Plan Overview

A Data Management Plan created using DMPTool-Stage

Title: CAREER: How does recalcitrant organic matter become bioavailable? Mechanisms and controls on sediment organic matter oxidation by microbial communities

Creator: Andrew Steen

Affiliation: University of Tennessee at Knoxville (utk.edu)

Principal Investigator: Andrew Steen

Data Manager: Andrew Steen

Funder: National Science Foundation (nsf.gov)

Funding opportunity number: 17-537

Template: BCO-DMO NSF OCE: Biological and Chemical Oceanography

Last modified: 07-01-2024

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customize it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

CAREER: How does recalcitrant organic matter become bioavailable? Mechanisms and controls on sediment organic matter oxidation by microbial communities

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).

This project will comply with the NSF OCE Data and Sample 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?

1.) Cruises will be planned by coordination with UNOLS and the Captains of relevant vessels. A site visit is budgeted and planned for the PI a year before the first cruise on each vessel, to speak with the Captain and tour the vessel. The PI will invite unfunded collaborators to fill extra berths, and will coordinate via regual Skype meetings and email prior to the cruise.

2.) Sample collection on the boat will be by gravity corer and box- or multi-corer, which do not collect data. Therefore only standard underway measurements will be collected (salinity, temp, fluorescence) and stored via R2R.

3.) The cruise event log will be recorded via R2R application.

4.) A preliminary cruise report will be prepared within 1 month of the end of the cruise report.

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.

1.) Observational data:

Observational data will include geochemical and biological parameters relating to cores to be collected from 4 stations at each of three field sites. White Oak River sites will be sampled 4 times, and Mid-

Atlantic Bight samples will be collected 2 times. Data will include oxygen, methane, DIC, anion, and cation concentrations, as well as DIC and methane 13C ratios, direct microscopic counts, and enzyme assays.

DNA and RNA sequences will also be collected from each site.

2.) Experimental data: Geochemical experimental data will mainly consist of concentrations of small molecules released by treatment of sediment with enzymes.

Assessments of student learning for the educational plan will also be collected. These data for these will not be publicly released in order to maintain privacy. Aggregated data will be made available in a white paper to be posted on the PI's website and possibly a published paper, if IRB approval is granted.

3.-4.) No model simulations or derived data sets are expected.

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?

Geochemical and experimental data raw data will be stored as .csv files, in tidy format to the extent possible (Wickham 2014, Journal of Statistical Software 59:10). Sequence reads will be stored as .fastq files. Each file will be paired with a README text file that will contain metadata and explanations (units, abbreviations, brief methods, etc). To the extent practical, all code (e.g. R scripts, in-house packages, and batch files or makefiles) used to convert raw data into final results, tables, and figures will be included.

Separately, to the extent practical, all raw data and code will be posted to the PI's Github site. This has been the practice of the Steen Lab since about 2013.

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?

During the project, data will be stored and shared in three redundant ways: via the Steen Lab's Dropbox for Business account, which can be accessed by any member of the project. Ability to modify files is

restricted to the 'owner' of the data set in order to prevent accidental modification or deletion (this is practical only for datasets smaller than ~50 GB). Secondly, all data will be stored on the Steen Lab's 12 GB RAID-6 array, which resides in a locked drive array in our office space. Finally, data will be backed up to University of Tennessee's Advanced Computing Facility data storage system. Geochemical and epxerimental data will be "small" - less than 1 GB - but sequence data may total > 100 GB.

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?

1. All data will be made available shortly ahead of publication, or at the end of the project, whichever comes first. Geochemical and experimental data, as well as "reduced" results from analysis of sequence data (e.g., lists of taxa and genes) will be archived with BCO-DMO according to policy. Raw sequence data will be archived at NCBI's Sequence Read Archive.

2.-3. Results of student assessments are considered sensitive, and will not be released publicly. All other data will be released publically without embargo.

3. PI Steen will own IP rights to the data, and will release them publicly (e.g. under an appropriate Creative Commons license).

4. Other biological and chemical oceanographers may be interested all of the data sets, particularly the sequence data which may be incorporated into other sequence data sets. We will take care to advertise the existence and informally facilitate the reuse of all data.

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?

All "small" data will be archived with BCO-DMO with appropriate documentation and metadata. Sequence data will be archived at NCBI's Sequence Read Archive. Educational assessment data will not be archived beyond the end of the project due to privacy concerns.

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?

As the sole PI, PI Steen is responsible for compliance with the Data Management Plan.