

Data Management Plan

Our data management plan is based on guidelines established by the National Science Board and the National Science Foundation and covers dissemination and sharing of materials and data that are expected to be collected as part of the research described in the project proposal. This data management plan has been agreed to at the institutional level and represents our overarching philosophy of open access to all information and data gathered during any and all publically funded scientific research. We intend to make our data as open access as possible in the shortest amount of time that is needed for securing publications.

I. Types of Samples and Data, and their Distribution.

This project entails analyzing samples that have already been collected as part of IODP Expedition 331 and includes discrete microbial sediment samples, mineral assemblages, rock specimens and fluids. All of these sample types have been previously subsampled by the different user groups for specific types of analyses as described in the expedition report (Takai *et al.*, 2010). Remaining archived core samples are presently available (post-moratorium) through JAMSTEC and the Kochi Core Depository.

The primary types of samples and data we anticipate utilizing as part of this project are:

- 1) Microbial sediment samples, mineral assemblages and rock specimens collected in small discrete quantities from whole-round core sample collected during IODP expedition 331.
- 2) Nucleic acid extracts derived from these samples.
- 3) Isolated strains of FeOB will be deposited in Bigelow Labs' Center for the Culture of Marine Phytoplankton (CCMP) as frozen stocks, along with appropriate metadata.
- 4) The greatest amount of data that will be acquired will be genomic. Genomic data generated using nextgen sequencing technologies for SCG and metagenomic analysis will be placed in the NCBI's short read archive (<http://www.ncbi.nlm.nih.gov/sra>). Curated metagenomic data may also be uploaded to the Joint Genome Institute's IMG/M system that hosts metagenome data, as well as submitted to GenBank. All SSU sequences or sequences of other functional genes of interest confirmed by conventional DNA sequencing will be released to GenBank prior to publication as is standard practice.
- 5) Metadata will be archived by PI directly to hard disk and by using a cloud backup service, e.g., Dropbox. Logbooks from Exp 331 have been scanned and digitally archived. A selection of personal photographs taken during field operations showing different aspects of our sampling activities have been also been archived.

II. Individual Scientist Responsibilities.

PI is responsible for maintaining all data associated with this project, in accordance with any and all institutional determined requirements. Basic practices for key areas are outlined below.

Lab notebooks. All information connected with initial experimentation, procedures, analysis, and results will be kept in a lab notebook. When studying genomes and metagenomes, for example, data are too plentiful to record by hand in a paper notebook. In such cases, digital notebooks may be preferable. Regardless of media, these notebooks will be stored as well. Hard-copy notebooks will be stored and archived to enhance institutional retrieval. All individual lab notebooks remain the property of the host laboratory and institution (WWU).

Field notebooks. As outlined above, all logbooks from this expedition were scanned and archived. Precise sample locations were also recorded and are made available in our expedition report (Takai *et al.*, 2010).

Data security. In addition to data hosted on local computers, it is advised that all important data be stored on off-network mobile devices (e.g., hard disks) or offsite cloud resources. Password protection is strongly encouraged. These data (and associated passwords) must be made available to senior institution officials in the case that any institutional liability issues should arise.

Data backup. Stored data shall be regularly backed up, preferably weekly. The frequency with which back up shall occur will be monitored by the PI.