

## Data Management

**Data types.** During this project we will obtain data on a variety of geochemical parameters, process rates, as well as gene expression profiles. We will create an environmental database that will contain all the obtained geochemical and physicochemical parameters, as well as the measured process rates.

**Research Expedition.** During the research expedition, the actual sampling events will be recorded on paper logs (scanned into PDF documents) and in a digital event log. The event log and the analytical data collected during the expeditions will be stored on a dedicated directory on a local computer and shared (password-protected) through a local network for easy data access and entries. The directory will be backed up on a daily basis on an external hard drive located in a separate room. An expedition report will be prepared at the end of the cruise, and two copies of the data sets will be transported separately back to WHOI by two different cruise participants. Soon after the completion of the expedition, the original underway data will be contributed by the vessel operator to the UNOLS central data repository at <http://www.rvdata.us/catalog/> managed by the Rolling Deck to Repository (R2R) project. Also, R2R will ensure that the original underway measurements will be archived permanently at NODC and/or NGDC as appropriate for the data type. The measurements made by the science parties will be managed by the Biological and Chemical Oceanography Data Management Office (BCO-DMO) and the data sets will be available online from the BCO-DMO data system (<http://bco-dmo.org/data/>). BCO-DMO will also archive the data they manage at the appropriate national archive facility, such as NODC and NGDC, following NSF guidelines. Cruise data and data collected as part of this project will also be deposited at the Marine Geoscience Data System Ridge 2000 Data Portal ([www.marine-geo.org/portals/ridge2000/](http://www.marine-geo.org/portals/ridge2000/)).

**Data availability and archives.** During this project we will obtain a large amount of sequence data as well as data on a variety of geochemical parameters, and process rates. E-mail and teleconferencing will be used to maintain vigorous communications among the parties involved in this project. Upon publication, the generated data will be submitted to publically accessible databases, such as NCBI ([ncbi.nlm.nih.gov/](http://ncbi.nlm.nih.gov/)) for the sequence data, the DSMZ ([dsmz.de/](http://dsmz.de/)) for obtained cultures, and the Marine Geosciences Data System ([marine-geo.org/](http://marine-geo.org/)) for the geochemical data. The metagenomic and –transcriptomic data will also be made publically available at sites such as MG-RAST and IMG/M. Samples and derived products obtained during the proposed studies will be appropriately stored and made available to other researchers upon request, the latest after publication of data derived from these samples. All notebooks will be scanned and stored electronically, as well as other information relevant to the collection, processing, and analyses of the samples. Inventories and data information will be kept on the computers of the respective PIs, who all have established robust institutional data management policies, which define general protocols for data use, access, share, formats, security, backups and long-term archival, taking into consideration research needs, legal obligations and ethical responsibilities.