

## Data Management Plan

The data that will be produced from this project will include results from laboratory culture experiments with phytoplankton isolates, field work from a seasonal study in coastal California, and earth systems modeling output. These efforts will be focused on understanding how temperature and nutrient interactions affect phytoplankton Elemental Use Efficiencies, and ultimately primary production by important phytoplankton groups. Measurements in both lab and field studies will include rates of growth, elemental quota data (C,N,P,Fe). CO<sub>2</sub> fixation, N<sub>2</sub> fixation where appropriate, cell counts with flow cytometry and microscopy, and chlorophyll. Communities in the field incubation experiments will be characterized using 16S/18S rRNA amplicon sequencing, flow cytometry, and microscopy. This project will thus produce data that will include experimental, biogeochemical, physiological, genomic, and modeling datasets.

**Sample Archiving:** All biochemical and genomic material produced will be stored at -20 or -80 °C, as appropriate for the sample type. Cell count samples will be stored in 4% buffered glutaraldehyde. A database containing the location of all materials will be maintained for use by all PIs, and submitted along with the data to BCO-DMO.

**Laboratory and Field Data Archiving:** All data collected during the laboratory- and field-based experiments will be securely stored in multiple redundant formats (hand-written lab notebooks, backed-up files on discs/hard drives) indefinitely in the PI's offices. 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 storage will be maintained by the PIs and submitted to BCO-DMO.

**Model Data Archiving:** Model data to be generated during this project includes new FORTRAN code of the EUE paramterisations and netcdf files containing key results and their boundary conditions. The FORTRAN files pertain to any modifications to the core PISCES code that is stored on a svn server by the Institute Simon LaPlace in Paris. Thus, only the modified routines will be archived, via the PISCES Gitlab page ([gitlab.in2p3.fr/](http://gitlab.in2p3.fr/)) and are small (~KB) files. Key results and boundary conditions are larger files (GB-TB size range) and, as is standard for all NERC-supported science, will be archived by the British Oceanographic Data Centre in Liverpool, UK. In addition to this core archiving, model results and code that underpin publications arising from the project will also be archived on GitHub and Zenodo, generating citable digital object identifiers.

**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. Raw nucleic acid sequencing reads will be deposited to the National Center for Biotechnology for Information (NCBI) Sequence Read Archive ([ncbi.nlm.nih.gov/sra](http://ncbi.nlm.nih.gov/sra)) and Gene Expression Omnibus ([ncbi.nlm.nih.gov/geo/](http://ncbi.nlm.nih.gov/geo/)) databases within three months of the end of this project, with appropriate links from our project metadata page at BCO-DMO. Furthermore, all raw and processed amplicon sequence data will be backed up and stored at the University of Southern California High-Performance Computing Cluster ([hpcc.usc.edu](http://hpcc.usc.edu)).

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. 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.