

## **Data Management Plan**

### *Types of Data/Materials to be produced*

We anticipate about 10 samples will be collected the first year and about 100 samples each year for the second and third years, a total of about 200 samples. All data related to diversity, function, growth, nutrients, DOM and standard oceanographic properties, such as solar irradiance, attenuation of light within the water column, temperature and salinity measurements will be considered primary data (see below).

### *Standards for data and metadata*

All data collected will be in compliance with the current Minimum Information about a Genome Sequence (MIGS) specification. All data and metadata be collected according to the latest technologies.

### *Data access and sharing*

All primary data will be recorded in laboratory notebooks and results compiled in files stored on local hard drives. The data will be routinely backed up to external hard drives. Extracted nucleic acids and associated samples will be archived in an -80°C ultralow freezer. The freezer is fitted with an alarm and phone notification system in the event of failure. All relevant nucleic acid sequence data will be deposited in the NCBI GenBank database as well as CAMERA, MG-RAST and IMG/M public databases (Field and Kyrpides 2007; Markowitz et al. 2008; Meyer et al. 2008). We will set up a website for the project that will link these data with other data sets, including the transcriptomic data and the biogeochemical properties.

Results of this project will be shared with the research community through a series of publications in peer-reviewed journals. The PIs and graduate students will participate and present results at national and/or international conferences and symposia. A description of the project with our objectives and results of our research will be made available to the general public on a website. The research will also be relayed to the general public at Coast Day, a public outreach effort at UD College of Earth, Ocean, and Environment, which is attended by over 10,000 people annually.

### *Distribution of derivatives*

When possible, nucleic acids or libraries/clones produced during this project will be made available to other researchers upon completion of this project.

### *Plans for archival of data*

Primary data will be archived on local computers stored in the PIs laboratory and in laboratory notebooks for at least three years after the end of the project. Backup hard drives will also be stored in a separate location.