Data Management Plan

Deep-sea research has grown in importance and become increasingly data-intensive, with uses ranging from conservation to resource management. We adhere to open-access data principles and will seek to maximize the findability, accessibility, interoperability and reusability of our data. In particular we will seek to share information with Alaska native tribal and government officials responsible for environmental management. We will adhere to the Division of Ocean Sciences Data and Sample Policy and will archive data and/or links to data repositories within the 2-year timeframe as stipulated. All data types will either be submitted directly through the Rolling Deck to Repository program, directly to BCO-DMO, via UNOLS, or made available by PIs via meta-data linked to BCO-DMO within two years after generation. We will establish a collective Data Dryad repository for our project data (http://datadryad.org/).

A. Data Generation Activities: This proposal will include the following types of data:

(1) Video and still photographic data from dives will be generated and stored in raw and processed formats according to the specifications of the submergence assets.
(2) Standard oceanographic (temperature, salinity, depth) data plus dissolved oxygen, and pH, alkalinity, and calculations of aragonite saturation state. These will all be georeferenced typically in .csv or .txt formats.
(3) Porewater geochemical, isotopic, metabolic turnover rate, and cell count data from sediment cores and carbonates will be organized in .txt and .csv formats.
(4) Species count data with lowest possible taxonomic ID in .csv and .txt formats.
(5) Sequence data (16S rRNA amplicon) and image datasets consisting of nanoSIMS (.im files) and fluorescence images (e.g. HCR-FISH) as .tif.

B. Research Cruise Data: Levin as lead PI will organize and distribute meta data as well as cruise videos. Pre- and post-cruise data meetings will be held with all cruise participants to coordinate data responsibilities, generation and format of metadata, and data holding locations. During the cruise one member from each of the 4 labs involved will be responsible for updating a community metadata file after each dive or event (e.g., CTD) indicating key descriptors such as dive or cast number, site, time of sampling, latitude, longitude, depth, sample type and ID numbers, and distribution. These metadata will be held in a project drive (Google) and will be submitted to BCO-DMO for each cruise. CTD data and bathymetric (multibeam sonar) data will be deposited via the Rolling Deck to Repository program.

C. Faunal Specimen management: Samples will be sorted into morphospecies and photographed, with vouchers kept for long-term vouchering and subsamples taken for stable isotope and the molecular work (95% ethanol and/or freezing). Unique numbers are assigned to each specimen or specimen lot and this is tracked though all subsequent processing, including DNA sequencing. For material studied in detail, voucher specimens will be submitted to the Benthic Invertebrate Collection (BIC) at Scripps Institution of Oceanography (SIO).

F. Disseminating data to the Ocean Biogeographic Information System (OBIS). To reach a wider audience, all specimen data generated from this project will be disseminated to the deep-sea node of OBIS, based at Senckenberg, Germany.

G. Molecular sequence data: All DNA data generated by direct (Sanger) sequencing will be deposited on GenBank with appropriate referencing to the voucher specimens held at SIO-BIC. Illumina-generated nucleotide sequence data will be deposited into the Sequence Read Archive at NCBI (http://www.ncbi.nlm.nih.gov/Traces/sra/). Metadata associated with nucleotide sequence information will conform to MIxS (Minimum Information about Any Sequence) standard, which encompasses the MIGS, MIMS, and MIMARKS standards for genomes, metagenomes, and marker genes, respectively.

H. **Geochemical and Metabolic Turnover Rate Data** - Geochemical and metabolic turnover rate data will (in addition to the submission to BCO-DMO) be uploaded to the World Data Center PANGAEA, who is a member of the ICSU World Data System.

I. **Image datasets**: Image data from nanoSIMS experiments (.im files) for individual ions, processed ratio images, and segmented images (processed in Matlab or python) relevant to publications will be deposited and made available with a unique and permanent DOI via the CaltechDATA repository (https://data.caltech.edu/). Caltech has committed to maintaining this archive indefinitely (https://data.caltech.edu/terms). Publications will clearly state where data has been deposited, to enable validation of research findings.

J. **Protocols:** Any new protocols developed during this program will be deposited with the open access Protocols.io database (https://www.protocols.io/)

K. **Data Quality:** Data QA/QC will follow appropriate plans for each type of data generated, with attention to metadata. Careful attention will be paid to sample custody and metadata during the cruise. Because this is such a large, complex, inter-disciplinary project, it is simply impossible within the space provided to detail all of the QA/QC procedures for genetic, oceanographic, chemical, ecological, and other data sets and analyses.