

## DATA MANAGEMENT PLAN

Our goal is to have all raw project data in the public domain a maximum of 6 months past the end date of the project. Reasonable requests from interested parties for early data distribution will be discussed by all PIs at the biweekly meetings outlined in the Overall Management Plan (see Project Description), and, if deemed appropriate, honored.

### **Chemical and biological oceanographic data**

Oceanographic data will be collected in conjunction with the SPOT program. Some data will be collected during sampling; additional chemical assays will be performed on collected water samples in the laboratory. This aspect of the project is a continuation of the existing Microbial Observatory work, and protocols for data handling are well developed. The data is provided to cruise participants when it is ready, and currently is incorporated into the USC Microbial Observatory website (<http://www.usc.edu/microbialobservatory>). Biological assays on the samples will be performed in the Caron and Fuhrman labs. We will work with the SPOT program to coordinate this data and submit it to the Biological and Chemical Oceanography Data Management Office's (BCO-DMO) Biological and Chemical Oceanography Database. While this database was created to support projects in other NSF directorates, we feel the information is congruent with the BCO-DMO's mission and of sufficient interest to merit inclusion.

#### in situ oceanographic measurements

pressure, temperature, salinity, oxygen, turbidity, chlorophyll

#### chemical and nutrient measurements

salinity, oxygen, phosphate, nitrate, nitrite, silicate, chlorophyll, ammonium

#### community composition

viruses, bacteria/archaea, phototrophic picoeukaryotes, nanoplankton, microplankton

#### productivity

primary productivity (<sup>14</sup>C bicarbonate uptake), phytoplankton growth and mortality rates, bacterivore activity, secondary productivity (<sup>3</sup>H-thymidine and -leucine incorporation)

### **Sequence data**

With the recent demise of NCBI's SRA repository, solutions for archiving sequence data are being explored. Pyrosequence data generated for the project will be submitted to both the CAMERA and MG-RAST databases. Metadata describing the samples and sample sites will be provided in a format compliant with the Genome Standards Consortium MIMS/MIENS standard (Field et al., 2008).

pyrotag rDNA (bacteria, archaea, protists) (SPOT, 5 depths, monthly samples)

prokaryotic metatranscriptome (quarterly transect, year 1 only)

target gene sequences (quarterly transect, years 2-4 only)

prokaryotic pyrotag rRNA (quarterly transect)

### **Data storage, distribution, and presentation**

Data distribution can be an issue in projects that bring together so many labs. We will have two mechanisms to ensure open and timely distribution of raw data and results from analysis. One will be to have a private data server to which all member labs are expected to upload project data, and from which any member lab can download data. Funds for equipment that can serve this purpose have been requested in the Budget. In addition, as part of the biweekly project-coordination meetings outlined in the Overall Management Plan, one agenda item will be Data, under which we will discuss what data has been generated and its availability. Data will not be distributed to outside labs without discussion and consensus at one of the group meetings.

A project-specific public website will be developed as another mechanism for data archiving and distribution. Raw project data will be available for viewing and download. Sample site chemical and oceanographic data will be distributed as tab-delimited text files. Graphical representations of the data will be integrated into the site. Sequence data will be distributed as fasta files. OTU groupings and taxonomic identifications will be available as tab-delimited text files. Precalculated trees and graphs of correlation and diversity analysis will be presented.

#### REFERENCES

Field, D., Garrity, G., Gray, T., Morrison, N., Selengut, J., Sterk, P. *et al.* (2008) The minimum information about a genome sequence (MIGS) specification. *Nat Biotechnol* **26**: 541-547