

## **Data Management Plan**

In accordance with the NSF policy on dissemination and sharing of research results, all data collected as part of the work funded through this proposal will be made available to the broader scientific community. The PIs are committed to producing and maintaining high-quality and accessible datasets that will enable future investigations to utilize these data.

### **Physical samples**

Sediment samples resulting from different experimental treatments will be generated through this work. These samples will be preserved in a freezer at  $-80^{\circ}\text{C}$  which is under the personal supervision of PI Malkin and equipped with a temperature sensor and an on-line communication system. Portions of these samples will be consumed during the proposed work, but the remaining sediment samples will be kept as a part of an ongoing reference archive at UMCES. Upon completion of this project, the physical samples will be available to the scientific community by request. All collaborators on this project are committed to preserving and providing access to the physical samples analyzed during this project for the benefit of the larger scientific community.

### **Molecular sequencing data: format and archiving**

Molecular data generated through this project will consist of 16S rRNA gene amplicon sequences, metagenomes, and metatranscriptomes. Metatranscriptomic and 16S rRNA gene sequence data will be collected and curated according to the Genomic Standards Consortium ([http://genc.org/gc\\_wiki/index.php/Main\\_Page](http://genc.org/gc_wiki/index.php/Main_Page)) Minimum Information about a (Meta)Genome Sequence (MIGS/MIMS) standards ([http://genc.org/gc\\_wiki/index.php/MIGS/MIMS](http://genc.org/gc_wiki/index.php/MIGS/MIMS)). These outline a standardized format for the minimum information required to accurately describe 16S rRNA gene and metagenomic data, including metadata, with the goal of facilitating inter-study comparisons and transparency.

In compliance with MIGS/MIMS, 16S rRNA gene sequences, metagenomic and metatranscriptomic data will be deposited at the Sequence Read Archive at the National center for Biotechnology Information (NCBI). Furthermore, metagenome and metatranscriptome sequence data will be annotated, stored and these assembled, curated data will be made available to the public via the Department of Energy's Argonne National Laboratory metagenomics analysis server, MG-RAST. This server currently hosts more than 259,095 metagenomes. We will provide NCBI accession numbers for uploaded sequences in all publications. All accompanying metadata will be uploaded to the National Science Foundation's Biological and Chemical Oceanography Data Management Office (BCO-DMO), in ASCII format. PI Malkin maintains responsibility for management, retention and submission of data to these resources throughout the duration of the project.

### **Data dissemination**

The main vehicle for dissemination of results will be through peer reviewed publications. Upon publication, sequencing data will be available through NCBI and annotated curated data will be available through MG-RAST, with metadata available through NSF's BCO-DMO portal. Data deposited to MG-RAST become available after 6 months, regardless of publication status. Supporting data arising from this work will include CARD-FISH images which will be provided in publications in summary format, in supplemental materials, and available from the PIs. Preparation of publications will commence during the project in accordance with the work plan. Data will be released publically 2 years after the end of the project.

### **Data storage and preservation of access**

Sequence data will initially be stored on servers in the PIs and Co-I's labs. All geochemical data data will be made available through peer-reviewed publications, and once published, these data will also be archived in BCO-DMO where the data will be freely available to the public. Data corresponding to a publication will be made available as supplementary material and will be available through the website of the relevant journal. As discussed above, sequence datasets will be made publically available through NCBI and MG-RAST. Description of the samples (metadata) that are available will be sent to BCO-DMO to disseminate information on sample availability.