

Data Management Plan

Introduction:

Data management will be coordinated by the Principal Investigators, and carried out by all project participants. The plan encompasses three areas: use policies, standards, and data preservation and archival. We will leverage existing systems wherever possible. All data will be communicated in a timely fashion following the NSF policy.

Types of data to be collected:

In terms of field data, we will measure abundance of *Synechococcus* and dominant diatom species at the BATS reference station and station thereabouts. We will also collect field data on size distribution of biogenic and lithogenic silica, elemental composition of *Synechococcus* and dominant diatom species (in terms of biomass) and molecular characterizations of *Synechococcus* communities. In addition, we will also produce a large amount of laboratory data on the accumulation of Si by many strains of *Synechococcus* under different culture conditions, the chemical nature of the Si associated with *Synechococcus* cells, the effect of Si uptake on growth and size of cells, the subcellular distribution of that Si within Si accumulating clones and correlations between Si uptake and cell densities. Bioinformatic analyses will identify candidate genes related to Si uptake and attempts will be made to develop molecular markers that can predict silicic acid uptake.

Data access and sharing policies:

Data collected under the project will be made available to the public with as few restrictions as possible. For data collected during BATS cruises, our policies will cohere with standing LTER policies regarding access. For experimental work, we plan for publication of most data with metadata after or in conjunction with primary publication of results, or at most two years after the completion of the study on a server hosted by SBU.

Plans for archiving and preserving data:

Hydrographic data collected during the BATS cruises will be managed and archived as per LTER policies. Information from cruises generated by research related directly to this proposal, including biomass estimates for diatoms and *Synechococcus*, measurements of size fractionated biogenic silica, final estimates of cellular elemental composition for diatoms species and picocyanobacteria, and molecular data on *Synechococcus* communities will be deposited as addenda to the LTER data for that date. We will establish a server at SBU for distribution of SXRF spectral data, with metadata and stand alone IDL Virtual Machine software required for inspection and analysis. Molecular data will be deposited in appropriate databases, depending on the nature of that data. In all our efforts we will work with the Biological and Chemical Oceanography Data Management Office (<http://www.bco-dmo.org>) to archive the data and to ensure our metadata conform to their standards.

Standards and formats to be used for metadata and data:

We will conform to the metadata standards of BCO-BMO. Data will be archived with metadata, unless there are proprietary restrictions (as described above). As much as possible, data will be archived in ASCII format, which is the most flexible and readable over the long term. We will archive data in tabular formats that have been proven successful when sharing data among the

project collaborators. In some cases, as with the SXRF and XANES/EXAFS spectral data, we will have to develop our own metadata standards to promote easy and fruitful access to these data. These standards can build on those employed by the national labs.