

Data Management Plan

The proposed work will generate the data related to optical properties of FDOM, molecular structure of DOM from FT-ICR MS, and high throughput sequencing data from metagenomics and metatranscriptomics. These data will come from cultures of cyanobacteria and incubation experiments. The chemical data will be deposited by Michael Gonsior, while the biological data (Next Generation Sequencing) will be managed by Feng Chen.

Chemical data

The Biological and Chemical Oceanography Data Management Office (BCO-DMO) serves PIs funded by the NSF OCE Biological and Chemical Oceanography Programs. Optical property and molecular signature data of DOM will be deposited in BCO-DMO data as the primary data management source. If funded, PI will contact BCO-DMO and discuss the format of data that will be acceptable to BCO-DMO.

Biological data

The metagenomics data will be deposited in GenBank under NCBI (National Center for Biotechnology Information). NCBI collects submissions of data for the world's largest public repository of biological and scientific information. GenBank is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences. GenBank comprises the DNA DataBank of Japan (DDBJ), the European Nucleotide Archive (ENA), and GenBank at NCBI. These three organizations exchange data on a daily basis. The most important source of new data for GenBank is direct submissions from scientists. The GenBank database is designed to provide and encourage access within the scientific community to the most up to date and comprehensive DNA sequence information. Therefore, NCBI places no restrictions on the use or distribution of the GenBank data.

Metotranscriptomic data will be deposited in GEO (Gene Expression Omnibus) under NCBI. GEO accepts all types of data: metadata, raw data and processed data. Data will be submitted using the GEOarchive spreadsheet format. GEO is a public functional genomics data repository supporting MIAME-compliant data submissions. GEO accepts next generation sequence data that examine quantitative gene expression, gene regulation, epigenomics or other aspects of functional genomics using methods such as RNA-seq, miRNA-seq, ChIP-seq, RIP-seq, HiC-seq, methyl-seq, etc.