Data Management Plan

Two general types of data will be produced from this project:

1. **Culture data:** This portion of the proposed work will include results from many laboratory experiments on the model diazotrophs *Trichodesmium* and *Crocosphaera.* These efforts will be focused on developing gene markers for balancing limitation that are tuned to biogeochemically relevant rates (i.e., growth rate and N_2 fixation). We anticipate this resulting many types of data, including: experimental physiological, proteomic, transcriptomic and biochemical.
2. **Field data:** Here we will be surveying natural diazotrophic populations for evidence of balancing limitation during a Y2 R/V Atlantis cruise across the Western North Atlantic. Pre-cruise planning will be done via teleconferencing and a PI planning workshop. Detailed plans for station locations, instrument deployment, water sampling strategy and water sample allocation will be written up as a science implementation plan for the cruises. The actual sampling events will be recorded on paper logs (scanned into PDF documents) and in a digital, searchable event log, which can be provided to BCO-DMO as metadata when the data are submitted there. During this cruise we will not only collect water-column biogeochemical, physical, and rate date but also archive samples for gene expression, metaproteomic, metatranscriptomic, and chemical measurements.

Data Archiving:

Upon receipt of the award we will contact the Biological-Chemical Oceanography Data Management Office (BCO-DMO: http://www.bco-dmo.org/) to register our project. We will submit all data collected from the field and laboratory portions to BCO-DMO for archiving per NSF guidelines. All biochemical and genetic material produced will be stored at -20 or -80 °C, as appropriate for the sample type. In addition to our findings, a searchable database containing the location of all processed and non-processed materials in long term storage will be maintained by the PIs and submitted to BCO-DMO.

Data Release:

Access to data will be given once it is quality controlled and published, or at any rate always within 2 years of collection. Availability will be in accordance with NSF guidelines for data accessibility. We will keep NSF abreast of our compliance with data management through our annual reports and all data will be made available to the general public as expeditiously as possible. We will work closely with the BCO-DMO to ensure that data used in our analyses and outcomes from the proposed experiments are publicly available according to NSF guidelines. Further, all data made available will be accompanied by compliant metadata. DNA sequences generated from this work will be deposited in Genbank at the National Center for Biotechnology for Information (NCBI) within three months of the end of this project (http://www.ncbi.nlm.nih.gov/geo/). Our proteome and metaproteome data will be archived in the Ocean Protein Portal EarthCube Building Block Prototype (recently recommended for funding), a unique proteomics database being developed by PI Saito in collaboration with BCO-DMO at WHOI, allowing broad access to these unique data and contributing content to the portal prototype.

As we have done in the past, we will continue to make the results of our work available to the marine science community through timely peer-reviewed publications and professional meeting presentations. Importantly the budget contains publication and travel funds that will cover the costs of publications and oral and poster presentations at professional venues like ASLO/Ocean Sciences, and Gordon Research Conferences.