

DATA MANAGEMENT PLAN

Data description: Field sampling will take place in conjunction with the Narragansett Bay Long Term Time Series and the Bermuda Atlantic Time Series. The time series collect and make public all core measurements (e.g., <http://www.gso.uri.edu/phytoplankton/>). The proposed project includes isolation and characterization (physiological and genomic) of diatoms and cyanobacteria from each field-sampling event. Additionally, incubation experiments will be conducted using the whole plankton community collected at each sampling event. These data will be combined and used as the basis for the modeling effort. In short there are four major data types **1) physiological data, 2) sequence data, 3) biological cultured isolates and 4) model code and output.**

Data release: All the appropriate data will be deposited to the Biological and Chemical Oceanography Data Management Office (BCO-DMO) website within two years of collection as required by the OCE Policy. **1)** Physiological data from both laboratory and field incubation experiments will be uploaded, along with tags to the publicly available time series data, to the BCO-DMO website, where it will be released to the wider community as the data are published. **2)** All appropriate field data products associated with the sequencing will be submitted to BCO-DMO (<http://www.bco-dmo.org/>) as highlighted above with a special link to the sequence archive. Both amplicon sequences of rDNA and partial genome sequences used for SNP analyses will be uploaded to the National Center for Biotechnology Information (NCBI) sequencing read archive. Links to these accessions will be uploaded to the BCO-DMO project page. **3)** Cultured isolates will be a) submitted to anyone who places a request pending sufficient material and b) archived in the Provasoli-Guillard National Center for Marine Algae and Microbiota (NCMA—formerly CCMP). **4)** The eco-evolutionary models developed and model output will be made freely available with user-friendly annotation upon publication through our MSU lab server or as supplementary information to publications. GCMs produce significant amounts of data, presenting challenges for data distribution. Consequently, only model output related to published results will be made available, except on request. Data will be released in netCDF format, through BCO-DMO and/or another public repository such as Dryad (<http://datadryad.org/>) or a lab or personal server if necessary. In some cases, a compromise can be struck by reducing file sizes through judicious averaging to reduce spatio-temporal resolution. In this eventuality, original data sets will remain available by request. In addition to model output, the models created through this project represent important research products. In keeping with GFDL's long history of facilitating public access to code and documentation, models that we develop adding evolutionary dynamics to GFDL's GCM will be made publicly available. This will be facilitated through either GFDL's public ftp site, or comparable web services at Princeton, following the publication of the model and results in peer-reviewed journal(s).

Data archiving: **1)** Upon receipt of the award we will contact BCO-DMO to register our project. We will submit all data collected from the field to BCO-DMO for archiving as highlighted above. Physiological and community structure data from the incubation experiments will be assembled and organized in electronic spreadsheets and stored on local and backup servers, prior to submission to BCO-DMO. We will submit all data upon publication to BCO-DMO for archiving in a searchable project format. 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. **2)** We will be generating and storing moderate amounts of sequence data and the associated analytical files. The data will be stored on a 15TB Raid 5 server and to external hard drives that are stored in a separate location. In this manner there is redundancy in preserving the raw data and the associated analytical files. Public archives will be

used where possible (e.g. NCBI). **3)** Diatom species isolated from our field studies will be cultured in at least two labs as a bulwark against their loss. Further, cultures will be deposited in the Provasoli-Guillard National Center for Marine Algae and Microbiota (NCMA—formerly CCMP) whenever possible. **4)** The models and model output will be archived in several ways: on our MSU lab server, as the supplementary information in publications, the BCO-DMO site or the GFDL public ftp site.