

Data Management Plan

In Year 1, Senior Personnel (CVD, FG, JMB, and SW) will establish a sample naming convention that will encompass all forms of material and data collected.

DATA POLICY COMPLIANCE

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

DESCRIPTION OF DATA TYPES

The project will produce several observational and experimental datasets which are listed below. Samples will be collected from 4 stations along the Central California margin. Two cruises will be completed, one during the first year of the study and one during the second year. Each cruise will last 5 days and include sampling at 4 sites.

Observational datasets:

1. **CTDO₂/Hydrographic Data:** Hydrographic data (depth, temperature, salinity, & dissolved oxygen) collected on Cruises 1 & 2 will be collected by CTDO₂ and associated with tows and physical collections where appropriate and archived in the BCO-DMO database upon the completion of the project. File types: spreadsheets (.xls, .csv). Repository: BCO-DMO.
2. **Assemblage Data:** Foraminiferal assemblages will be quantified from each MOCNESS tow. Abundance data will be archived on the BCO-DMO database upon the completion of the project. File types: spreadsheets (.xls, .csv). Repository: BCO-DMO.
3. **MOCNESS Samples:** Each net from MOCNESS tows collected on Cruises 1 and 2, will be assigned a unique identifier and stored in borate buffered formalin. Splits used for foraminifera assemblages as well as any residual material will be archived at NCSU. The preservation approach has been selected in keeping with standard practice in biological oceanography and to ensure archived material can be maintained for a substantial period of time. Archived material will be made available upon request. Repository: NCSU.
4. **Cellular Ultrastructure:** Individual foraminifera, from OMZ-affiliated species, that appear to be laden with cytoplasm will be isolated from the MOCNESS tows and analyzed with TEM. Transmission Electron Microscopy (TEM) will be completed on properly preserved samples. At the most, we plan to analyze 36 specimens. Several images will be used in publications; unpublished TEM micrographs will be hosted on BCO-DMO, upon completion of the project. File types: images (.tif, .jpg files). Repository: BCO-DMO.
5. **Underway Data:** Data collected while the research vessel is underway and on site (e.g., weather data, ADCP data, navigation data) is automatically uploaded for each cruise to the R2R repository (Rolling Deck to Repository). This information is automatically done by the ship's operator, and is fully accessible. File types: Varied. Repository: R2R.

Experimental datasets:

1. **Discrete water sampling:** Water samples will be collected at each station with a rosette with 10-L Niskin bottles. From selected Niskins, water samples (~20 ml) will be filtered (0.2 μ m) and stored frozen until lab analysis for nitrate concentration, and $\delta^{15}\text{N}_{\text{NO}_3^-}$ and $\delta^{18}\text{O}_{\text{NO}_3}$. Resulting nitrate concentration, and the $\delta^{15}\text{N}_{\text{NO}_3^-}$ and $\delta^{18}\text{O}_{\text{NO}_3}$ will be published, and made available on BCO-DMO upon completion of the project. File types: spreadsheets (.xls, .csv). Repository: BCO-DMO.
2. **Foraminiferal Denitrification Rate and Intracellular Nitrate:** The denitrification rate of foraminifera will be measured shipboard, in addition to which the nitrate concentration as well as the $\delta^{15}\text{N}$ and $\delta^{18}\text{O}$ of nitrate will be measured in ~120 specimens. A synopsis of resulting data,

and the $\delta^{15}\text{N}_{\text{NO}_3^-}$ and $\delta^{18}\text{O}_{\text{NO}_3}$ will be published with data made available on BCO-DMO. File types: spreadsheets (.xls, .csv). Repository: BCO-DMO.

3. **Metagenomics:** Data will be generated by Illumina NextSeq, and analyzed by the middle of year 2 from specimens preserved shipboard. Metagenome sequences will be deposited to GenBank and MG-RAST, upon completion of the project. All codes and scripts that will be used to analyze the data will be deposited to GitHub. File types: Fasta files, Microsoft Word, Notepad. Repository: GenBank and MG-RAST, GitHub.

DATA AND METADATA FORMATS AND STANDARDS

Geochemical data will be maintained in XLSX Excel files, stored in plain CSV text files and the associated metadata in read.me files. Analytical procedures will follow standard, published protocols for QA/QC and be compared to certified standards. All sensors will be calibrated according to manufacturer specifications, but will be the responsibility of the ship operator. All appropriate data will be geo-referenced.

DATA STORAGE AND ACCESS DURING THE PROJECT

The investigators will store all project data (e.g., spreadsheets, images, text files, scanned field/lab logs) on computers which are backed up daily. The Principal Investigator (PIs) will also establish a shared account (Cloud-based storage) for data storage and sharing among project members. Personal computers will be backed up daily to external hard drives.

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

The investigators of the project will hold the intellectual property rights to the data. However, we will not place any restrictions on data access, use and re-use, except that we be attributed when data are used for future studies. To ensure public access, we will be depositing data in publicly accessible online databases (e.g., BCO-DMO). The majority of the data we will collect do not present any privacy issues and can be made publicly available without restrictions, within 2 years of analysis. We plan on publishing results in peer-reviewed open-access journals to ensure that there are no restrictions on access to either data or published results.

PLANS FOR ARCHIVING

In addition, the PIs will work with BCO-DMO to ensure that data is archived appropriately, including any associated documentation. Genetic sequencing data will be stored in a publicly accessible, permanent database (e.g., NCBI).

ROLES AND RESPONSIBILITIES

Each PI will be responsible for sharing his/her subset of data to all research members in a timely fashion. The lead PI, C.V. Davis, will coordinate data management and sharing among investigators as well as submission of project data to BCO-DMO (metadata) or another qualified archive.