

DATA MANAGEMENT

We will comply with the NSF Division of Ocean Sciences' open access data policy (NSF 04-004). The data that we generate will be published and made available publicly through a number of data repositories. Shipboard hydrographic data will be deposited to R2R (<https://www.rvdata.us>) and to the Biological and Chemical Oceanography Data Management Office (BCO-DMO; <http://bco-dmo.org/>) within six months of the cruise.

This project will generate water column and sediment geochemical, physical, and microbial and viral data from sites in the Gulf of California. Our cruise track will depend on hydrographic, weather, and other conditions, but we plan to sample 6 different areas. At each site we will carry out ALVIN dives, SENTRY mapping surveys, and multiple CTD casts. In all, we expect to generate ~400 water column samples (CTD and gas tights), and 400 pore fluid samples (we will profile ~8 cores/area \times 6 areas \times 8 depths ~ 384) and solid phase sediment samples during the cruise. We will conduct AOM and SR rate assays in all profile cores and examine alternate modes of AOM in a subset of these cores with the highest methane concentrations. For the water column, we will assess methane oxidation, dark CO₂ fixation and bacterial production at all depths; other rate assays will be conducted at 3 depths with the greatest plume impact plus one depth well above that. We will generate environmental biogeochemical data, activity data and genomics data for sediment and water samples.

The specific types of data we will generate include:

- A. *Hydrographic Data*: We will obtain standard hydrographic, chemical, and optical data from the CTD (routine C, T, D, O₂, chlorophyll fluorescence, and beam attenuation). Additional hydrographic data will be collected on ALVIN (via a hull mounted CTD) during research dives and via AUV Sentry during mapping dives.
- B. *Biogeochemical Data*: We will measure nutrient concentrations, major ions, dissolved organic matter, volatile low molecular weight alcohols and acids, dissolved gases (alkanes, H₂, CO), and redox metabolites in sediment samples and water column plume samples. Sampling details and methods are given in the Project Description.
- C. *Activity and Biological Data*: Achieving our objectives requires that we constrain patterns of microbial activity and determine microbial and viral community composition. We will quantify microbial biomass and document the abundance and diversity microorganisms and viruses using genomics. We will carry out rate assays with radioisotopes and stable isotopes, including stable isotope probing experiments to give independent measure of active microorganisms and assess patterns of activity of particular organisms across sites and over depth. Rates of methane oxidation, dark CO₂-fixation, and bacterial production will be quantified routinely as will ammonium, sulfide, and CO oxidation at select depths. Metagenomics, metatranscriptomics and BONCAT-FACS/BONCAT-FISH data sets will be generated for a select subset of samples. Sampling details and methods are given in the Project Description.

I. Data and metadata standards

- A. *Hydrographic Data*. Data and metadata for these analyses will be submitted to BCO-DMO. Both raw and processed CTD data will be uploaded with relevant metadata components and all analyzed data fields. Metadata will include detailed information on data collection (method of collection, position, platform, site characteristics, weather conditions) and analytical techniques. (instrumentation used, calibration data and date, standards and blanks used, QA/QC protocols, etc).

DATA MANAGEMENT

- B. *Biogeochemical Data*. Data and metadata for these analyses will be submitted to BCO-DMO. Metadata will include detailed information on data collection (method of collection, position, platform, site characteristics, weather conditions) and analytical techniques (instrumentation used, calibration data and date, standards and blanks used, QA/QC protocols, etc).
- C. *Activity and Biological Data*: Data and metadata for plankton abundance and activity measurements will be submitted to BCO-DMO along with the biogeochemical data as described above. Metadata will include detailed information on data collection (method of collection, position, platform, site characteristics, weather conditions) and analytical techniques (instrumentation used, calibration data and date, standards and blanks used, QA/QC protocols, etc). All molecular (16S rRNA gene, metagenomic, and metatranscriptomic) data will be curated according to the Genomic Standards Consortium. Raw reads and processed genome data will be submitted to the Sequence Read Archive and WGS databases at NCBI Genbank, respectively. Meta-omics data will also be made available publicly through DOE's Argonne National Laboratory MG-RAST. All protocols and scripts for data processing will be made publicly available through protocols.io and the PI's lab github pages. NCBI accession numbers for all molecular data will be provided in publications and in spreadsheet form in association with the paired rate measurement data.

II. Policies for accessing/sharing and provisions for data protection

We anticipate that our work will attract interest from a broad range of scientists and we will make our data accessible to others as soon as possible. The main avenue for sharing our data and results will be through publication in scientific journals and presentations at conferences, but we will make our data widely available to other researchers to minimize duplication of effort. To accomplish this, we will upload these data to BCO-DMO as well as our own data server to assure optimal data availability. Data will be freely downloadable from these sources within two years of collection. This initial embargo period will allow PIs and graduate students to explore the data set for original publication before opening it up for wider use. We will work closely with BCM-DMO to maximize the legacy of our observations.

III. Policies and provisions for Re-Use and Re-distribution.

Once our final, quality-checked data are uploaded to BCO-DMO and our own websites, no permission restrictions will exist. We anticipate that the primary users of these data will be researchers working on deep sea microbiology and biogeochemistry but the results will be of interest to any person curious to learn more about microbial oceanography. Downloadable data and metadata will be freely accessible to anyone who is interested.

IV. Archiving plans and preservation of access

Many of our sampling and analytical activities are destructive in nature, preventing reasonable archiving of many of our field samples. As possible, samples will be archived and made available to other PIs, upon request. We plan to keep data for 10 years past the lifetime of the project. Since this portion of the data occupy a relatively small amount of space relative to our computing capabilities, the long-term preservation will be easily accomplished by keeping several copies on local computers at UGA, Montana and Wisconsin.