Data Management Plan

Summary

The PI will comply with the data management and dissemination policies described in the NSF Award and Administration Guide (AAG, Chapter VI.D.4) and the NSF Division of Ocean Sciences sample and data policy. Results from all experiments will be submitted as comma-separated value files to the BCO-DMO data repository within two years of generation or upon manuscript submission, whichever is earlier. All metagenomes and metagenome-assembled metagenomes will be made publicly accessible within two years of generation or upon publication, whichever is earlier, through an established web-accessible data archive center (namely, NCBI GenBank). Educational data will be maintained locally as crypto-protected spreadsheets with recursive backups.

Cruise Participation

Pre-Cruise planning will be done via teleconferencing and email. Detailed plans for station locations, water sampling, water budgets etc. will be written up as a science implementation plan for the cruise in concert with the Chief Scientist (Ward). Actual station logs will be recorded on paper logs and scanned into PDF documents.

Types of Data

This project will generate spatially resolved growth rate and associated chemical time series data, including nitrate, nitrite, nitric oxide, nitrous oxide, oxygen, and pH. Data will be analyzed predominantly in Matlab, ImageJ, and Nikon Elements. All data will be stored locally at MIT on duplicate hard drives, and in the cloud with automated online backups from the primary data source on the Code42 CrashPlan PROe that has unlimited no-cost storage for MIT faculty. All data will be made available to the public through the BCO-DMO database archive.

The project will additionally generate metagenomes and metagenome-assembled genomes. The original fastq files generated by the sequencing runs at the MIT Bio-Micro Center will be stored on our in-house RAID server that maintains automated online backups.

Laboratory reports generated by students in class will be stored as PDF files with redacted personal identifying information (e.g., names). Summary data regarding performance, activities post-class and tracking of majors will be maintained as a crypto-protected Excel database.

Data Format

Comma-separated value spreadsheets of all quality-controlled measurements and associated metadata will be made available following established guidelines for oceanographic data sets (e.g. templates established by the Biological and Chemical Oceanography Data Management Office (BCO-DMO) based at Woods Hole Oceanographic Institution. The original fastq files generated by the sequencing runs will be submitted to GenBank at the National Center of Biotechnology Information (NCBI) as outlined in their data submission protocols. Individual metagenome-assembled genomes will be submitted to NCBI GenBank according to their established submission protocols.

Access to Data and Data Sharing Practices and Policies

After quality control and assurance, all chemical data will be made publicly available through archiving on BCO-DMO within two years of acquisition. Data sets will be published in peer-reviewed journals and included in supplemental materials as appropriate. Long-term access to raw data (e.g. raw reads, microsensor voltage values, specific microscopy images and movies) for other researchers will promptly be made available as requested. Genetic data will be publicly available through GenBank within two years of generation.

Educational data of student performance and student curriculum and career progress will be available as anonymized, group statistics in order to maintain confidentiality, and provided only upon reasonable request. Summarized educational statistics will be submitted as supplementary data alongside its publication.

Policies for Re-Use, Re-Distribution
Once released, the data will be open and freely available for use without copyright restrictions.

Archiving of Data
Processed, quality-controlled biogeochemistry results will be archived at a national database center within two years of acquisition (BCO-DMO). Genetic data, both raw reads and processed metagenomes, will be archived in GenBank. The original (raw and processed) data as well as any published papers will be fully archived with redundant off site back up of computer files at MIT. Original data log sheets and (electronic) laboratory notebooks will be maintained by the PI.