraction of the derived data from R2K field and laboratory programs have been submitted. This is despite the need for access to these data to facilitate integration and synthesis and conform to the stated NSF grant requirements pertaining to data submission, (the R2K and NSF data policies are available at: <http://www.ridge2000.org/science/info/data.php#policy>) and the NSF web link: [http://www.nsf.gov/publications/pub\\_summ.jsp?ods\\_key=nsf04004](http://www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf04004)).

To stimulate integration and synthesis of multidisciplinary research both within and between the three ISSs and to make comparisons to TCS and analog sites outside the ISSs, the R2K Steering Committee and Program Office have been charged by NSF to develop an action plan for compliance with existing NSF and R2K Data Policies, as outlined in the most recent R2K Program Announcement ([http://nsf.gov/funding/pgm\\_summ.jsp?pims\\_id=5513](http://nsf.gov/funding/pgm_summ.jsp?pims_id=5513)). This Data Compliance Plan encourages investigators to submit available data to the DMO immediately, to make provisions for timely submission of recently collected data, and to ensure that the spirit and terms of each research grant are consistent with NSF policy and the R2K Program goals. ***The objective of this plan is to bring the R2K Program into data compliance by summer 2009 and to ensure data compliance in the future.***

## 2. Responsibilities of Ridge 2000 Principal Investigators

The NSF Data Policies outline responsibilities for Chief Scientists of field programs (i.e., cruises), Principal Investigators, and the DMO. The Chief Scientists are responsible for submitting Level 1 and Level 2 metadata, while the PIs are responsible for submitting Level 3 and Level 4 metadata and all environmental and “other” derived data<sup>1</sup>. Environmental or basic field data, including raw and processed shipboard, vehicle and sensor data should be submitted within 6 months of collection; and derived data, including grids, mosaics, laboratory analyses, modeling results, interpretations (e.g., shape files), and tabular data should be submitted within 1 year of acquisition<sup>2</sup>. This includes, but is not limited to, data that have been vetted by peer-review and are included in journal publications. Specialized data that have been submitted to other repositories (e.g., GenBank, PetDB) must also be discoverable in the R2K Data Portal, which has the capability to provide direct linkages to those data repositories to further facilitate data integration and synthesis.

In some cases, for multi-year, multi-investigator projects, some of the analyses may not be complete in this time frame. In these cases, it will be important to submit additional information about planned and ongoing analyses to the DMO so anticipated data sets can be identified until final results are available (see below: ii. Shore-based Data Analysis Plan). ***The submission of metadata does not exempt investigators from making the actual raw and/or derived data available in a timely manner.*** PIs should be aware that starting with the 2009 R2K competition, NSF will use data compliance as a strong consideration in making award decisions and that every proposal to R2K will need to specify the status of the PIs compliance with the R2K Data Policies.

Submissions to the Data Portal by investigators can be reasonably divided into the following three categories:

#### **i. Field Data (raw and processed) and Level 1 & 2 Metadata**

This includes environmental data such as bathymetric data, temperature observations, vent field maps, data from standard equipment on NDSF vehicles or UNOLS ships, as well as all data from science party instrumentation.

#### **ii. Shore-based Data Analysis Plan**

This describes the plans for producing data (the anticipated products), and is vital for other researchers to develop plans for using the data. This metadata may include:

- Specific types of analyses expected to be done on each sample or with each data set.
- Storage facility and preservations status (e.g., dried, frozen, chemically treated) for each sample, with name and contact information of the person to whom requests for samples can be sent.
- Information on whether students are working on samples or data, and the expected dates of degree completion indicated.
- Estimated delivery date for outstanding derived data.

#### **iii. Final Data**

These are data products used as basis for publication and relate to Level 3 & 4 Metadata.

- Data products include bathymetric and sidescan sonar grids, seismic sections, velocity models, chemical and temperature probe time-series records, results of laboratory analyses, interpretations, etc., **but are not limited to data presented in the publication.**
- Tables of data, analyses completed and verified, to be included in, or already published in journal articles.
- Model metadata and data: should include model parameters or input metadata, computer code when non-proprietary, or software version metadata when proprietary software is used, and summary of model results (e.g., schematics, summary figures etc) including, but not limited to, those published in journal articles.
- Final data submission should include references to publications and electronic copies of all data tables that are either in the papers or resulting from the research. If possible, relevant cruise IDs and data sets should be noted.
- Information about data held in external databases (e.g., GenBank accession number) should also be provided to enable linkages between the R2K Data Portal and other databases.

