

Data Management Plan

All data generated by this project will be made available to the scientific community via the World Wide Web.

Types of data. This project will collect water samples for extraction of DNA and RNA, as well as biogeochemical concentration and flux measurements and omics-type sequence datasets. Extracted nucleic acids will be stored in freezers at UGA for two years after the project ends or until the final manuscripts are published. We do not anticipate creating any long-term curated physical collections.

Data and Metadata Standards and Release. Environmental context data and metadata will be made available through a project web site and through DataONE, the USGS NBII Metadata Clearinghouse, or comparable NSF-wide data archiving site such as BCO-DMO if available. Nucleic acid sequence data will be subjected to QC pipelines and formatted in NCBI-approved formats for submission to GenBank and CAMERA (<http://camera.calit2.net/>). Appropriate metadata will be deposited according to the Genomic Standards Consortium specifications for the “Minimum Information about a Genome Sequence” (MIGS) (for RNA-seq data) and “Minimum Information about a Metagenome Sequence” (MIMS) (for metatranscriptomic data). Analytical laboratory data will be archived in native format on servers at UGA or DISL, and derived data products will be distributed in tabular format. Data will be made available at the time of the first publication that uses the dataset or within 2 years of collection.

Policies for Access and Sharing

This project will conform to NSF standards for data access and sharing. Metadata will be available immediately after data are archived, and data files will be openly and freely available on the web within two years from date of collection. Primary (raw) data and model output will also be archived along with the finalized data, and links to web-accessible files will be provided in the data set metadata. RNA and DNA will be stored for at least two years in a frozen sample repository initially established in duplicate in the Moran and Whitman labs.