DATA MANGEMENT PLAN

Data types: The project will generate multiple types of data from discrete samples collected during the incubation experiments as well as samples collected during the cruise (Table 1). During the cruise, we will use paper logs to keep track of the station locations, sample collections, sampling times, and descriptions of the samples. If an E-log system is available on the ship, we will use it as our primary logging system.

For the laboratory experiments and the field experiments, we will follow the best management practices for metadata and data outlined by the Biological and Chemical Oceanography Data Management Office (BCO-DMO). We will work with the BCO-DMO staff to manage the data, and the data generated by the project will be contributed to the BCO-DMO system. All data submitted to a public repository will be accompanied by as much metadata as possible, meeting the appropriate MIMARKS / MIxS / MIFlowCyt standards (and exceeding minimum requirements where possible).

Table 1. Description of the types of data to be generated during the present project.

Type of samples	Brief description	Expected file format	Expected file size	Repository
Discrete samples	Discrete samples will be collected and processed to obtain concentrations of the inorganic (nitrate + nitrite, nitrite, and phosphate) and organic (dissolved/total organic carbon, total dissolved nitrogen, dissolved organic nitrogen) components of each water sample. Particulate organic carbon samples will be analyzed from the incubation experiments. Bacterial respiration rate data will be collected from heterotroph incubations.	.tsv	100 Kb	BCO-DMO
Microbial diversity	16S/18S rRNA amplicon sequencing will be conducted on the natural microbial communities during incubations with vesicles, other carbon sources, or controls.	.fastq	1-5 Gb	NCBI SRA; BCO-DMO
Gene expression	Culture transcriptome (RNAseq) data	.fastq, .tsv	10-15 Gb	NCBI SRA, GEO; BCO-DMO
Flow Cytometry	Cell population measurements during growth on different carbon sources	.fcs	~0.5 Gb	FlowRepository

Data availability and archives: All data will be available upon request, and made fully publicly available upon publication. No restrictions will be placed on the use or reuse of the datasets, as long as standard practices to properly cite the original source of the data are followed. All electronic data (both primary data files and derived work files) generated will be stored on multiple computer hard drives or portable hard drives, and will be further backed up to cloud-based backup services provided courtesy of our institutions. All work products in the lab will be recorded in lab notebooks, which will be stored and maintained within the lab as per standard practices. For computational analyses, records of all processing steps and information about file contents will be maintained either in a physical or virtual lab notebook.

Laboratory notebooks containing primary data will remain in the laboratories of the project's lead investigators. Immediately following the cruise, the underway data collected from the ship's flow through system will be contributed by the vessel operator to the UNOLS central data repository at http://www.rvdata.us/catalog/ which is managed by the Rolling Deck to Repository (R2R) project. R2R will ensure that the original underway measurements are archived permanently at National Oceanographic Data Center (NODC). The remaining measurements will be managed by the Biological and Chemical Oceanography Data Management Office (BCO-DMO) and the data sets will be available online from the BCO-DMO data system (http://bco-dmo.org/data/). BCO-DMO will also archive all the data they manage at NODC.

Molecular sequence data will be shared in standard INSDC repositories and made available within 2 years of generation or upon manuscript publication. Original and unprocessed metagenome and transcriptome sequence reads will be submitted to the NCBI Sequence Read Archive (https://www.ncbi.nlm.nih.gov/sra), and processed transcriptome data from cultures will be deposited in the Gene Expression Omnibus database (https://www.ncbi.nlm.nih.gov/geo/). Any computational scripts, pipelines or tools developed as part of this project will be placed into a public archive at Github (http://github.com) along with appropriate documentation. Any relevant flow cytometry data files will be shared in FlowRepository (http://flowrepository.org), DataDryad (https://datadryad.org/), or FigShare (https://figshare.com). Sequence data and flow cytometry data will be linked to the associated BCO-DMO record for the project.

The final accepted version of all peer-reviewed publications from this project will be submitted, as required, into NSF's Public Access Repository.

Distribution of cultures, DNA, and other information: We donate our cultures to the NCMA at the Bigelow Labs (https://ncma.bigelow.org/) for distribution, and also supply them directly to researchers upon request. We also supply DNA from our cultures upon request.

Educational materials: Curricula, lesson plans and exercises developed around the analysis and interpretation of these datasets by co-PI Biller will be made publicly available on Wellesley College's website.