

## DATA MANAGEMENT PLAN

**Project:** From hot to cold in the dark - shifts in seafloor massive sulfide microbial communities as physical and geochemical conditions change after venting ceases

**NSF Division:** OCE

**Solicitation Info:** Biological Oceanography

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. Description of data types

The goal of this proposal is to collect active and inactive hydrothermal sulfide chimneys, sample some immediately and incubate the remainder on the seafloor for a few weeks and a year, after which they will be collected. The proposed project includes recovery of vent deposit samples on *HOV Alvin* or *ROV Jason* during dives to two vent fields along the East Pacific Rise near 9N. The focus is on identifying shifts in microbial communities as physical and chemical conditions change after hydrothermal venting ceases. All samples to be collected will be subsampled by 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 designations.

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

Sample metadata by dive, which will be compiled and submitted shipboard to the Marine Geosciences Data System (MGDS) for immediate release; all recovered samples will be assigned International Geo Sample Numbers (IGSN) through SESAR (System for Earth Sample Registration). The PIs will work with the MGDS to archive and make all data sources publically available.

Shipboard/submersible data including still images, video, and navigation data will be submitted to the MGDS for immediate release.

Recovered vent samples: Any remaining vent deposit sample will be archived in the WHOI SeaFloor Sample Lab (SFSL) dry storage room with location/sub-sampling metadata in SEDCORE2000 linked to the MGDS.

X-ray diffraction (XRD) and X-ray adsorption spectra (XAS) data will be deposited in BCO-DMO. Exoenzyme (alkaline phosphatase, leucine aminopeptidase) activity rates will be collected using a Tecan Spark 10M multimode microplate reader. Data will be deposited to BCO-DMO and all data will be included in table format (tab delimited text files) as supplements in published manuscripts.

Sequence data: MiSeq PE 300bp iTag data will be generated for bacteria and archaea from all samples in our collection. These data will be deposited to GenBank and made publicly available by the end of the project period or upon publication, whichever occurs first.

Metagenome data generated with Illumina NextSeq for selected samples will be deposited to IMG Metagenome Expert Review (IMG/MER) and Metagenomics-RAST (MG-RAST), and metatranscriptome data (also Illumina NextSeq) will be archived on the NCBI –Gene Expression Omnibus (GEO) database and made publicly available by the end of the project period or upon publication, whichever occurs first. Links to all data will be provided through BCO-DMO.

Educational outreach materials: Blogs and videos generated during the cruises will be maintained on Sylvan's Wordpress blog website and Google Drive website. Videos will be

additionally be uploaded to Youtube and/or Vimeo.

## **II. Policies for access, sharing and archiving**

Data collection and analysis processes as well as contextual details (sampling site location, time, etc.) will be documented in individuals' laboratory notebooks and in their publications.

The PIs will work with the MGDS and BCO-DMO to archive and integrate data and metadata from this project. Data archived with all three of these repositories are publically available and easily searchable. Data will be submitted to the MGDS, PetDB, and BCO-DMO by the completion of project Year 3. Links and references to the data sources will be available in associated publications and presentations.

All data will be shared through peer reviewed publications, including mineral identifications and porosity determined by reflected and transmitted light petrography and SEM)

## **III. Policies and provisions for re-use, re-distribution**

There will be no embargo periods for political/commercial/patent reasons. Further, there will be no permission restrictions placed on the data. Data dissemination will be noted in the publications within the Materials and Methods section to inform the scientific community of the data availability and accessibility. All data and metadata will be made available through the MGDS and BCO-DMO. We will retain the right to hold data prior to publication only if a conflict of interest seems warranted.

The dissemination of data to be collected for this proposed research will not be restricted by any ethical or privacy issues, copyright concerns or restrictive licenses. As discussed above, all the data collected will be made readily available to the scientific community through various datacenters, published manuscripts in peer-reviewed journals, and by request to the affiliated researchers.

## **IV. Data and metadata formats and standards**

1. Observational data will be submitted as spreadsheet files (.xlsx).
2. Genetic data formatting will follow the standards of NCBI, IMG, and MG-RAST and links to these data will be provided through BCO-DMO.
3. Enzyme activity data will be submitted to BCO-DMO as spreadsheet files (.xlsx).

## **V. Data storage and access during the project**

All experimental data generated at WHOI will be stored on our laboratory computers and associated Network Accessed Storage devices that are backed up daily by WHOI's automated backup service. All experimental data generated at TAMU will be stored on laboratory computers and in Google Drive, for which TAMU faculty have unlimited file storage; Sylvan maintains data storage currently for all files/data generated in his lab on Google Drive.

## **VI. Roles and responsibilities**

Lead PI Sylvan will ensure compliance with this Data Management Plan. He will be responsible for deposition of all genetic data generated at TAMU. WHOI PI Tivey and UMN PI Toner will be responsible for deposition of geochemical and mineralogical data to MGDS.