

## **Data Management Plan**

Data management and sharing are essential to the goals of the proposed research. The PIs are committed to working with the Biological and Chemical Oceanography Data Management Office (BCO-DMO), an NSF-funded data repository, archiving and making all data sources publically available. The PIs have corresponded with BCO-DMO about the anticipated project data types, and are assured that they will be able to assist with archiving and storing data generated from this proposed project. During all years of this project, the PIs have budgeted salary time for organizing, managing, and publishing data from this project.

This project will produce large amounts of data in the form of: (i) data and photographs from field and laboratory experiments, (ii), genomic amplicon sequences, raw and processed and (iii) peer-reviewed publications.

### **Raw data management, archival and storage**

Data collection and analysis processes as well as contextual details will be documented in individuals' laboratory notebooks with daily photo back-ups of these pages. All field and experimental metadata will be electronically recorded and managed by data type using the software Microsoft Excel, stored on laptop computers and backed up daily using external hard-drives (in the field) and to a remote locations at the Arkansas State University (A-State) and at the Woods Hole Oceanographic Institution (WHOI).

The raw and processed next-generation sequencing data (amplicon genes) will be stored on a server located in Apprill's laboratory, which is backed-up nightly. Raw amplicon sequence data will be submitted to the National Center for Biotechnology Information (NCBI) Sequence Read Archive depository (<http://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi>). Processed amplicon data, will be submitted to the Marine Biological Laboratory's Visualization and Analysis of Microbial Population Structures (VAMPS) database (<https://vamaps2.mbl.edu>), a visualization and sharing platform for microbial amplicon data. This submission will include metadata about the sequences that conforms to MIMARKS, the Genomic Standards Consortium (<http://gensc.org>). Accession numbers for all sequences will be made available in respective publications.

### **BCO-DMO archival and integration of project data**

The BCO-DMO project website will serve as the primary link to all project data. All project metadata will be submitted to BCO-DMO under the award number of this project. PIs Sikkel and Apprill will work with BCO-DMO to archive, integrate, and link the genomic repositories to the project data and make them available for use.

### **Publications**

Project results will be published in open access, peer reviewed publications with links available to the data (NCBI or VAMPS).

### **Data use policies**

There will be no permission restrictions placed on the data, nor any embargo periods for political, commercial or patent purposes. Data collected from this project will be made readily available to the scientific community through several data servers and published manuscripts in open access peer-reviewed journals.