Data Management Plan

Introduction:
Data management will be coordinated by Marchetti 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.

Several types of data will be generated in the proposed activities. In the lab, types of data include phytoplankton isolate physiological measurements (i.e. growth rates and photophysiology), cellular elemental quotas and transcriptome sequences. Types of field data include oceanographic survey data (e.g. ship underway and CTD data), chemical (i.e. dissolved and particulate nutrients) and biological measurements (phytoplankton cell counts and biomass, NO_3 uptake rates, photophysiology, etc.) and transcriptome and genome sequences.

Data access and sharing policies:
We are committed to making all data types publicly available through peer-reviewed publications and public databases with as few restrictions as possible. As sequence libraries are created and samples are analyzed, data will be processed, and raw and processed data will be uploaded to networked servers maintained at UNC-Chapel Hill. These servers are backed up weekly, and can be made available to collaborators needing access to the data. Data will be transferred to BCO-DMO following processing, and public access will be granted to data following its publication or at most two years after its collection.

Formats to be used for metadata and data:
We will conform to the metadata standards established by the Biological and Chemical Oceanography Data Management Office (BCO-DMO). As much as possible, data will be archived in ASCII format, which is the most flexible and readable over the long term. In the case of MS data, calculated particulate element concentrations will be reported to BCO-DMO along with detailed information on the number and types of blanks (e.g., analytical instrument blanks, digestion blanks, filter blanks, etc.).

Plans for archiving and preserving data:
Sequences obtained through Illumina high-throughput sequencing platforms will be stored in long-term storage space provided to Marchetti by UNC-CH Research Computing and deposited in public sequence databases: all individual gene sequences will be deposited in the National Center for Biotechnology Information (NCBI) GenBank and all large sequence libraries will be deposited into the Sequence Read Archive (SRA) managed by NCBI. All other data will be archived at BCO-DMO. 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.