Data Management Plan

I. Types of data

Materials produced from this research will include cultivated microbial consortia, DNA and RNA sequences from various sites and water depths, and (meta)transcriptomic data. Experimental measurements of the concentrations of reactive oxygen species and dissolved Hg species will be obtained at several stations per cruise. Experimental data will consist of production and decay rates for ROS and Hg redox reactions in water incubations conducted under various environmental conditions. Microbial community analysis of the waters will be simultaneously investigated, resulting in a host of observational data, including abundance (cells/mL) of microbial groups (autotrophs, heterotrophs) from flow cytometry, microbial community composition via DNA pyrosequencing (~150-bp average length), and microbial gene expression inferred from mRNA transcripts of functional genes. DNA and cDNA sequences will be queried against the NCBI-nr and GOS peptide databases using BLASTX.

II. Data and metadata standards

The chemical data sets generated by this project will be small and generated by two groups. Data collection and analysis processes as well as contextual details (sampling site location, time, etc.) will be documented in these individuals' laboratory notebooks and in their publications. Raw data will be stored in the form of Microsoft Excel spreadsheets. All cruise details and chemical data will be submitted to the Biological and Chemical Oceanography Data Management Office (BCO-DMO). Metadata sequences will be submitted to NCBI according to NCBI's standards, which prescribe the necessary formatting and supplementary data to assist researchers in using the data, including the site location/description, predicted function, and associated publication.

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

The experimental data obtained in this research, including biological ROS production and Hg redox rates, and biological information (e.g., biomass of autotrophs, heterotrophs), will be made available to the scientific community and general public through publication in peer-reviewed journals. Any data of interest to environmental and consortia DNA and cDNA pyrosequences will be deposited in the NCBI, and accession numbers will be reported in subsequent publications. Further, microbial consortia will be maintained in the Hansel Lab microbial collection and available upon request from interested researchers. The availability of the cultures will be documented in subsequent publications of the microbial research conducted herein.

There will be no embargo periods for political/commercial/patent reasons. Chemical and biological data will be made available following collection and analysis. Data dissemination will be noted in the publications within the Materials and Methods section to inform the scientific community of the data availability and accessibility. All chemical, physical, and some biological (e.g., chlorophyll) data will be made available through the BCO-DMO. All DNA and RNA sequence data will be available through NCBI. These data pipelines are free of charge and open to the public. We will retain the right to hold data prior to publication only if a conflict of interest seems warranted.

The dissemination of the chemical and biological data to be collected for this proposed research will not be restricted by any ethical or privacy issues, copyright concerns or restrictive licenses. As discussed above, all the data collected will be made readily available to the scientific community through various datacenters, published manuscripts in peer-reviewed journals, and by request to the affiliated researchers.

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

There will be no permission restrictions placed on the data. The data that will be collected through the proposed research will be of interest to a broad spectrum of environmental scientists, including chemical and biological oceanographers, biogeochemists, microbial ecologists, and environmental chemists. Particularly, the metagenomic and transcriptomic data collected within this research will add to the growing meta-datasets of microbial community composition, activity, and abundance within ocean systems. This data will be available for researchers to include in their analysis of marine microbial ecology and biogeochemistry.

V. Plans for archiving and preservation of access

As discussed above, all sequence data will be deposited in publicly available databases (e.g, NCBI, BCO-DMO), which will include indefinite archiving of this data. All relevant data will also be published through peer-reviewed journals provided in a usable manner. All microbial cultures will undergo long-term storage and archiving within the Hansel lab and, if appropriate, the ATCC (for fully characterized organisms). These organisms will be made publicly available via the center websites and upon request from the Hansel laboratory. They will be stored indefinitely. Depositing these organisms and information in multiple locations will ensure their availability to future researchers. All known relevant information (e.g., site of collection, affiliated publications, geochemistry, etc.) will be submitted alongside the sequence data and microbial cultures. As one of our stations during the offshore cruise will be the same as that of the U.S. GEOTRACES program, we will archive in an appropriate manner our remainder water samples for researchers interested in re-analysis of samples from that region of the NW Atlantic. Availability will be advertised through the U.S. GEOTRACES website.