

Data Management Plan

I. Title and contact information

1. *Project title*: “Collaborative Research: Regional variation of phytoplankton diversity and biogeochemical functioning in the subtropical Indian Ocean”

2. *Points of contact*:

Overall data products and molecular data – Adam Martiny, amartiny@uci.edu, (949) 824 9713.

Biological and Biogeochemical datasets - Michael Lomas, mlomas@bigelow.org and Ben Twining, (btwining@bigelow.org), (207) 315-2567.

II. Types of data

A. Data description

Processed carbon, nutrient, iron, pigment, and cellular stoichiometry in Excel files, CSV files and raw text files containing chemical analysis, output files from instruments and output files from model runs. The expected size of collection is less than 100GB.

B. Existing and ancillary data

Synoptic data (T, S, nutrients, oxygen) used in this study will be derived from data collection efforts underway from the ship. No special tools or software are needed to use these data.

III. Data and metadata formats, standards, and organization

A. Formats

ASCII formats, with comma or tab-separated values and/or in spreadsheet format, will be generated, maintained and submitted to Biological Chemical Oceanographic-Data Management Office (BCO-DMO). Data generation will not be large so there is no need to compress files using binary format. ASCII data is universally readable by most if not all analysis software. The complete data product from the proposed work will be made available through the BCO-DMO.

All genetic data will be submitted to GenBank. The metagenomic data will also be posted on MG-RAST where it is possible to include extensive metadata. We will also adhere to recommendation from Genomic Standards Consortium and include the maximum number of contextual INSDC fields for metadata. We will link the data by a unique identifier to the BCO-DMO data to ensure easy integration

B. Metadata

We will use the BCO-DMO metadata authoring tool to prepare our information for submission to the archive. We will work with BCO-DMO to include the biochemical parameters that we have used in this study are properly described in their system.

C. Data organization

Preplanning will be coordinated between the PIs and a sample plan for the cruise will be drafted to finalize the following questions: a) All necessary types of data and metadata to be collected to achieve the project goals; b) Sampling strategy to obtain the best possible sample coverage. c) Data management during the cruises of opportunity. During the cruises, this project will be directly responsible for collecting and storing raw output from measurement of biogeochemical properties. A sample log will be kept in paper copy and transcribed to a spreadsheet recording available meta information, including the time/location/physical conditions during water sampling to build the first stage of a data archive. The spreadsheet and other at sea cruise files

will be backed up daily on the ship's data storage repository. At the end of each cruise, a cruise report will be prepared and all digital data will be backed up to CD prior to disembarking from the ship. Data will be stored on a desktop computer and backed up to a remote storage system at UCI and Bigelow. Post-cruise analysis will generate data stored in ASCII format and backed up on the UCI and Bigelow data storage systems. Laboratory notebooks and sample logs will be scanned and digital files stored in the same way as described above. All file names include DDMMYYYY and are renamed as they are re-saved.

D. Data quality

Routine instrument calibration for measurement of all biogeochemical parameters is conducted using appropriate calibration standards reference materials. Equipment will be regularly calibrated. All notebooks and written sample logs will be scanned and stored electronically, as well as other information relevant to the collection, processing, and analyses of the samples. Data files and information will be kept on the computers of the respective PIs, who all have established robust institutional data management policies. The designated person of responsibility is Adam Martiny, amartiny@uci.edu, (949) 824 9713.

III. Data access and sharing

The data will be of exclusive use by the PIs until publication or one year after the end of the grant. The project will upload data files and collection-level metadata to BCO-DMO once at the end of the project, once submitted for publication. These are experimental data. BCO-DMO will make the data publicly available through their system immediately upon receipt and subject to any approved embargo period. Data may be parsed into subsets that work with the format of the data. For example, the molecular data will be a separate dataset from the chemical data. All datasets for the project will be connected such that anyone looking at the project will see all the data as we've done with prior research projects.

IV. Data Reuse

Scientists interested in the linkages between phytoplankton physiology, ecological function and biogeochemistry in the Indian Ocean should find these data useful. The data will be of interest to modelers attempting to model the carbon cycling in the Indian Ocean and other ocean basins limited by iron and nitrogen. We will maintain a link on our institutional website to the location of the data on BCO-DMO as well as indicate this in our publications. Data will be described in accordance with developing BCO-DMO standards. The investigators will work closely with BCO-DMO curators to ensure accurate and complete documentation in accordance with the BCO-DMO designated level of service, if appropriate. To facilitate tracking of reuse and fair credit to data providers, BCO-DMO will provide a recommended formal citation for the data set, including a persistent identifier and the contact person's last name.

V. Data Preservation

During the field and laboratory work, a collective sample log of samples collected and allocated for each measurement type will be kept in spreadsheet form to build the first stage of a data archive. The event log and the analytical data collected during the lab work will be stored on multiple physical disks locally. This plan for backups will permit restoration in the event of a hard disk failure, fire or other incident, which might affect multiple computers in one physical location. After the completion of the lab work, a metadata file and discrete data files will be submitted to BCO-DMO.