

DATA MANAGEMENT PLAN

- (1) Types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project

Field experiments will generate data from processing of physical, chemical and biological samples as well as data from *in situ* measurements. All relevant field data will be supplied to the Biological and Chemical Oceanography Data Management Office (<http://bco-dmo.org/>). Data to be submitted to BCO-DMO will include: DIC, pH, nutrients (NO₂, NO₃, PO₄, SiOH₄, NH₄), salinity, chlorophyll, temperature, oxygen, Secchi depth, turbidity and flow cytometrically measured microbial populations (*Prochlorococcus*, *Synechococcus*, picophotosynthetic eukaryotes, bacterioplankton) as well as all meta-sampling data. All molecular data (nucleic acid sequences) will be deposited at the National Institutes of Health's NCBI (<http://www.ncbi.gov/>). Software or analysis tools or schemes will be made available to the scientific community through a project website, publications, upon request or deposited to BCO-DMO where appropriate. Other data relevant to publications that does not conform to the above databases will be deposited in the Dryad database, (<http://datadryad.org/>). Samples from this project will be made available, as requested, to the wider scientific community if this does not interfere with the goals of this project. Educational and outreach materials (e.g. teacher blogs, lesson plans, etc.) will be made publically available through a project website or through broader educational material repositories (e.g. <http://www.learnnc.org/> and <http://opened.creativecommons.org/>).

- (2) Standards to be used for data and metadata format and content

Following standard biological, chemical and physical and oceanographic practices, all data will be collected using certified reference materials (CRMs) where available. These include CRMs for nutrient analyses (NRC Institute for National Measurement Standards, <http://www.nrc-cnrc.gc.ca/eng/services/inms/reference-materials.html>), carbon analyses (<http://andrew.ucsd.edu/co2qc/>) and chlorophyll (http://www.turnerdesigns.com/t2/doc/appnotes/998_0058.html) among others. Where CRMs are not available, other best practices will be used including using internal standards or appropriate controls. Data will be formatted in accordance with archival database repositories using requested formats. Specifically, all molecular (i.e. DNA sequence) data will be submitted to NCBI using BankIt or Sequin (<http://www.ncbi.nlm.nih.gov/guide/howto/submit-sequence-data/>). Environmental data including metadata from research cruises will be submitted to the Biological and Chemical Oceanography Data Management Office following their submission procedure (<http://bco-dmo.org/>). Our past NSF-sponsored research projects have used all of these approaches and data repositories and their inclusion and execution is a standard component of our project management.

- (3) Policies for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements

Using established communications channels (e.g. Cisco telepresence, Skype, data servers, etc.) all PIs and their students and staff will have equal access to all data generated from this project as is reasonably achievable. Data and communications will be housed and conducted on secure data servers and channels or located in locked rooms while under active analysis. When appropriate and ready for a wider distribution, data and project products (i.e. papers, software, etc.) will be publically disseminated using the above archival databases (e.g. BCO-DMO, NCBI, etc.). Given

the nature of this research, security or intellectual property rights are not a concern. Data access rights will follow standard NSF project management rules.

(4) Policies and provisions for re-use, re-distribution, and the production of derivatives

Once publically disseminated to the archival databases described above, all data and research products will be available for unrestricted re-use or re-distribution. Acknowledgement of the original source of the data will be requested, but not required for its use.

(5) Plans for archiving data, samples, and other research products, and for preservation of access to them

As outlined above, all data and research project products will be archived to publically accessible database repositories or scientific journals. All samples will be archived in deep-freezers (-80°C or -196°C) for the duration of the project and while manuscripts or data quality control are underway. To maximize efficiency in sample processing and storage, all samples will be processed (e.g. nucleic acid extraction) and stored at Duke University but made available to Georgia Tech upon request. Beyond this period, samples will be stored as required by the project or for the period that they are still valuable (i.e. have not degraded beyond their utility).