### **3. Responsibilities of the Ridge 2000 Program Office**

The R2K Program Office will facilitate the implementation of the Data Compliance Plan in several ways. It will directly contact PIs and PI groups who have had field or laboratory based programs at each ISS to clarify the status of various datasets, schedule for submission of data to the DMO, provision of secondary metadata if needed, and publication schedules for research papers. This effort will commence in March 2009. As part of this effort, it will also contact investigators with “available” datasets, identified in the reports of the three ISS Integration Synthesis Workshops, held in September 2008. The R2K Office has also begun to collect bibliographic information and electronic copies of papers on mid-ocean ridge research funded by NSF, particularly by R2K. It is hoped that this collection of papers will also be useful for identifying missing datasets in the Data Portal that could be easily submitted by the investigators and ingested by the DMO. If necessary, the R2K Office may also provide opportunities for data compilation and submission workshops, where PIs working at each ISS or in thematic areas can meet and work together to compile and verify data, and submit data on-site to the DMO to achieve compliance. The timing of these data ingestion workshops would be determined based on discussions with R2K PIs working at the various ISSs and subject to budgetary constraints.

### **4. Responsibilities of the Ridge 2000 Data Management Office**

The DMO at LDEO has the responsibility to provide data archiving, storage and access capabilities for the Ridge 2000 Program, as well as other national research programs (e.g., MARGINS and Antarctic and Southern Oceans programs, see: <http://www.marine-geo.org/>). The URL for the R2K DMO is: <http://www.marine-geo.org/portals/ridge2000/>.

DMO staff is available to help PIs with questions regarding data submissions and will, in general, facilitate data ingestion into the database. They have developed an extensive and innovative array of software and web-based tools to both assist in searching for and utilizing data in the database. The primary contact for R2K at the DMO is Dr. Vicki Ferrini ([ferrini@ldeo.columbia.edu](mailto:ferrini@ldeo.columbia.edu)); PIs should not hesitate to contact her regarding questions about submitting data to the R2K Data Portal or problems with existing data listings or metadata. DMO staff will be available to attend the data integration workshops as needed.

### **5. Restatement of NSF Ridge 2000-Specific Reporting Requirements**

Open access to Ridge 2000 data is crucial to the success of the integration and synthesis phase of the program. Therefore, all annual and final reports and requests for no-cost extensions, must discuss progress in fulfilling the Ridge 2000 Data Policy (<http://www.ridge2000.org/science/downloads/Ridge2000DataPolicy.pdf>). In this regard, NSF expects that all requirements stated in the Data Policy will be met according to its specified timetable. All proposals submitted to R2K competitions from 2009 onward will need to specify the status of the PI(s) data compliance with the R2K data policy on previously funded projects. Because the availability of data is viewed as a critical requirement to enable integration and synthesis activities, NSF will use data compliance as a strong consideration in making award decisions from 2009 onward. To prevent delays in fulfilling this requirement, PIs should examine their metadata spreadsheets and reports in the R2K DMO Portal in advance of proposal deadlines to assure timely entry and submission of all required meta- and derived datasets. Any questions concerning this policy should be directed to the cognizant program officers<sup>3</sup>.

**Footnotes:**

<sup>1</sup> Four levels of metadata: “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 set 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.” From the “Division of Ocean Sciences Data and Sample Policy”, National Science Foundation, NSF 04-004, p. 11

<sup>2</sup> “All basic environmental data and metadata should be submitted to the DMO for inclusion in the DMS [*sic* Data Portal] *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 [*sic* Data Portal] include procedures, techniques, model parameters and computer codes. Historical data that would increase the value of the DMS should also be submitted promptly.” From the “Division of Ocean Sciences Data and Sample Policy”, National Science Foundation, NSF 04-004, p. 11

<sup>3</sup> NSF Ridge 2000 Program Officers are:

Richard Carlson, telephone: (703) 292-7978, email: [rcarlson@nsf.gov](mailto:rcarlson@nsf.gov)

David Garrison, telephone: (703) 292-7588, email: [dgarrison@nsf.gov](mailto:dgarrison@nsf.gov)

Barbara Ransom, telephone: (703) 292-7792, email: [bransom@nsf.gov](mailto:bransom@nsf.gov)