Data Management Plan

Data and Sample Types

We will generate bulk dissolved organic nutrient and inorganic nutrient measurements, mass spectrometry data, fluorometry scan data, cell abundance data, microscopy images (FISH), amplicon gene sequences and metagenomics sequence data from field sampling and field incubation experiments. Tab delimited files (MS Excel) will be used for data storage, which can be imported into a variety of statistical programs and data storage programs or facilities. Image files will be stored as both program specific files (to the imaging software program) and as more accessible files including PDF and TIFF. Sequence files will be stored as FastA files, accessible by any text editing program.

Laboratory notebooks will be archived in the PI's laboratory and available for review by all collaborators and interested scientific entities.

Standards

Datasets will comply with standards such as Minimal Information about a Metagenomic Sequence (MIMS) and Minimal Information about a Marker Sequence (MIMARKS), which provide a curated standard format layer for the acquisition and display of information associated with sample acquisition, processing, handling, sequencing, and analysis. These are community standards, agreed using consensus and updated where necessary by routine annual meetings of the GSC (www.gensc.org).

Metabolomics data (LC-MS) will follow guidelines by the Metabolomics Standards Initiative (MSI), including Core Information for Metabolomics Reporting, which provides guidelines for reporting metabolomics experiments, and specifically for data from an Environmental Analysis Context.

For all data, including images, scan data from the scanning fluorometer, and nutrient data, the associated metadata including sample location, time, depth, and ancillary experimental data (e.g., treatment, time, temperature, etc.) will also be provided when these data are made available.

Data Dissemination

All field data will be made available through the biological and chemical oceanography data management online (BCO-DMO) prior to publication or within two years of collection. Sequence data generated in this project will be deposited in NCBI's Sequence Read Archive (SRA) for next generation sequencing projects (www.ncbi.nlm.nih.gov/sra). All policies for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, and other rights or requirements will be adhered to where necessary.

Fluorometry data (EEMs of PARAFAC components) will also be deposited to the OpenFluor database (http://www.openfluor.org/). Metabolomics data will be deposited in the MetaboLights database within the European Bioinformatics Institute (EMBL-EBI; https://www.ebi.ac.uk/metabolights/).

We will attempt to publish our data in peer reviewed international scientific journals in a timely manner following the proposed timeframe in the project description, making all data publicly available at least by the time of manuscript submission.

Data Archiving

Whenever possible, public data posting will use persistent URLs linking to the archival location (BCO-DMO, NCBI, etc) and appropriate accession numbers. Furthermore, announcements that completed metabolomics and gene sequence data relevant to (marine, sponge, coral reef, microbiology) research communities will be made through available research list servers such as the sponge and coral reef list serve (http://coral.aoml.noaa.gov/mailman/listinfo/coral-list/), professional conferences or data links from the website generated for this project.

Short-term storage, or back-up, of data will be performed over the duration of the grant using computational resources (computer and external drives) in the PI's laboratory. Long-term storage, or archival, will preserve data as long as is necessary to satisfy the legal and moral responsibilities at Appalachian State University and Woods Hole Oceanographic Institution.