

Data Management Plan

Data Policy Compliance

The project investigators 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.

Pre-Cruise Planning

Pre-cruise planning will be conducted by conference calls and email. Sampling instruments will include the ships CTD system (hydrographic sensors and bottle data), Saito Lab Trace Metal Rosette (hydrographic sensors and bottle data), AUV Clio (hydrographic sensors and particulate and water samples), and McLane pumps (backup to Clio). The actual sampling events will be recorded on paper logs (scanned into PDF documents) and/or in a digital event log using the R2R event logger application (if available). A final report will be prepared describing cruise activities and deposited to BCO-DMO and the foreign clearance nations.

Description of Data Types

This project will produce observational and experimental datasets.

Experimental datasets

1. Metalloproteomic experimental datasets (ICP-MS and LC-MS and enzyme activity) will be produced.

Observational Datasets:

1. CTD and Niskin bottle data: CTD data collected using a SeaBird SBE CTD package; processing to be done using SeaBird's SeaSave software; data will include standard environmental measurements (such as pressure, temperature, salinity, fluorescence). File types: Raw (.con, .hdr, .hex, .bl) and processed and .cnv, .asc, .btI) ASCII files. Repository: BCO-DMO.
2. Event log: Cruise scientific sampling event log; will include event numbers, start/end dates, times & locations of instrument deployments. Will be recorded using the R2R event logger (if available) and on paper log sheets. File types: Excel file converted to .csv; scanned PDFs. Repository: BCO-DMO and Rolling Deck to Repository (R2R).
3. Cruise underway data: Routine underway data collected along the ship's track (including meteorological data, sea surface temperature, salinity, fluorescence, ADCP). Will be collected by the shipboard instrumentation. File types: .csv ASCII files. Repository: BCO-DMO and R2R.
4. Macronutrient, dissolved metal, particulate metal, POC/PON, shipboard ammonia and nitrite data, and rate measurements, and iron uptake rates will be collected, compiled and submitted to BCO-DMO.
5. Metaproteomic data: Raw files collected in Thermo's proprietary raw format will be converted to nonproprietary format (either mzIdentML- or PRIDE XML) for submission. Processed datasets will be generated from our informatic pipeline as excel spreadsheets or CSV files.
6. qPCR and Metagenomic data: sequence data will be submitted to NCBI and linked to from the BCO-DMO project page, and qPCR results will be submitted to BCO-DMO.

Data and Metadata Formats and Standards

Field observation data will be stored in flat ASCII files, which can be read easily by different software packages. Field data will include date, time, latitude, longitude, cast number, and depth, as appropriate. Quality flags will be assigned according to the ODS IODE Quality Flag scheme (IOC Manuals and Guides, 54, volume 3; http://www.iode.org/mg54_3). Metadata will be prepared in accordance with BCO-DMO conventions (i.e., using the BCO-DMO metadata forms) and will include detailed descriptions of collection and analysis procedures.

We will endeavor to use standardized parameter names from the onset of data collection to facilitate data sharing and ingestion into BCO-DMO based on the SeaDataNet Parameter Usage Vocabulary.

Data Storage and Access during the Project

The investigators will store project data on laboratory computers equipped with automated backup systems. Field sensor data will be stored on portable hard drives and transferred to lab and cloud computers upon return to the laboratory. Mass spectrometry data is archived at WHOI on the Saito

laboratory's 40TB server in the institution's climate controlled computer room as well as being deposited to PRIDE. Project data will also be shared using Google Drive cloud services.

Mechanisms and Policies for Access, Sharing, Re-Use, and Re-Distribution

Immediately after completion of the research cruise, underway data and metadata will be submitted to the Rolling Deck to Repository (R2R) project. DNA sequences will be deposited in the National Center for Biotechnology Information (NCBI) database, GenBank, upon submission of manuscripts. GenBank accession numbers will be provided to the Biological and Chemical Oceanography Data Management Office (BCO-DMO) in an Excel spreadsheet or a .CSV file and metadata will be provided using the BCO-DMO Dataset Metadata submission form. Raw mass spectrometry data will be submitted to ProteomeXchange upon submission of manuscripts (through EBI's PRIDE or San Diego's MASSIVE repository system). Processed metaproteomic data will also be submitted to the EarthCube Ocean Protein Portal via a BCO-DMO ingestion using standard template. The Saito laboratory and BCO-DMO collaboratively created the Ocean Protein Portal (OPP) prototype with the intention of making it lightweight enough so as to be sustainable long-term, and it hosts processed ocean metaproteomic data from around the world. The data reuse policy for the OPP is modeled after the GEOTRACES data use policy, where if substantive use of data occurs, reaching out to the data generator is encouraged to foster collaboration, proper data use, and to encourage continued participation of data generators in the project.

Data produced by this project may be of interest to chemical and biological oceanographers, and climate scientists interested in the role of biogeochemistry in the global climate system. Protein data is also of interest to biochemistry (metals in biology) researchers who are interested in mechanistic aspects of protein biochemistry.

We will adhere to and promote the standards, policies, and provisions for data and metadata submission, access, re-use, distribution, and ownership as prescribed by the BCO-DMO Terms of Use (<http://www.bco-dmo.org/terms-use>).

Plans for Archiving

R2R will ensure that the original underway measurements are archived permanently at NCEI and/or NGDC as appropriate. BCO-DMO will also ensure that project data are submitted to the appropriate national data archive. The PI will work with R2R and BCO-DMO to ensure data are archived appropriately and that proper and complete documentation are archived along with the data.

All processed data will be submitted to BCO-DMO. Sequence data will be submitted to NCBI and raw mass spectrometry data will be submitted to ProteomeXchange.

Roles and Responsibilities

Each PI will be responsible for sharing his/her subset of data among the project participants in a timely fashion. Saito will be responsible for cruise event logs, dissolved metal data, the mass spectrometry and proteomic datasets. Santoro will be responsible for qPCR, metagenomic, rate data, and shipboard experimental datasets. Breier will be responsible for C lio hydrographic sensor data, POC/PON, and particulate metal dataset. Jakuba will be responsible for C lio engineering data. Saito will be the primary PI responsible for ensuring compliance with the data management plan.