

## Plan Overview

---

*A Data Management Plan created using DMPTool*

**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, 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:** 07-21-2017

### **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**

This project will comply with the NSF OCE Data and Sample Policy.

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 regular 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.

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  $^{13}\text{C}$  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.

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.

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 experimental data will be "small" - less than 1 GB - but sequence data may total > 100 GB.

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.

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.

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

---