

Data Management Plan

1. Types of data

Data for this project will include physical collections for biological and chemical samples. Specifically, physical samples will include concentrated water samples and extracted total community DNA and cultured bacterial isolates as well as filter-captured aerosol particulates. Hundreds of samples will be collected (bulk water) and over 1,000 bacterial isolates are expected to be obtained. Physical samples will be stored in a redundant system to ensure that samples are not lost in the case of power outages or other unforeseen failures.

Other data produced will include nutrient concentrations, trace metal concentrations, particulate size class and concentrations, chlorophyll *a* concentrations as well as other parameters measured *in situ* including pH, temperature, salinity, dissolved oxygen and turbidity. Satellite images (MODIS) will also be obtained for tracking dust transport. Raw fluorescence curves will be generated for each qPCR assay and digital images of any electrophoretic gels for conventional PCR will be stored. Finally, several thousand genetic sequences will be produced through 454 pyrosequencing.

2. Data and Metadata Standards

Metadata will be recorded during every sample collection including date and time of collection, tide, weather conditions, wind speed and direction and wave height, and for any notable conditions experienced. Among the bacterial collection, metadata on colony morphology and growth characteristics will also be noted.

Experimental and field data and metadata will be recorded into appropriate laboratory notebooks at the time of collection/analysis (i.e., waterproof pages for field work) with duplicate carbon copy pages. All recorded measurements will be transferred to Excel spreadsheets or databases within one week of collection. Established data exploratory exercises will be performed on databases to scan for errors, missing data, or outliers. Only after rigorous quality assurances of databases have been performed will they be entered into a metadata archive for official analyses. Full methodology for each procedure will be documented in database metadata and in refereed publications.

Master copies of all notebooks will be maintained in the offices of the respective PI or co-PI and will be kept for at least 5 years after the end of the project. Additionally, all entered data or digital images will be backed up daily onto University servers under password protection. Data will also be backed up onto CDs and external hard-drives at least weekly.

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

The specified embargo period associated with these data extends from the projected beginning date until six months after project conclusion date. Upon cessation of the embargo, all experimental data and metadata will be submitted to the Biological and Chemical Oceanography Data Management Office (BCO-DMO) for assistance with data stewardship, dissemination and storage; and, will be made available upon request. We expect collections of ocean physical, chemical and biological data and associated metadata will be submitted to the National Oceanographic Data Center. All aspects of the data management plan will be coordinated with the BCO-DMO.

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

Data and representative physical samples will be made available within a reasonable time frame. Data sharing will be coordinated with the BCO-DMO. Any bacterial isolates will be maintained in Dr. Lipp's culture collection and will be made available to interested researchers in the scientific community. Limited quantities of concentrated water samples can be made available; although the small amounts collected will limit the ability to share samples widely. Any genetic sequences obtained for the cultures obtained during this study will be uploaded into the NCBI's GenBank system; all accession numbers will be reported in any published work.

Through the course of this project period, and beyond, we will publish our results in appropriate peer-reviewed journals and scholarly texts, using open access formats whenever feasible. Results will also be reported broadly through presentations at local, state, regional, national and international venues, including meetings of stakeholder groups as well as scientific societies.

5. Plans for archiving and Preservation of access

Physical samples will be curated and archived in the respective laboratories of the PIs. Detailed spreadsheets (or databases) will be maintained to note sample collection date, storage date, condition and location. Periodically, spreadsheets will be checked for accuracy against the physically stored samples and updated as required. Physical samples will be stored in a redundant system to ensure that samples are not lost in the case of power outages or other unforeseen failures (funds are requested to offset the cost of an additional ultra-low temperature freezer). Electronic data produced in this study will also be stored in a redundant manner on both external hard drives and university servers, as described above. Data files will be password protected and only accessible to project personnel. Data will be archived for at least 5 years beyond the final publication for this project.