

DATA MANAGEMENT PLAN

1. Types of Data and Samples

This project will generate several types of data from *in situ* ship-based measurements and laboratory experiments; we anticipate the following general types: (1) water column biogeochemical and hydrographic parameters; (2) rate measurements and PCR-dependent estimates of microorganism abundances and transcriptional activities; and (3) DNA sequences. University of California guidelines for data management (<http://www.cdlib.org/services/uc3/datamanagement/>) will be followed for all data types.

Discrete sample data will be generated according to established sampling and analytical protocols, with methodologies for each measurement included with the accompanying metadata. Water column data will include CTD depth-profiles of temperature, salinity, pressure, oxygen concentrations, photosynthetic active radiation, light transmission and fluorescence collected as part of the cruises proposed as part of the project. In addition, nutrient concentrations (NH_4^+ , NO_3^- , NO_2^- , PO_4^-) and dissolved oxygen concentrations will be measured from discrete samples collected as part of these cruises. Rate measurements will include ^{15}N -based determination of ammonia oxidation rates, and oxygen-based measurements of production and respiration rates, from both *in situ* samples and perturbation experiments. QPCR and RT-QPCR gene and transcript abundances will be generated from discrete water column samples and experiments.

Samples and subsamples will be physically archived in appropriate locations (e.g., freezers, climate-controlled facilities) for subsequent (re)analysis. Types of samples will include: frozen seawater (125-500 ml), plankton concentrates (on filters) for subsequent rate determinations and nucleic acid extractions, and nucleic acid extracts.

2. Data and Metadata Standards

Data quality will be assured through proper analysis of replicate samples, standards, and blanks/controls. Our quality control protocols are stringent and we will discard data that do not meet all requirements. Data will also be archived in multiple locations, including hard copies, laboratory computers, and cloud-based servers. All data will be compiled in commonly-used database management programs, and metadata and primary data will be submitted to archiving data repositories as appropriate. Hydrographic and other oceanographic metadata will be supplied to the National Oceanographic Data Center and the Biological and Chemical Oceanographic Data Management Office (BCO-DMO) established through the Ocean Carbon and Biogeochemistry program at the Woods Hole Oceanographic Institution. All DNA sequence data will be deposited into the GenBank public archive. Data and metadata will be submitted to publicly available databases following the standards required by the specific repository (e.g., BCO-DMO, GenBank, NODC).

Experimental data and observations not appropriate for archiving in national data repositories will be reported in peer-reviewed publications, either as tabulated data in the publication or in supplementary data tables. Where appropriate, data will also be made available on laboratory websites for download. To increase accessibility to project data and the dissemination of our research findings—particularly among scientists from developing countries—we will make every effort to publish our results as open-access articles, or within open-access journals, using reduced rates negotiated between the University of California and select journals (see http://osc.universityofcalifornia.edu/alternatives/submit_work.html including PLoS [20% discount] and PNAS [25% discount]). Ph.D. or M.S. theses associated with the project will also be made available electronically.

Metadata needed to support our observations include time and day of year, latitude, and longitude, units of measure, accuracy and precision of measurements, methods of measurement and sampling, investigator, and data processing protocols. Cruise-level metadata will include name of vessel, start and end dates for the cruises, PI names and contact information, cruise plans, cruise reports, data inventory

lists, and event logs. These metadata will be entered into a computer spreadsheet program and stored as ASCII text. We will use the metadata standards outlined in the document titled “Data Management and Guidelines Manual” produced by the NSF-funded Biological and Chemical Oceanography Data Management Office (BCO-DMO).

3. Data Sharing, Reuse and Redistribution Policies

We will share and archive data collected as part of this research project in compliance with the Division of Ocean Science Data and Sample Policy. We will submit quality-controlled data to suitable databases (BCO-DMO, CAMERA, GenBank) that make these data and associated metadata publicly available. Data will be made available as soon as possible after collection; but within 1 year of collection. All data will be publicly available within six months of the project end date. The original data collector/creator/principal investigator does not retain the right to use the data before opening it up to wider use. There are no ethical and privacy issues with the proposed data and no human research subjects are included as part of this study (negating the need for an IRB protocol). The datasets from this project will not be copyrighted.

4. Policies and provisions for re-use, re-distribution

All data from this project are considered within the public domain for all not-for-profit uses and there will be no permission restrictions placed on use of the data.

5. Plans for archiving and preservation of access

Hydrographic and biogeochemical data submitted to BCO-DMO will be maintained in perpetuity through archival at the National Oceanographic Data Center (NODC). Microorganism abundances and transcriptional data will be made available through a central project website. Nucleic acid sample aliquots will stored in ultracold freezers in perpetuity.