<u>Data Management Plan</u>

Title: Collaborative Research: Investigating the relationship between size and the balance between carbon acquisition modes in mixotrophic protists

Data Policy Compliance
The project 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, and to all institutional guidelines for data management and dissemination.

Description of Data Types
<u>Experimental Laboratory Data</u>: The primary experimental data products for this project will include flow cytometry derived bacterial and mixotroph cell abundances, and bulk ^{13}C measurements (POC, DIC, and DOC). Data collected from all laboratory experiments (growth limiting experiments and ^{13}C experiments) will be published in peer-reviewed journals and raw data will be available upon request. Laboratory samples will be stored in amber bottles with Lugol's iodine solution, in cryovials with buffered paraformaldehyde at -80 °C, in glass or borosilicate amber EPA vials with septa caps at 4 °C, and all GF/F filters will be dried and stored at -80 °C.
<u>Images and Morphometric Data:</u> Cell images from the laboratory experiments will be obtained with imaging cytometry and microscopy. Morphometric data including basic dimensions (length width, etc), as well as cell surface area and biovolume, will be derived from these images. Morphometric data for each species analyzed will be published in peer-reviewed journals. The raw data and cell images will be available upon request.

The data generated by this work does not include any information on human subjects, socially controversial topics, or financial matters, and thus there are no personal privacy, confidentiality, security, or property rights concerns.

Data and Metadata Formats and Standards
<u>Data Quality</u>: The laboratory experiments will be quality controlled by running triplicates of every treatment and checking the standard deviation of the average. We will set-up and rerun specific experimental treatments that produce outliers or otherwise questionable values to ensure it is not an artificial result. The flow and imaging cytometry instruments will be quality controlled (QC'd) using bead standards (for both fluorescence and size) prior to daily use.
<u>Data Documentation</u>: 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. Cell count data will be entered and stored in spreadsheets (Microsoft Excel).

Data Storage and Access During the Project
Notebook and paper data sheets will be scanned and saved as pdfs and transcribed to spreadsheets or databases and these electronic files will be archived and backed up by automated institution storage. Initial electronic archiving of these data will be on the PIs desktop computers and backed up to a separate network storage device. These data will be made available to all project personnel. Data exchange between labs will take place using DropBox and Google Drive.

The computer code and analysis workflow for cell morphometric data will be made publicly available on Github as it is finalized. The code will be updated and annotated so that the latest version is always available.

Bigelow Laboratory Local Storage: This project is anticipated to consist of approximately 0.5 TB of data, including spreadsheets of laboratory experiment results, cytometry image files, written protocols, and image analysis code. All will be stored by the end of the project in the Bigelow Laboratory data warehouse. This warehouse consists of over 200TB of high performance, available storage. Bigelow Laboratory has a robust, multi-tiered data backup and retention strategy in place. Local data is stored on a scalable enterprise NetApp system that utilizes redundant power from uninterruptible power supplies with generator backup, dual high availability (HA) controllers, and redundant networking to ensure uptime. RAID DP allows for two disk failures per raid group without losing data, providing time to replace the disks. Hourly read-only snapshots provide the ability to rapidly recover from a ransomware attack or accidentally deleted files. A second similar system is co-located at an off-site data center for full redundancy in case of a physical disaster affecting the local system. Volumes are mirrored to the backup location daily.

Mechanisms and Policies for Access, Sharing, Re-Use, and Re-Distribution
Data provided to BCO-DMO will be in an Excel spreadsheet or .CSV file. Files of the sample metadata, cell count and environmental data for each sample date will be deposited using the BCO-DMO Dataset Metadata submission form. Data sets produced by the science party will be made available through these repositories immediately after publication or within two-years from the date of collection, which ever comes first. Data, samples, and other information collected under this project can be made publically available without restriction once submitted to the public repositories.

Plans for Archiving
The PIs will work with BCO-DMO to ensure data are archived appropriately and that proper and complete documentation are archived along with the data. In addition to the specified data repositories, raw and processed information will be managed and maintained at all three institutions, as both hardcopy and electronic files (Excel or other database program). Data will be retained throughout the careers of the Principal Investigators.

Roles and Responsibilities
Each PI will be responsible for sharing her subset of data among the project participants in a timely fashion. Millette will take responsibility for coordinating data management, with the assistance of the co-PIs, Bigelow postdoctoral researcher, and VIMS Ph.D. student. All PIs will work with the Bigelow postdoc and VIMS student, and include them in the data management process as part of their training.