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.

Four main types of data will be generated in the proposed activities: transcriptome sequence libraries, targeted gene expression, particulate metal concentrations, and cell-specific metal quotas. Transcriptomic sequence libraries will be created using the Illumina HiSeq 2000 platform. Gene expression analysis will be performed using quantitative PCR. Particulate trace metal samples will be analyzed by ICPMS, producing particulate concentrations for a suite of trace elements (Al, Ti, P, Mn, Co, Ni, Cu, Zn, V, Cd, and Mo) in addition to iron. Cellular concentrations of a subset of these metals (P, Mn, Fe, Co, Ni, Cu, Zn) will be measured by SXRF in diatoms collected from the mixed layer at stations, as well as from incubation experiments and laboratory experiments. Iron quotas of diatom isolates will be measured using a dual label $^{55}Fe/^{14}C$ radiotracer technique. Other data that generated in this project includes physiological measurements on phytoplankton growth characteristics and phytoplankton composition (e.g. light microscopy counts and flow cytometry) of field samples.

Data access and sharing policies:
We are committed to making both data types publicly available through peer-reviewed publications and public databases with as few restrictions as possible. As sequences 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 and Bigelow Laboratory. 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.

Standards and 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 ICPMS data, calculated particulate trace element concentrations will be reported to BCO-DMO along with detailed information on the number and type of blanks (e.g., analytical instrument blanks, digestion blanks, filter blanks, and process blanks that integrate the entire sample collection, digestion and analysis process). We will also report concentrations of commercially-available standard reference materials (PACS-2: Marine Sediments for Trace Metals, NRC Canada; BCR-414: Trace elements in plankton, Community Bureau of Reference, Institute for Reference Materials and Measurements, Belgium) that are digested and analyzed in the same batches as the cruise samples. In the case of the SXRF data, which does not readily conform to existing oceanographic data formats, we are developing our own metadata standards to promote easy and fruitful access to these data. These standards can build on those employed by the national labs. Mean trace element quotas for each diatom group analyzed will be

reported, delineated by depth and station, along with the number of cells analyzed, the range and standard deviation of the results.

Plans for archiving and preserving data:
Sequences obtained through Sanger DNA sequencing or 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 (e.g. ferritins) will be deposited in the National Center for Biotechnology Information (NCBI) GenBank and all sequences obtained through Illumina sequencing will be deposited into the Sequence Read Archive (SRA) managed by NCBI. All other data will be archived at BCO-DMO. Additionally, we will establish a server at Bigelow Laboratory for distribution of SXRF spectral data, with metadata and stand-alone IDL Virtual Machine software required for inspection and analysis. This server will be backed up regularly to ensure data safety and integrity. 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.