

## Data Management Plan

The data management procedures in this project will adhere to and are guided by the requirements of the Division of Ocean Sciences Sample and Data Policy as described in document NSF 17-037 (<https://www.nsf.gov/pubs/2017/nsf17037/nsf17037.jsp>). Environmental datasets will be made available to the Biological and Chemical Oceanography Data Management Office (BCO-DMO). Ecophysiological datasets will be made available to Dryad and multiomics datasets will be made available to the National Center for Biotechnology Information (NCBI), each with reference to environmental datasets at BCO-DMO.

### Data collection

**Field studies:** Field studies will generate both observational and experimental datasets. Environmental data collected at each site will include GPS data, PAR light profiles, water chemistry, and water temperature. Observational ecophysiological datasets will include PAM fluorescence, sample surface area, growth rate, DNA metagenomic and RNA metatranscriptomic datasets. All data will be deposited in the databases discussed below. All observations and data from each sampling event will be transcribed from dive slates, instruments, and photographs into field notebooks (e.g., GPS locations, additional comments) or recorded in an excel file if field measurements are obtained from digital photograph analysis (i.e., surface area, growth rate) or instruments (i.e., PAM fluorometry). All field notebooks will be digitized to PDF format.

**Lab studies:** Lab studies will generate experimental datasets on field-collected samples cultured under varying environmental manipulations. Temperature and light profiles from each growth chamber will be logged and linked in experimental metadata. Physiological experiments will generate photosynthetic data, PAM fluorescence, sample surface area, growth rate, DNA metagenomic and RNA metatranscriptomic datasets, and nutrient and metabolic product profiles. All data will be deposited in the databases listed below. Lab notebooks will be kept in the McCoy Lab and digitized to PDF format. Data obtained from digital instruments will be recorded in an excel file.

**Data analysis:** The code for statistical analysis and modeling of the environmental and physiological data generated through the work will consist of annotated R scripts (code) and R dataframes (output) with associated metadata. Bioinformatics workflows used in the manipulation and analyses of the multiomics data will also be preserved and made available with associated metadata.

### Documentation and metadata

Metadata of field-collected datasets will include date, time, GPS coordinates and a description of collected samples/species, allowing users to identify the collection history of each sample and/or data point. Metadata of lab-generated experimental datasets will include date, time, a description of relevant samples, and information about where and when those samples were collected that references the appropriate field dataset. Metadata will follow the established guidelines for environmental data (i.e. following NSF LTER protocols for consistency with other datasets). Metadata files will link raw data files to processed data files, and field sampling, and additionally link to digitized field notes for reference. Metadata will be compiled and provided to the Biological and Chemical Oceanography Data Management Office (BCO-DMO). A Dataset Metadata Form will be completed for each dataset contributed to BCO-DMO. Ecophysiological data submitted to Dryad and multiomics data submitted to NCBI will be similarly prepared and

metadata will include citations for the environmental dataset at BCO-DMO. All data will be submitted using SI units and commonly used units appropriate to each measured parameter.

### **Data availability**

Metadata files, full data sets, derived data products and physical collections will be made publicly accessible within two years of collection. The Biological and Chemical Oceanography Data Management Office will be used as the data management archive for environmental oceanographic data. When the award is initialized, BCO-DMO will be contacted the project will be registered by submitting project metadata. Dryad will be used as the primary data management archive for physiological data, with reference to the BCO-DMO dataset. This will make each dataset most readily available to the relevant research communities. All data in Dryad are released to the public domain without legal restrictions on reuse, through a Creative Commons Zero waiver. Updates on the status of metadata and data archival will be included in the Annual Project Reports, and the compliance with this Data Management Plan will be documented in the Final Project Report. For data submissions that are due after the Final Report, we will report plans for final data submission.

### **Publication of data and analyses**

Results of this project will be presented at national and international science conferences (e.g. Phycological Society of America, Ocean Science Meeting) and published in peer-reviewed papers submitted to international scientific journals. With these publications, analytical scripts and subsets of the post-processing data will be made available also as in Web Appendices that several journals now manage (e.g. L&O).

### **Data formats and dissemination**

All laboratory/field notebooks and datasheets will be scanned and stored electronically, along with other information relevant to the collection, processing, and analysis of the samples. Metadata and measured data will be stored in electronic spreadsheets (e.g., .csv for use with Microsoft Excel, Access, or other programs). Meta-data from instruments (MS, sensors) will be downloaded and stored on McCoy lab servers hosted at UNC Biology.

### **Educational and mentorship assessment**

Datasets obtained from educational and mentorship assessments related to the educational plan and broader impacts will not be made publicly available. These assessments will be made anonymously and will not be used outside of the explicit purposes of internal assessment and improvement. Course data collected for participation in SEISMIC educational research is likewise collected anonymously. IRB approval for SEISMIC research associated with *Environmental Microbiology* and *Marine Ecology* is pending at UNC and has already been given at the lead institution (U. MN).

Preliminary data on fish behavior was exempted from animal use ACUC protocols by Florida State University review. The current project proposes to study only simulated herbivory based on previously collected data.

We do not anticipate that there will be any significant intellectual property issues involved with the acquisition of the data. In the event that discoveries or inventions are made in direct connection with these data, access will be granted upon request once appropriate invention disclosures and/or provisional patent filings are made. Key data relevant to the discovery will be preserved until all issues of intellectual property are resolved. The data acquired and preserved in the context of this proposal will be further governed by the UNC's policies pertaining to intellectual property, record retention, and data management.