Collaborative Research: Chemical and microbiological studies of water-soluble alkanes in the ocean

Data Policy Compliance

Identify any published data policies with which the project will comply, including the NSF OCE Data and Sample Policy as well as other policies that may be relevant if the project is part of a large coordinated research program (e.g. GEOTRACES).

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

If the proposed project involves a research cruise, describe the cruise plans. (Skip this section if it is not relevant to your proposal.) Consider the following questions:

1. How will pre-cruise planning be coordinated? (e.g. email, teleconference, workshop)
2. What types of sampling instruments will be deployed on the cruise?
3. How will the cruise event log be recorded? (e.g. the Rolling Deck to Repository (R2R) event logger application, an Excel spreadsheet, or paper logs)
4. Will you prepare a cruise report?

Pre-cruise planning will be done via teleconferencing and planned visits among the 3 different co-PIs. Sampling instruments on the cruise will consist of CTD rosettes, submersible vehicle instruments for oil/gas sampling, and atmospheric sampling via drone/weather balloon.

Description of Data Types

Provide a description of the types of data to be produced during the project. Identify the types of data, samples, physical collections, software, derived models, curriculum materials, and other materials to be produced in the course of the project. Include a description of the location of collection, collection methods and instruments, expected dates or duration of collection. If you will be using existing datasets, state this and include how you will obtain them.

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, .btl) 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. 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. **Atmospheric measurements of volatile hydrocarbons:** Measurements will be made using unmanned aerial systems fit with sorbent tubes for sampling C3-C8 hydrocarbons. Using an automated thermal desorption system coupled with a gas chromatograph and time-of-flight spectrometer we will recover, separate, and quantify linear, branched, and cyclic short-chain hydrocarbons from tubes packed with carbopack x sorbent. File type: Excel files. Repository: BCO-DMO.

Experimental Datasets:

1. **Water-soluble alkane respiration:** Incubations with water-soluble alkanes will be conducted with water collected from the deep Gulf of Mexico. Measurements of oxygen use will be conducted with optical oxygen sensors and an oxygen meter. Data will be stored in laboratory notebooks and scanned into PDFs. File type: Excel files. Repository: BCO-DMO.
2. **Metagenomic reconstruction of alkane degraders:**DNA will be extracted from incubations showing metabolic activity with water-soluble alkanes. Sequencing will be performed at UC Davis DNA Technology Core in Davis, CA after the cruise and DNA extractions are completed. File types: Short-read archive (.sra) and .fasta files. Repository: NCBI, accession numbers will be provided to BCO-DMO.

Data and Metadata Formats and Standards

Identify the formats and standards to be used for data and metadata formatting and content. Where existing standards are absent or deemed inadequate, these formats and contents should be documented along with any proposed solutions or remedies. Consider the following questions:

1. Which file formats will be used to store your data?
2. What type of contextual details (metadata) will you document and how?

3. Are there specific data or metadata standards that you will be adhering to?
4. Will you be using or creating a data dictionary, code list, or glossary?
5. What types of quality control will be used? How will data quality be assessed and flagged?

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. 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.All data will be assessed and flagged for quality issues before it is added to BCO-DMO.

Data Storage and Access During the Project

Describe how project data will be stored, accessed, and shared among project participants during the course of the project. Consider the following:

1. How will data be shared among project participants during the data collection and analysis phases? (e.g. web page, shared network drive)
2. How/where will data be stored and backed-up?
3. If data volumes will be significant, what is the estimated total file size?

The investigators will store project data (including spreadsheets, ASCII files, images from ROV and AUVs on laboratory computers that are backed up on servers on campus. Personal computers will be backed up daily using Apple Time Machine to an onsite external hard drive. All data will also be uploaded to cloud storage systems by all PIs monthly.

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

Describe mechanisms for data access and sharing, and describe any related policies and provisions for re-use, re-distribution, and the production of derivatives. Include provisions for appropriate protections of privacy, confidentiality, security, intellectual property, or other rights or requirements. Consider the following:

1. When will data be made publicly available and how? Identify the data repositories you plan to use to make data available.
2. Are the data sensitive in nature (e.g. endangered species concerns, potential patentability)? If so, is public access inappropriate and how will access be provided? (e.g. formal consent agreements, restricted access)
3. Will any permission restrictions (such as an embargo period) need to be placed on the data? If so, what are the reasons and what is the duration of the embargo?
4. Who holds intellectual property rights to the data and how might this affect data access?
5. Who is likely to be interested in re-using the data? What are the foreseeable re-uses of the data?

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 .CSV file and metadata will be provided using the BCO-DMO Dataset Metadata submission form. Data sets produced by the science party will be made available through the BCO-DMO data system within two-years from the date of collection. The project investigators will work with BCO-DMO data managers to make project data available online in compliance with the NSF OCE Sample and Data Policy. Data, samples, and other information collected under this project can be made publically available without restriction once submitted to the public repositories.

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. 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

Describe the plans for long-term archiving of data, samples, and other research products, and for preservation of access to them. Consider the following:

1. What is your long-term strategy for maintaining, curating, and archiving the data?
2. What archive(s) have you identified as a place to deposit data and other research products?

Samples collected on the cruise well be stored at the appropriate temperature in the dark using fridges, freezers, and incubators conencted to back-up generators in the case of power loss. All raw data generated from analysis of these samples and CTD data with hydrographic and water chemical properties is immediately uploaded to a data server and marked as read-only. The finalized forms of these data will be added to public repositories as per NSF guidelines. Once uploaded to public archives data will be perpetually available to the public. All community analysis data will be uploaded to BCODMO and NCBI. All geochemical measurements will be uploaded to BCO-DMO, NODC, and NGDC databases. R2R will ensure that the original underway measurements are archived permanently at 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.

Roles and Responsibilities

Describe the roles and responsibilities of all parties with respect to the management of the data. Consider the following:

1. If there are multiple investigators involved, what are the data management responsibilities of each person
2. Who will be the lead or primary person responsible for ultimately ensuring compliance with the Data Management Plan?

David L. Valentine is the lead PI and the primary person responsible foe ensuring compliance with the Data Management Plan. He is responsible for all data related to the microbiology of alkane degraders. Robert Swarthout (co-PI) is responsible for collecting data related to atmospheric measurements of volatile hydrocarbons. Christopher Reddy (co-PI) is responsible for collecting data related to water-column measurements of water-soluble alkanes. Each PI will be responsible for sharing their subset of data among the project participants in a timely fashion.