DATA MANAGEMENT PLAN

Our data management effort will begin with an all-PI meeting in Year 1 before the first small boat sampling campaign, where we will finalize the sampling strategy and establish coordinated data collection protocols within the group. We will also formalize sample naming and numbering conventions for the small boat and cross-shelf campaigns. During each cross-shelf cruise, we will generate a coordinated sampling event log (paper and digital). Per the policy of the NSF Division of Ocean Sciences, we will submit our data no later than two years after they are generated to the online **Biological and Chemical Oceanography Data Management Office (BCO-DMO)** data repository for public use.

I. Types of data

The proposed research project will generate data from the analysis of field and experimental samples collected during the two UNOLS project cruises as well as four small boat sampling efforts. The field data will include water column, estuarine, groundwater, and riverine macronutrient concentrations (nitrate+nitrite, nitrite, phosphate, silicate, ammonium, dissolved organic nitrogen, dissolved organic phosphorus), nitrogen isotopic ($\delta^{15}N$) measurements of nitrate+nitrite and dissolved organic nitrogen (and ammonium, when >5 µM), and concentrations and isotopic composition of suspended particulate organic carbon and suspended particulate organic nitrogen; experimental datasets for these parameters will be generated from the proposed shipboard incubations. Additionally, micronutrient concentration measurements (barium, dissolved iron and other trace metals), iron speciation (iron-binding ligand concentrations and conditional stability constants), and dissolved iron isotope ($\delta^{56}Fe$) measurements will be made on water column, estuarine, groundwater and riverine field samples, as well as on incubation samples. Surface, bottom, estuarine, groundwater and riverine samples will also include radium isotope measurements (^{223}Ra, ^{224}Ra, ^{226}Ra, ^{228}Ra and parent isotopes ^{227}Ac and ^{228}Th). We will measure chlorophyll *a* concentrations, carbon and N_2 fixation rates, and quantitative PCR data from molecular markers monitoring gene abundance and gene expression of iron stress genes of both *Trichodesmium* clades from the water column samples and shipboard incubations. Finally, we will use liquid chromatography tandem inductively coupled plasma mass spectrometry and electrospray ionization mass spectrometry (LC-ICPMS-ESIMS) to quantify and confirm the identity of metal binding organic compounds including siderophore. The LC-ESIMS data will also be processed in CoreMS to identify nitrogen containing organic molecules to generate 'feature lists'. Data generated from these analyses include raw data, converted open-source data files (mzXML), and processed data stored as CSV files. The USGS marine groundwater wells will be equipped with sensors that measure and continuously log conductivity, temperature, and pressure of the marine groundwater for the duration of the project to provide annual context of each shelf-wide and coastal sampling. The datalogging sensors will be replaced during each coastal sampling cruise. The raw data will be published as an official USGS Data Release following USGS Fundamental Science Practices and per federal mandate will be open access.

II. Data and Metadata Standards

The water column chemical and biological data generated from the cruises will be integrated with supporting hydrographic (CTD and bottle) data to create a merged seawater data set. All field and shipboard incubation data and associated metadata will be submitted by project PIs to the Biological and Chemical Oceanography Data Management Office (BCO-DMO), to be made publicly available online from the BCO-DMO data system following standard NSF requirements. All of the PIs have prior experience with this data system. All discrete samples will be assigned a unique sample identifier (IGSNs), using the IEDA IGSN assignment tool. Metadata standards and data management compliance will be further explained to NSF through our annual project reports.

DATA MANAGEMENT PLAN

III. Policies for access and sharing and provisions for appropriate protection/privacy

Water column, estuarine, riverine and groundwater macro- and micro-nutrient concentration and isotopic data, including radium isotope data, will be deposited to BCO-DMO as described above. LC-ICPMS-ESIMS data (both raw and processed) will be available in open-source format in the searchable public database maintained by the UCSD Center for Computational Mass Spectrometry (http://massive.ucsd.edu/ProteoSAFe/static/massive.jsp). The MassIVE repository provides accession numbers for these data sets that will be linked to BCO-DMO and cited in publications that make use of the data. The PIs have a history of reporting data in figures as well as in supplementary tables for ease of use by other researchers, and will continue to publish data in this format as well.

IV. Policies and provisions for re-use, re-distribution

There will be no permission restrictions for these data. The data may be of interest to chemical, geological and biological oceanographers. The intended and foreseeable users of the data are oceanographers, geologists, modelers, ecologists, and microbial ecologists within academia and government labs. It is anticipated that other scientists will compare their sequence information to nucleic acid sequences generated in this study via NCBI.

V. Plans for archiving and preservation of access

Initially, all data will be archived on computers in the respective labs of the PIs, and backed up on remote servers and/or external hard drives. Data will be submitted to public databases (BCO-DMO and MassIVE), where they will be permanently archived to preserve access to the public. A hard copy of all notes (i.e. lab notebooks) will be retained in the laboratory. All relevant metadata associated with genomic libraries will be submitted along with the nucleic acid sequences themselves. Research publications generated from this work will include all relevant data and refer readers to public databases where data is permanently archived.

Biological and Chemical Oceanography Data Management Office (BCO-DMO)

The Biological and Chemical Oceanography Data Management Office (BCO-DMO) was created in late 2006 to serve PIs funded by the NSF Geosciences Directorate (GEO) Division of Ocean Sciences (OCE) Biological and Chemical Oceanography Sections and (with augmented funding in 2010) Office of Polar Programs (OPP) Antarctic Sciences (ANT). BCO-DMO manages and serves oceanographic biogeochemical, ecological, and companion physical data and information developed in the course of scientific research and contributed by the originating investigators. The BCO-DMO data system facilitates data stewardship, dissemination, and storage on short and intermediate time-frames.