Plan for Data Management, Archiving, and Community Access

This plan adheres to the open access data policies of the NSF – Division of Ocean Sciences (17-037) and the Bigelow Laboratory for Ocean Sciences, including consideration of copyright, intellectual property, and dissemination of data and other research products. Orcutt will be responsible for ensuring that all data are archived appropriately. In addition to local storage on Bigelow Laboratory's computer server system (described in Facilities documentation), all data will be made available on the BCO-DMO data portal (Orcutt project ID 565799, which is part of the C-DEBI BCO-DMO Program).

1. Data Categories

The primary data categories include (1) technical documentation of *in situ* experimental system components, (2) flow cytometry data, (2) nanoSIMS cellular elemental data, (3) radioisotope incubation data, (4) potentiostat time-series current records, (5) *in situ* dissolved gas concentration time-series measurements, (6) fluid chemistry data; (7) DNA sequence data, and finally, (8) metadata on sample collection, incubation conditions, preservation, and calibrations. Secondary support datasets include (1) pressure time-series data from CORK observatory network, (2) standard shipboard underway data, and (3) standard ROV underway data. There will also be temperature-sensitive biological fluid samples collected and archived.

2. Data Descriptions

Technical documentation: The incubation system (i.e. modified flow through chimney and syringe samplers) developed as part of this project will be documented with a technical report, including drawings, photographs, and data presented in plots.

Flow cytometry data: BONCAT experimental samples sorted by flow cytometry at the Bigelow Laboratory Facility for Aquatic Cytometry will generate hundreds of ~16 MB flow gram .fcs files containing light scatter and fluorescence property characterization of sorted particles. Raw data, as well as .png or .jpg image files of sorted particle regions will be archived.

nanoSIMS cellular elemental data: Ion microprobe analysis by the Caltech CAMECA NanoSIMS 50L with generate hundreds of high resolution maps of elemental abundance at the single cell level of cells mounted on gold-coated filters. Image maps of cellular element abundance with corresponding mass information will be archived.

Radioisotope incubation data: Incubations of fluids with radiotracers will generate scintillation counts per sample measured on a scintillation counter, which will be converted to radioactivity based on comparison of counts to known standards and converted to ASCII and/or spreadsheet files using standard formats. Metabolic rates will be calculated based on activity of injected substrate compared to activity in product pool, the substrate pool concentration, and time of the incubation.

Potentiostate time-series current data: Oxidation-reduction potential electrode data will potentially generate thousands of individual voltage potentials that will be analyzed shipboard and in the laboratory, which will be auto-analyzed in near-real time using computer algorithms that track changes in current measurement, and confirmed with manual review of data upon recovery.

Dissolved gas sensors: The dissolved gas sensors used in this project will generate time-series records of voltage corresponding to dissolved gas concentrations as a function of time, which will be stored on a laptop connected to the sensor interface in the ROV Jason van and

analyzed remotely. These data sets will be formatted as ASCII files and/or spreadsheet files using standard formats and archived along with laboratory documentation (calibration tests).

Fluid chemistry: Major, minor and trace element concentrations will be measured on select discrete syringe samples of CORK fluids using ICP-MS. Metadata will include sample location, sample time, dilution volume, and analysis method.

DNA sequence data: Environmental and sample DNA extraction and sequencing will generate DNA concentrations per volume material, and Illumina Miseq or NexSeq 16S rRNA gene V4 hypervariable region amplicon sequence data files and shotgun metagenome sequence data. Sequence data will be provided as forward and reverse read .fastq files from the commercial sequence facility; raw .fastq data will be made publicly available through the NCBI Sequence Read Archive, with relevant link information provided to BCO-DMO. Amplicon data processing using the *mothur* program will generate ASCII and/or spreadsheet formatted files of sequence abundance, operational taxonomic unit abundance and phylogeny, various image files; this will be accompanied by an HTML-formatted markdown file documenting software script entries and data files generated. Shotgun metagenome data will be processed and assembled into contigs and/or genome bins in .fasta format.

Metadata and underway data: Primary metadata will include information about the cruise and sample collection. This information will be summarized in a PDF cruise report and associated sample spreadsheets. Laboratory notebooks and log spreadsheets will serve as hard copy backup, which will all be digitized for archival at the primary institution. Orcutt will work with ship and ROV technicians to ensure underway data submission to the Marine Geoscience Data System, the Rolling Deck to Repository (R2R) program, and the video archive for ROV Jason at the Woods Hole Oceanographic Institute.

Biological samples: When possible, samples of fluids will be preserved as archive material for sharing with external users and made available upon request.

3. Policies for Access and Sharing

Data will be made available to the general public as described above within two years of data collection, Data archiving and long-term storage will be accomplished in collaboration with the groups and database systems listed above. Papers resulting from proposed work will be linked to dataset through these archives, as well as through the C-DEBI website, and (where and when permitted) posted in Open Access venues. Progress of data archiving will be reported in each annual report of the project, with appropriate links provided.