Data Management Plan

Data Policy Compliance: The project investigators will comply with the data management and dissemination policies described in the *NSF Award and Administration Guide* (AAG, Chapter VI.D.4) and the *NSF Division of Ocean Sciences Sample and Data Policy*.

Pre-Cruise Planning: Pre-cruise planning will occur through email and teleconferences. The cruises will utilize a deep submergence vehicle (DSV Alvin, Woods Hole Oceanographic Institution). Preliminary dive plans will be written prior to the cruise, and updated and amended prior to each dive based on cruise events.

Cruise event logging: Detailed dive reports for each Alvin dive will be digitally compiled as Microsoft Word documents during the cruise. Each dive report will include site information, times of launch and recovery, sampling events, and watch-stander logs. These, along with each dive plan, participant information, and samples logs will be compiled into a cruise report at the conclusion of the cruise and shared among the cruise party.

Description of Data Types

This project will produce observational datasets through the ship's underway sensors, DSV Alvin sensors and cameras, deployed autonomous chemistry sensors, and derived from analysis of collected biological specimens. Observational data will be collected on research cruise at the East Pacific Rise planned to take place during Jan-April of two different years.

Cruise underway data: Standard underway data collected along the ship's track (e.g., sea surface temperature, salinity, etc.). File types: .csv ASCII files; Repository: BCO-DMO and the Rolling Deck to Repository (R2R).

Alvin data: Routine sensor data (e.g., temperature), video, and images collected by the DSV Alvin, as well as dive event logs, are made available through WHOI dives FrameGrabber system at http://4dgeo.whoi.edu/alvin and archived at the National Deep Submergence Facility (NDSF) at Woods Hole Oceanographic Institution. File types: .csv ASCII, .jpg, .mpeg; Repository: NDSF archive

Field colonization: Data on the abundance, weight, size, and fate of collected species will be recorded in Excel spreadsheets. After the cruise, genetic species ID data will be added to these datasheets after sequencing and analysis. File types: Excel file(s); Repository: BCO-DMO.

Autonomous sensor data: We will carry out *in situ* measurements of sulfide (total and free sulfide), oxygen, and pH at each sampling site using cyclic voltammetry with Au/Hg microelectrodes and potentiometric miniaturized glass electrodes. File types: .csv ASCII; Repository: Marine Geosciences Data System.

Genomic data and gene sequences: Microbial genomic, bacterial 16S rRNA gene sequences, and metazoan mitochondrial CO1 gene sequences will be derived from experimental and field-collected biofilm samples, metazoan larvae and settlers, and metazoan adults. The resulting assembled and raw data will be made publicly available. File types: Short-read archive .sra, .fastq read files, .fasta files; Repository: NCBI, MG-RAST; accession numbers and associated sample metadata will be submitted to BCO-DMO.

Proteomic data: Microbial metaproteomic data will be derived from experimental and field-collected biofilm samples. The resulting raw and assembled data will be made publicly available on the ProteomeXchange database (http://www.proteomexchange.org/). File types: .mgf files for MS spectral data, .fasta files for translated sequence data. Excel spreadsheet containing spectral data I.D. and corresponding metagenomic target (translated sequences) will be deposited to BCO-DMO repository.

Animal collections: Data on the abundance, weight, size, and fate of collected species will be recorded in Excel spreadsheets. File types: Excel file(s); Repository: BCO-DMO.

Larval samples: Larvae will be collected via high volume pumps and benthic tube traps. Taxonomic identification data and photographs derived from fresh and preserved samples will be recorded in Excel spreadsheets. After the cruise, genetic species ID data will be added to these datasheets after sequencing and analysis, when necessary. File types: Excel file(s); Repository: BCO-DMO.

Data and Metadata Formats and Standards: Genetic sequence data will be prepared in accordance with the minimum information about a metagenome sequence (MIMS) specifications and minimum information about a marker gene sequence (MIMARKS) developed by the Genomic Standards Consortium, and the Earth Microbiome Project standards and protocols for 16S rRNA gene amplicon sequencing. All other data will be stored in Excel spreadsheets, which can be submitted to the BCO-DMO. In addition, metadata will be prepared using BCO-DMO metadata forms and will include detailed descriptions of sampling and analysis protocols.

Data Storage and Access During the Project:
We will create a relational database (based on postgreSQL) containing all the obtained microbiological, metazoan and geochemical data. A searchable web interface hosted on the Department of Marine and Coastal Sciences' server (Vetriani lab) will be used by the PIs to access data and metadata associated with the project. This will facilitate meta-analysis of data and integration among tasks assigned to different PIs. Raw data will be backed up on PIs' institutional servers. Additionally, the investigators will store raw project data (including Excel spreadsheets, PDFs of dive logs, etc.) on a shared Microsoft Sharepoint directory that will be hosted by PI Arellano.

Mechanisms and Policies for Access, Sharing, Re-Use, and Re-Distribution: Immediately after completion of the research cruise, underway data and metadata will be submitted to the Rolling Deck to Repository. Within four months of the conclusion of the cruise, the final cruise report will be submitted to the BCO-DMO. All other data will be submitted to the appropriate repository (detailed previously) upon submission of manuscripts, or 2 years post-cruise, whichever comes first. Genbank accession numbers will be submitted to the BCO-DMO after sequences are submitted to Genbank.

We will adhere to the standards, policies, and provisions for data and metadata submission, access, re-use, distribution, and ownership as prescribed by the BCO-DMO Terms of Use (http://www.bco-dmo.org/terms-use). Resultant manuscripts will be published in open access journals where appropriate. All datasets, database and derived products obtained during the proposed studies will be appropriately stored and made available to other researchers upon request after publication of data derived from these samples. All custom scripts (R environment) for multivariate analyses will be made available through a repository in GitHub (https://github.com/) and assigned a permanent DOI using Zenodo (https://zenodo.org/).

Plans for Archiving: Rolling Deck to Repository will ensure that the original underway measurements are archived permanently at NODC and/or NGDC as appropriate. NDSF maintains permanent archive of ROV and AUV sensor data, images, and video. The NCBI Genbank maintains a permanent archive of sequence data. The lead PI will work with R2R, NDSF, BCO-DMO to ensure data are archived appropriately and that proper and complete documentation are archived along with the data.

Roles and Responsibilities: Each PI will be responsible for sharing his/her subset of data among the project participants in a timely fashion. S. Arellano and L. Mullineaux will be responsible for collecting and analyzing the field settlement and larval. C. Vetriani will oversee 16S rRNA gene and genomic sequencing, along with proteomic work, and will submit the resulting sequences to the National Center for Biotechnology Information's (NCBI) GenBank database. The Lead PI, S. Arellano, will coordinate the overall data management and sharing process and will submit the project data, including GenBank accession numbers, and metadata to the Biological and Chemical Oceanography Data Management Office (BCO-DMO) who will be responsible for forwarding these data and metadata to the appropriate national archive.