DATA MANAGEMENT PLAN

LIFE AFTER DEATH: DO INACTIVE SULFIDES FUEL A UNIQUE ECOSYSTEM AT THE DEEP SEAFLOOR?

DATA POLICY COMPLIANCE

The project investigators will comply with the data management and dissemination policies described in the NSF Award and Administration Guide and the NSF Division of Ocean Sciences Sample and Data Policy, including adhering to and promoting the standards, policies, and provisions for data and metadata submission, access, re-use, distribution, and ownership as prescribed by the Biological and Chemical Oceanography Data Management Office (BCO-DMO).

PRE-CRUISE PLANNING

This project plans for two cruises using assets of the National Deep Submergence Facility (NDSF, which at the present time includes HOV Alvin, ROV Jason and AUV Sentry). Planning will be done via teleconferencing and email communication. Detailed plans for station locations, vehicle deployment, habitat characterization, sampling protocols, and other science activities will be drafted with input from the other PI(s) involved in the collaborative cruises. Sampling events will be recorded on digital event logs (or paper logs scanned into PDF).

DESCRIPTION OF DATA TYPES

On the cruises the project will produce observational datasets, collect imagery and physical samples, and produce experimental datasets. In the laboratory on shore, additional observational datasets will be produced from processed samples and imagery, and derived datasets will result from statistical and bioinformatic analyses.

1 Observations at the seafloor will include NDSF vehicle data (e.g., navigation, bathymetry, temperature probe) and PI-provided data (e.g., sampling event logs).

2 Imagery at the seafloor will include NDSF vehicle imagery (e.g., 4K video) and PI-provided imagery (e.g., time-lapse still).

3 Physical samples will include animals, rocks, and water column samples. Rocks and animal specimens for which we extract gut contents will be assigned unique sample identifiers, important for provenance for subsamples to be analyzed for microbes.

4 Animal composition and function: Individuals will be sorted from collections or annotated from imagery, identified to lowest taxonomic level using morphological characters (and/or genetic sequence barcode for collections), and recorded into tabular datasets to also include assignment of functional trait modalities.

5 Animal movement ecology: Time-lapse imagery will be analyzed for movement of grazers with the context of substratum type.

6. Microbial cell density and primary productivity: Cell counts via standard epifluorescence microscopy will be recorded into tabular datasets. Primary productivity from multi-day shipboard incubations with 14C-bicarbonate measured via liquid scintillation counter will be recorded into tabular datasets.

7. Microbial community composition, including microbes in animal gut contents: Extracted DNA will be used for analysis of amplicon-based microbial community composition (16S for Bacteria and Archaea, 18S for single-celled Eukaryota). After quality control screening of raw reads (sequence datasets), amplicon sequence variants (ASVs) will be generated for taxonomic assignment (tabular datasets).

8. Microbial community function: Metagenome sequencing for a subset of sulfide and animal gut samples will be performed on the Illumina NovaSeq 6000 platform. After quality control screening of raw metagenomic sequences, we will generate contiguous sequences (contigs); assembled contigs will be binned into metagenome-assembled genomes (MAGs).

9. Minerals: Rock samples will be photographed and described. Subsamples will be prepared for petrographic thin sections for light microscopy to record minerals, grain size, textures, porosity and paragenetic relations (tabular datasets). Images taken through the microscope will be analyzed to quantify porosity and pore distribution.

10. Derived data: We will integrate the above data types into statistical and bioinformatic analyses that will result in tabular datasets for publishing in peer-reviewed journals.

DATA AND METADATA FORMATS AND STANDARDS

Field observation data will be stored in flat ASCII files, which can be read easily by different software packages. Observational and experimental data from processed samples will be stored mainly as

spreadsheets. Imagery from the seafloor will be stored in original resolution (this will be the biggest data volume in this project). Metadata will be prepared in accordance with BCO-DMO conventions (i.e. using the BCO-DMO metadata forms) and will include detailed descriptions of collection and analysis procedures.
Specifically for animals: We will utilize the World Register of Marine Species (WoRMS) taxonomic classification. We will standardize to Darwin Core format to provide occurrence data to OBIS and GBIF.
Specifically for microbes: Genetic sequence data will be prepared in accordance with the minimum information about a marker gene sequence (MIMARKS) and about a metagenome sequence (MIMS) developed by the Genomic Standards Consortium, with quality control screening of raw reads from 16S/18S rRNA sequencing and raw metagenomic sequences.
Specifically for minerals: Rock samples will be assigned International Geo Sample Numbers (IGSN) through SESAR (System for Earth Sample Registration).

DATA STORAGE AND ACCESS DURING THE PROJECT

Each PI will have a copy of their respective files on individual laboratory computers with backup for redundancy. The investigators will share project data (including ASCII files, still images, and spreadsheets) in shared cloud storage (Microsoft OneDrive) provided by WHOI's Information Services (IS) group. See Facilities statement for additional information about each institution's on-premises and/or cloud backup services and WHOI's cloud storage. Excluding the high-resolution video, we do not expect to exceed the 25 TB per WHOI OneDrive user capacity in this project. For the high-resolution video provided by the NDSF on external drives, we will make additional copies to external drives if needed.

MECHANISMS AND POLICIES FOR ACCESS, SHARING, RE-USE, AND RE-DISTRIBUTION

The repository for all PI-provided data is BCO-DMO, unless otherwise noted. Data sets will be made available through the BCO-DMO data system within two years from the date of collection. DNA sequences will be deposited in the National Center for Biotechnology Information (NCBI). NCBI identifiers and metadata will be provided through BCO-DMO. For biological samples we will archive voucher specimens if we discover new species, otherwise preserved specimens will be maintained in the PI's labs during analysis, and then transferred to BioSpecs facility at WHOI for long-term storage. Rock samples will be archived at WHOI SeaFloor Sample Lab (SFSL) dry storage room.
Data, samples, and other information collected under this project can be made publicly available without restriction once submitted to the public repositories. 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. Data analyses and resultant manuscripts will be published in open access journals where appropriate. Some scripts for data analyses will be made publicly available through GitHub repositories.

PLANS FOR ARCHIVING

The investigators will work with BCO-DMO to ensure that PI-provided data are submitted to the appropriate national data archive and with necessary documentation. The NCBI maintains a permanent archive of nucleic acid sequence data. Data collected by NDSF will be archived at the Lamont Doherty Earth Observatory-hosted Marine Geosciences Data System managed by IEDA. An event log may be used for vehicle deployments, which would be provided by research vessel operators along with underway data to the Rolling Deck to Repository (R2R). NDSF imagery data is archived at the Woods Hole Oceanographic Institution data repositories.

ROLES AND RESPONSIBILITIES

Each PI will be responsible for sharing among the project participants and submitting to BCO-DMO his/her subset of data, matched to Data Types section above:
1 Observations at seafloor: All PIs
2 Imagery at seafloor - Mullineaux and Beaulieu
3 Physical sample cataloging - Jones will obtain IGSNs for rock samples; Beaulieu will provide unique animal specimen identifiers. Each PI will catalog his/her respective subsamples associated to identifier when applicable.
4, 5: Data for animals - Mullineaux and Beaulieu
6, 7, 8: Data for microbes - Sylvan and Achberger
9: Data for minerals - Jones
10 Derived data from integrating 1-9: All PIs
Lead PI Mullineaux will ensure compliance to this data management plan.