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: proteomic mass spectral data, metabolomics mass spectral data, particulate metal data and cell specific metal quotas. We are committed to making all data types publicly available through peer-reviewed publications and public databases.

Data access and sharing policies:

Mass Spectrometry Data (proteomic and metabolite data): Tracking, software, and automated pipeline development procedures have been initiated for proteomic data collected from the South Atlantic (Morris OCE-0723866). Support requested in this proposal will strengthen these efforts and improve data dissemination. There is currently no standard public repository for raw mass spectrometry files. However, a minimum standard for data access will be to make these files publicly available. The Mass Spectrometry Center at the University of Washington has this capability and will make all raw mass spectrometry files available upon publication.

These data should also be available in a centralized public database. We will work closely with the proteomic community to enhance available work-flows that are available for analysis of these kinds of data, such as the DOE Systems Biology Knowledgebase or CAMERA. The Knowledgebase is an online resource that is under construction. CAMERA is a project funded to develop the necessary cyberinfastructure and tools needed to facilitate the integration and utilization of environmental and genomic data. Although CAMERA's initial focus is on DNA sequence data, storage and searching of proteomic data should be within their long-term mission. Metabolomics and vitamin data will be submitted to BCO-DMO as has been done previously by the Ingalls group. 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.

Metals data (particulate metal and cell quotas)
Particulate trace metal samples will be analyzed by ICPMS, producing concentration profiles of particulate Al, P, Ti, Mn, Fe, Co, Ni, Cu, Zn, Cd, and Mo. Cellular concentrations of a subset of these metals (P, Mn, Fe, Co, Ni, Cu, Zn) will be measured by SXRF in major taxonomic groups at each station.

Data collected under the project will be made available to the public with as few restrictions as possible. For data collected during the cruise, 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 the BCO-DMO website.

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 data, 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.