Data Management Plan

1. Types of data

Immediately upon funding, we will register the program with BCO-DMO. Two distinct data streams will be utilized in this project. On HOT cruises, we will tie our measurements to the ongoing time-series collections. On the independent research cruise, we will collect our own CTD, nutrient, and dissolved/particulate data as well. On both the HOT and our cruise, we will also be collecting high-resolution plankton imagery and environmental data via the dVPR and CPICS system. Project specific bottle collections will be TEP, CSP, and carbohydrate samples from the HOT cruises with the addition of nutrients, DON, DOC, O_2, and chl a for the independent reseach cruise. All data including ship's position and time, and will be recorded, stored, and backed-up in digital formats on board the ship. Data storage will be formatted for upload to BCO-DMO and also backed up for laboratory storage at both UTMSI and UNH. Standard HOT data will be uploaded as per their published data processing procedures.

Both the CPICS and dVPR imagery data are taken at a high scan rate and stored as digital recordings. The scanned images are then broken up into individual regions of interest (ROIs). The DICE image recognition and classification program can be used for both data streams and after creating the ROIs will then be used for classification/identification into targeted categories (e.g., *Ethmodiscus*, *Halosphaera*, *Rhizosolenia* mats, and others as needed). Both the ROIs and original digital images will be uploaded to BCO-DMO.

Model output will come from the three 1-D models detailed in the proposed activities: the VMF growth and migration model, the organic matter cycling model, and the preformed NO_3 and DIC model. Chemical analysis will consist of dissolved nitrate + nitrite, dissolved phosphate, dissolved silicate, dissolved oxygen, TEP, CSP, total carbohydrate, and particulate chlorophyll concentrations measured on CTD-collected seawater samples.

2. Data and Metadata Standards

The dVPR collects concurrent conductivity, temperature, depth, and fluorometry. These data will be available through the DICE user interface and are available for export to .cvs or .xls formats. These profiles will be uploaded with metadata to identify position and date/time. Individual CPICs ROIs are extracted and stamped with CTD and time data embedded in the file. The images derived from the dVPR will be stored in a standard PC file system format and uploaded to BCO-DMO as station specific files. Individual dVPR ROIs are extracted and CTD/time stamped in a similar fashion. CTD calibration sheets will be uploaded to BCO-DMO as metadata files with the final data sets. Nutrient data and chlorophyll collected by these PIs will be calibrated to standards from commercial sources. TEP, CSP, and carbohydrate data will be standardized to prepared standards as per the protocols listed in the relevant literature. Concentrations from bottle sample data will be stored in comma separated values, '.csv', format. Files will contain data and metadata related to the CTD sampling location and time, depths sampled, and chemical concentration data. Both input and output files of the 1-D models will be stored in the native data format of MATLAB® software, .mat, which is a matrix data format that stores matrices, values, or strings in a binary format. Metadata for these files will consist of user-generated 'readme' files describing the contents of each stored matrix.

3. Policies for access and sharing and provisions for appropriate protection/privacy

The PI will approve access to the data until such time as the data are to be presented to the public (see below). In the near term, the PIs retain the right to use the data before access is provided to the larger scientific community or the public. The data will be made public within a 2-yr time period following completion of the funded project. There are no ethical or privacy concerns. All intellectual property rights are reserved as per standard University policies at the PI's institutions.

Bottle collected data will be made available to the public after all sensor calibrations are completed and publication of results. Expected timeline is within a maximum of 2 year following the project end date. HOT data (TEP, CSP, carbohydrates) will be made available following their standard quality assurance procedures. Model output will be available immediately following the release of relevant publications upon request via email (robert.letscher@unh.edu) or the PI's website.

4. Policies and provisions for re-use, re-distribution

We are planning for permanent sharing and re-use of these data. It is likely that many researchers within the marine biology, biological, and chemical oceanography fields could utilize these data. DICE is a proprietary image extraction/analysis system; however, other processing software such as the free Visual Plankton is available for analysis of the original files. Original CTD/bottle data may be requested and re-plotted, or downloaded directly from BCO-DMO for re-analysis. No policies are imposed for reuse or production of derivatives from the model output other than citation of original publications.

5. Plans for archiving and Preservation of access

All data with its associated metadata will include specifics on data collection as well as instrument details (brand/model, calibration, resolution, accuracy) and uploaded to BCO-DMO. Archiving of dVPR and CPICS data will require discussion with BCO-DMO to determine how they wish to preserve the data and the proper format for the imagery. Once complete and quality-controlled, these data will also be archived in open-access national databases (BCO-DMO).

The primary archive for model results will be within published papers in peer-reviewed journals. Model code and output will be archived at the PI's institution (University of New Hampshire) as well as the model code with GitHub.