

## Data Management Plan

### I. Title and contact information

1. *Project title:* “RAISE: Linking the adaptive dynamics of plankton with emergent global ocean biogeochemistry”

2. *Points of contact:*

Overall data products and chemical data – Adam Martiny, [amartiny@uci.edu](mailto:amartiny@uci.edu), (949) 824 9713.

Proteomic analyses – Kate Mackey, [kmackey@uci.edu](mailto:kmackey@uci.edu), (949) 824-1133.

Trait Model analyses – Simon Levin, [slevin@princeton.edu](mailto:slevin@princeton.edu), (609) 258-6880

Ocean Biogeochemistry model analyses – Francois Primeau, [fprimeau@uci.edu](mailto:fprimeau@uci.edu), (949) 824-9435.

### II. Types of data

**Data description.** Processed chemostat concentrations of cells, nutrient, cellular quotas (CHNOPS), carbohydrates, lipids, proteins, nucleic acids in CSV files and raw text files containing chemical analysis. Processed proteomic data as counts pr. gene in CSV files. Raw Mass Spec Proteomic data in XML format. Model code and outputs will be in ASCII files. The expected size of collection is less than 100GB.

### III. Data and metadata formats, standards, and organization

**A. Formats.** ASCII formats, with comma or tab-separated values and/or in spreadsheet format, will be generated, maintained and submitted to Biological Chemical Oceanographic-Data Management Office (BCO-DMO). Data generation will not be large so there is no need to compress files using binary format. ASCII data is universally readable by most if not all analysis software. The complete data product from the proposed work will be made available through the BCO-DMO. XML data for proteomics will be maintained and submitted to the PRIDE Archive (<http://www.ebi.ac.uk/pride/archive/>). Model code will be in ASCII format and submitted to GitHub. Model outputs will be in ASCII format and submitted to BCO-DMO.

**B. Metadata.** As done in the past, we will use the BCO-DMO metadata authoring tool to prepare our information for submission to the archive. We will work with BCO-DMO to include the biochemical parameters that we have used in this study are properly described in their system. Similarly, we will work the submission protocol for the PRIDE Archive and GitHub.

**C. Data organization.** Data will be organized by PI Adam Martiny with help from . A sample log and laboratory notebooks will be kept in paper copy and transcribed to a spreadsheet recording available meta information, including the time/location/physical conditions during chemostat sampling to build the first stage of a data archive. Data will be stored on a desktop computer and backed up to a remote storage system at UCI in ASCII format. All file names include DDMMYYYY and are renamed as they are re-saved.

**D. Data quality.** Routine instrument calibration for measurement of all parameters is conducted using appropriate calibration standards reference materials. Equipment will be regularly calibrated. All notebooks and written sample logs will be scanned and stored electronically, as well as other information relevant to the collection, processing, and analyses of the samples. Data files and information will be kept on the computers of the respective PIs, who all have established robust institutional data management policies. The designated person of responsibility is Adam Martiny, [amartiny@uci.edu](mailto:amartiny@uci.edu), (949) 824 9713.

### III. Data access and sharing

The data and model will be of exclusive use by the PIs until publication or one year after the end of the grant. The project will upload data files and collection-level metadata to BCO-DMO once submitted for publication. BCO-DMO will make the data publicly available through their system immediately upon receipt. Data may be parsed into subsets that work with the format of the data. All datasets for the project will be connected such that anyone looking at the project will see all the data as we've done with prior research projects.

**IV. Data Reuse.** Scientists interested in ecological stoichiometry, biological models, and ocean biogeochemistry will be interested in the data. We will maintain a link on our institutional website to the location of the data on BCO-DMO as well as indicate this in our publications. Data will be described in accordance with developing BCO-DMO standards. As done in earlier projects, the investigators will work closely with BCO-DMO curators to ensure accurate and complete documentation in accordance with the BCO-DMO designated level of service, if appropriate. To facilitate tracking of reuse and fair credit to data providers, BCO-DMO will provide a recommended formal citation for the data set, including a persistent identifier and the contact person's last name.

**V. Data Preservation.** During the laboratory work, a collective sample log of samples collected and allocated for each measurement type will be kept in spreadsheet form to build the first stage of a data archive. The event log and the analytical data collected during the lab work will be stored on multiple physical disks locally. This plan for backups will permit restoration in the event of a hard disk failure, fire or other incident, which might affect multiple computers in one physical location. After the completion of the lab work, a metadata file and discrete data files will be submitted to BCO-DMO. Model will be submitted GitHub.