

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. We intend to make our data as open access as possible in the shortest amount of time that is needed for securing publications. For this collaborative project, the data management plan is agreed upon by all institutions, and may be considered common to all.

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

This project primarily entails sampling of discrete microbial mat and fluid samples using custom-built samplers (e.g., BMS sampler). In addition, hydrothermal fluids will be accessed using standard gas tight fluid samplers, major fluid samplers, semi-automated fluid and filtration systems as well as surrounding waters will be captured using standard CTD rosette Niskin bottles. All of these samples will be subsampled by the different PI lab groups for specific types of analyses. We do not anticipate collecting core, sediment, or dredge samples that need to be archived at NSF-approved repositories.

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

- 1) Seawater and hydrothermal vent fluid samples for chemical analysis; these will be distributed among the participating scientists as determined by PIs.
- 2) Microbial samples collected in small volumes and/or as discrete quantities on filters; these will be distributed among the participating scientists as determined by PIs.
- 3) Nucleic acid extracts; these will be generated and shared as determined by PIs from samples distributed in 1 & 2, above.
- 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. The geochemical and associated metadata will be logged in a way that is consistent with data acquisition protocols for the biological and chemical oceanography data management office (BCO-DMO) and deposited with BCO-DMO.
- 5) Metadata in the form of video and datalogging as part of the routine ROV and AUV operations is maintained in a digital archive and as part of the 'Virtual Van' for Jason at WHOI. In addition, the chief scientist will keep and maintain hard drives containing all raw video footage and still camera images. A master copy of these data will be made fully available to each PI. Select video highlights and stills will be placed on a central server for the project also accessible by each PI. Naming standards for dive, location, and specific sample type will be agreed upon prior to cruise operations, and maintained throughout. Sample logs for each dive series will be scanned following the dive and distributed to all cruise participants. A selection of personal photographs taken during cruise operations showing different aspects of cruise activities will be collected, and disseminated publically via educational outreach webpages and through social networking sites.

II. Individual Scientists Responsibilities.

Individual basic practices for key areas are outlined below:

Lab notebooks. All information connected with initial data collection, 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 individuals' lab notebooks remain the property of the host institution.

Field notebooks. As outlined above, all logbooks from any filed expeditions will be scanned and archived. Precise sample locations and any maps produced will be recorded and will be made available in our cruise report, also to be made publically available.

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. This 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 determined by the overseeing PI. Metagenomic data will be stored both locally and remotely (offsite and/or in the cloud).