

## **Data Management Plan**

### **I. Title and contact information**

1. *Project title:* Phosphonate Utilization by Eukaryotic Phytoplankton: Who, How, and Where?
2. *Points of contact:* M. Lomas, mlomas@bigelow.org; L. Whitney, lwhitney@bigelow.org;  
(207) 315-2567

### **II. Types of data**

#### **A. Laboratory Data description**

Processed cell counts, nutrient concentrations, alkaline phosphatase activity, particulate cell quotas, polyphosphate quantification, and methane measurements will be collated on Excel and CSV files by experiment/species. Raw text files containing chemical analysis, output files from instruments will also be available if needed. RNA-sequencing data from the laboratory culture experiments will include raw data files containing sequence reads and quality scores (fastq) and processed expression data (tab delimited and/or spreadsheet format). The expected size of collection is less than 1 TB.

### **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. RNA-sequencing data will be deposited to the NCBI Gene Expression Omnibus database. All genomic data will be listed on the data inventory at BCO-DMO with a unique accession number to link to the experimental data.

#### **B. Metadata**

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.

#### **C. Data organization**

Preplanning will be coordinated between the PIs and a sample plan for the laboratory experiments will be drafted and finalized to ensure that all necessary types of data and metadata will be collected to achieve the project goals and that the sampling strategy to obtain the best possible sample coverage will be used. Data will be stored on a desktop computer and backed up daily to a high performance compute cluster at Bigelow Laboratory. All file names include DDMMYYYY and are renamed as they are re-saved.

#### **D. Data quality**

Routine instrument calibration for measurement of all biochemical parameters is conducted using appropriate calibration standard 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 PIs, who have established robust institutional data management policies. The designated person of responsibility is Michael Lomas, mlomas@bigelow.org, and/or LeAnn Whitney, lwhitney@bigelow.org, (207) 315-2567.

### **III. Data access and sharing**

The data will be of exclusive use by the PIs until publication or one year after the end of the grant; data files and collection-level metadata will then be uploaded to BCO-DMO. BCO-DMO will make the data publicly available through their system immediately upon receipt and subject to any approved embargo period. 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**

The data will be of interest to modelers attempting to model nutrient cycling in a range of ocean basins. 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. The investigators will work closely with BCO-DMO data managers 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.