**Data Management Plan– Investigating Vitamin B1 Limitation and Advantageous use of B1-related compounds by marine bacterioplankton (PI- Paerl)**

**Data Policy Compliance**- PI Paerl, lab members and co-investigators will comply with the data management and dissemination policies described in the NSF Award and Administration Guide (AAG, Chapter VI.D.4) and the NSF Division of Ocean Sciences Sample and Data Policy.

**Description of Data Types -** The project will produce several observational and experimental datasets, described in the list below. In addition to the datasets described below, educational resources produced by the project, including data and images, will be made available for public use on the http://www.cosee-se.org/ COSEE SouthEast website, as well as, http://www.cosee.net.

**Observational Datasets:**
1. **Physiochemical and biological measurements:** Measurements of temperature, salinity, conductivity, chl a, dissolved oxygen, pH, nutrients, and fluorescence of surface water. The methods are those used by the local ongoing monitoring program MODMON http://www.unc.edu/ims/neuse/modmon/water_quality.htm. A YSI Sensor Sonde will be used to measure these parameters in surface water from station OB10 used for experimental incubations and filtrations. Raw data will be processed and converted to easy-access datatable files (.csv, .xls). Repository: BCO-DMO
2. **Event log:** Cruise scientific sampling event log. This will include event numbers, start/end dates, times & locations of water collections and physiochemical and biological measurements. Will be recorded on paper log sheets then entered into Excel files. File types: Excel file converted to .csv; scanned PDFs. Repository: BCO-DMO.

**Experimental Datasets:**
1. **Genetic sequencing:** DNA and RNA sequences will be obtained from field collected and experimentally incubated plankton. Plankton biomass will be collected at coastal UNC-IMS and NCSU-CMAST laboratories, as well as, the NCSU main campus (Jordan Hall) at different incubation time points. Extraction and sequencing prep will be performed in labs of PI Paerl and Gifford in Raleigh and Chapel Hill, NC. Data from experiments will be noted as such when deposited. File types: Short-read archive (.sra) and .fasta files. Repository: NCBI/Genbank; accession numbers to be provided to BCO-DMO.
2. **Proteomics:** Peptide sequences from bacterioplankton biomass; Datasets will consist of experimental metadata and amino acid sequences. File types: .raw . Repository: Proteome Xchange; Public repository Figshare (http://figshare.com).
3. **Cell abundances and growth (including bacterial production) rates**: Measurements of cell abundance, growth in bottle incubations of natural suface water, but also isolate cultures/chemostats; Flow cytometry data will be in .fcs format – example gates will be provided as image files, while enumeration data will be provided in common table formats (.xlsx or .csv). Dataset will include labeling

of treatments and time points.  File types: .csv, .xls, .jpeg (where applicable).
Repository: Figshare.

**Data and Metadata Formats and Standards-** Field observation data will be stored as
.csv or .xlsx files, common tabular formats. Field data will include date, time, latitude,
longitude, cast number, and depth, as appropriate. Metadata will be prepared in
accordance with BCO-DMO conventions (i.e. using the BCO-DMO metadata forms) and
will include detailed descriptions of collection and analysis procedures.

**Data Storage and Access During the Project**- Paerl lab members will store project data
(including spreadsheets, ASCII files, images, and PDFs of scanned logs) on laboratory
computers that are backed up by the University's central IT organization and/or an onsite
external harddrive (done weekly to biweekly). PI Paerl will generate a 'team' google
drive folder for data storage and sharing amongst all project participants.

**Mechanisms and Policies for Access, Sharing, Re-Use, and Re-Distribution**- All
cruise paper logs will be photocopied and/or converted to electronic electronic
copies.  Physiological and biological data from field samplings will be compiled and
made available through the BCO-DMO data system within two-years from the date of
collection.  DNA and mRNA sequences will be deposited in the National Center for
Biotechnology Information (NCBI) database GenBank upon submission of
manuscripts.  Proteomic data will be made available through Proteome Xchange upon
submission of manuscripts.  GenBank  and Proteome Xchange accession numbers will be
provided to the Biological and Chemical Oceanography Data Management Office (BCO-
DMO) in an Excel spreadsheet or .CSV file and metadata will be provided using the
BCO-DMO Dataset Metadata submission form. The project investigators will work with
BCO-DMO data managers to make project data available online in compliance with the
NSF OCE Sample and Data Policy. Data, samples, and other information collected under
this project can be made publically available without restriction once submitted to the
public repositories.  Data produced by this project may be of interest to chemical and
biological oceanographers, also microbiologists interested in physiological responses
related to thiamine by bacteria. We will adhere to and promote the standards, policies,
and provisions for data and metadata submission, access, re-use, distribution, and
ownership as prescribed by the BCO-DMO Terms of Use (http://www.bco-
dmo.org/terms-use).

**Archiving Plans**- Data will be made publically available through BCO-DMO or
appropriate public (and national) data archives. As mentioned above, NCBI/Genbank and
Proteome Xchanges Genetic will be used to archive genetic and proteomic data.  These
data will be deposited with Dr. E. Bertrand, a key collaborator on the project.   Figshare
will be used to share all other relevant experimental data.  The PI will work with BCO-
DMO to ensure data are archived appropriately and that proper and complete
documentation are archived along with the data.