Data Management Plan

1 **Roles and responsibilities**
All members of our research team will be responsible for preserving data quality through standardized field and lab practices. Gonsior will supervise data quality measures, file naming, data documentation and metadata creation and will have the oversight of the data management plan. Gonsior will also be the point of contact for long term archiving that is currently under development at UMCES. The post-doctoral scholar will work closely with undergraduate and REU students to maintain the standardized field and lab practices. Open access data repositories will facilitate the data entry from remote locations.

2 **Expected data**
This study will generate physico-chemical data that can be relatively easily handled, but it will also generate highly complex spectrometric and spectroscopic data that has to undergo drastic data reduction to be useful for general use. These data reduction procedures are partially in place but need to be refined throughout this study to make data accessible and useful for non-experts in mass spectrometry and optical spectroscopy. Quality assurance and quality controls are integral to all method validations and include instrument performance checks, a variety of blanks, replicate analysis and cross-validation. Many of our analyses are performed with internal standards and all data will undergo manual and semi-automated proofing and quality control measures.

3 **Period of data retention**
All data (processed and raw) will be kept indefinitely in currently developed repositories at UMCES. The exact layout of storing raw spectral data is currently under development and will be streamlined in the near future. All processed data will be made available online and all raw data will be available on request toward the end of the proposed project.

4 **Data formats and metadata**
Data reduction and analysis will occur primarily in spreadsheets and table (e.g., MS Excel, relational databases (e.g., MS Access) and statistical programs (e.g., R, PRIMER-E). The programs we use are capable of creating comma-delimited ASCII flat files which will be the principal data format used for long term storage and archiving. Comma-delimited ASCII flat files are compact, designed for simplicity, durability, accessibility and interoperability with the vast majority of data analysis/processing programs past and present.
Spectral metadata is typically stored in software specific files and pose a challenge to long-term storage. We anticipate designing meaningful export routines to also store spectral metadata in ASCII flat file format or similar simplistic file formats.

5 **Data dissemination and policies for public access, sharing and publication delays**
Physico-chemical data will be made instantly available through open-access repositories. Spectral data will first be made available through peer-reviewed publications and will then be stored in open-access databases where inquiries can be used to filter the data of interest. All data will be freely accessible at the end of the proposed study.

6 **Data storage and preservation of access**
Current practices at CBL are to back up all raw data and processed data on site and at a second location. A digital repository for scientific data including both near-line and archival storage is currently being expanded and a greater UMCES repository of mirrored sites among the different facilities is being developed. The raw data for this project will be archived and external public access will be granted. IT will provide a centralized preservation, storage and access

environment for the digitized data of current and future research activities.