Data Management Plan

Types of Data

Data management and sharing are essential to the goals of the proposed research and the Broader Impacts of this proposal. Thus, the PIs are committed to working with the Biological and Chemical Oceanography Data Management Office (BCO-DMO), an NSF-funded data repository, to archive and make all data sources publically available. Time during years 2 and 3 will be devoted to organizing, managing and publishing data from this project.

The expected data and metadata that will be produced and retained are:

1) Geochemical data collected via field, shipboard and lab analysis, including the concentrations of various species of manganese, nitrogen, iron, and iodine.
2) Geochemical data collected *in situ* using remotely operated submersible sensors and instruments including microsensor packages (dissolved gases, pH, T) and superoxide (via SOLARIS).
3) Microbial group abundance within waters via flow cytometry.
4) Shipboard underway digital data generated by sensor systems permanently installed on the ships and routinely maintained by the operator (e.g., navigation, multibeam sonar, sub-bottom profiler)
5) CTD data
6) Other complementary data in conjunction with research carried out by our IOW collaborators, including P concentrations, and particulate and dissolved trace metal concentrations and speciation.

Other samples to be obtained could include microbial cultures to be cultivated within the Baltic Sea depending on the results of the lab incubations and time/personnel availability. If cultivation is conducted, all microbial cultures will undergo long-term storage and archiving within the Hansel lab. These organisms will be made publicly available upon request. All relevant metadata (e.g., site of collection, affiliated publications, geochemistry etc.) will be submitted alongside the sequence data and microbial cultures.

Policies for access, sharing and archiving

Data collection and analysis processes as well as contextual details (sampling site location, time, etc.) will be documented in individuals' laboratory notebooks and in their publications. All biogeochemical data will be recorded and managed by data type using Microsoft Excel. The raw and processed data from these analyses will be stored on computers. Daily back-ups on the cruises will occur using external hard-drives and at WHOI these will be backed up daily to the cloud and a remote location.

The PIs will work with the BCO-DMO to archive and integrate all data (geochemical data, flow cytometry data, and metadata) from this project. Data archived with BCO-DMO are publically available and easily searchable, including by sample type, date of collection, type of data, and project name. Data will be submitted to the BCO-DMO within 1 year following each cruise. Links and references to the data sources will be available in associated publications and presentations.

Further, PI Hansel has provided reference X-ray absorption (XAS) spectra to a number of synchrotrons for dissemination to other users. These spectra now constitute the majority of the reference standard databases for Fe and Mn at a number of synchrotron facilities. Any new reference spectra collected as part of this project will be disseminated throughout the synchrotron libraries and made freely available to other users.

Policies and provisions for re-use, re-distribution

There will be no embargo periods for political/commercial/patent reasons. Further, there will be no permission restrictions placed on the data. 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 biological data will be made available through the BCO-DMO. If microbial cultures are obtained, all 16S rRNA data for isolate identification will be made 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 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.