

Data Management Plan

Products of the Research: A series of ethanol and acetaldehyde concentration measurements and rate measurements will be made across a range of coastal waters and seasons using a purge and trap GC/MS system. Concentrations will be calculated from the peak areas on chromatograms and concentrations of ethanol and acetaldehyde will be plotted as a function of time to determine rates. Rate constants will be calculated by fitting the concentration vs. time data with first order and second order kinetic fits using Origin software. Concentrations of ethanol and acetaldehyde in air will also be measured using a combination of HPLC and GC/MS systems. Once again, concentrations will be determined from chromatogram peak areas. We will also generate standard water quality measurements (temperature, salinity etc.) and measurements of bacterial counts using fluorescence microscopy on filtered samples and RNA gene sequencing of microbial populations on material extracted from filters. Water samples will not be permanently archived as they are not stable over lengthy time periods due to changing bacterial populations. Filters for microbial measurements will not be archived after processing due to changes associated with processing methods.

Data Format: Data files will contain a sample identifier and typical data associated with a chromatography file (i.e. peak area and retention times). Chromatography files will be named systematically to indicate point of origin, time sampled and data treatment if any (eg. filtering, sodium azide etc). Ancillary data will be treated the same way. Data will be analyzed as soon as collected to monitor data quality and preemptively identify any issues with instrument or sample anomalies that would require repeating the experiments. Standard statistical tests will be used to assess data quality eg. student's t-test, p-test. Data from field sites will be collected in Excel spreadsheets with sufficiently detailed column headings to be descriptive and plotted in Origin software, with figures exported as wmf files for manuscript preparation. Origin files are saved in an opj format that can be read by the software program.

Dissemination of Results: The proposed research will be published in peer-reviewed journals, including undergraduate students as co-authors and acknowledging support from the NSF. Publications will include the concentration, time, bacterial count data and RNA sequencing as appropriate. Supporting material (such as peak areas on chromatograms) will be made available as supplementary material to the papers when published. The PIs have published 8 papers with undergraduate student co-authors in the last 5 years in Tier 1 and Tier 2 peer-reviewed journals. Representative figures will be shown in the papers and extensive data tables provided. We will aim to publish results within 2 years of collection and completion of data analyses per NSF data management guidelines. We will be happy to share our data with anyone requesting these. We will post links to published papers (pending copyright) on our web-site. The results of the proposed work will also be presented at regional and national scientific meetings. In recent years, the PIs and undergraduate students have attended and presented at regional and national meetings of the American Chemical Society and the national Ocean Sciences meetings.

Access to Data and Data Sharing Practices and Policies: Per the sample and data policy of the Division of Ocean Sciences of the NSF, data will be uploaded to the Biological and Chemical Oceanography Data Management Office (BCO-DMO). Once the project is funded, the PIs will submit a project metadata form to register the project with BCO-DMO. Data will be uploaded to the database in excel spreadsheets with dataset metadata forms once calculations have been completed and the data quality checked. Per NSF data management policy, data will be available within 2 years of collection as a maximum time frame, and as quickly as processing allows for (ideally within 3 to 5 weeks). Archived data will include ethanol and acetaldehyde concentration vs. time data, bacterial counts and ancillary measurements like salinity and dissolved organic carbon. The RNA gene sequences from genetic analyses providing information on microbial population genetics will be uploaded to the National Center for Biotechnology Information (NCBI) database. Sequence accession numbers will be given in the BCO-DMO project database to link to the NCBI database. The interpretation and results of the proposed project will also be made accessible through peer-reviewed publications and the accompanying supplementary material. The PI's research websites will include web links to publications and the BCO-DMO and NCBI websites. If someone requests further information, the PI will communicate with them to determine what additional materials to send. Typically, the information needed to replicate the results in publications will be provided in sufficient detail in the database management systems and papers and supplementary material.

Policies for Re-Use, Re-Distribution, and Production of Derivatives: The PIs will respect the re-use and re-distribution policies of the journals in which the research products are published. Specific results from publications will not be presented on the PI's research website. Instead, web links will be provided on the PIs' web sites to publications. Publications from the project will also be given on the BCO-DMO database. Reuse will be subject to the policies of the journal websites.

Archiving of Data: Data files will be stored on the PC associated with the instrument and backed up weekly to an external hard drive which will be permanently archived in the PI's office. All experimental data will also initially be recorded in hard copy laboratory notebooks at the time of collection. These will be archived in the PIs' offices after the project is completed. Data will be processed and transferred to standard spreadsheet software which will be archived on a lab pc and backed up weekly to an external hard drive. Research-related Excel and Origin files for the project will also be saved on the PIs office desktop PCs and backed up to Dropbox monthly. Files on Clark's PC will be backed up weekly to a chemistry department server, and to Microsoft One Drive monthly. Locally archived files and lab notebooks will be available for at least 10 years after the date of publication of papers. The data will also be archived on the data management systems of BCO-DMO and the NCBI.