DATA MANAGEMENT PLAN

Data Archive and Dissemination. The proposed project will generate several types of data including (i) oceanographic data, (ii) chemical ionization and electron impact data, (iii) NMR data and (iv) isotopic data and thermographs.

1. The cruise planned for March 2018 (courtesy of Dr. Ken Smith) will generate some biogeochemistry data. With the exception of the mass spectrometry data, the PI will make all other datasets publically available through the CCE-LTER Datazoo platform (oceaninformatics.ucsd.edu/datazoo/) and also submit these datasets to be managed by the Biological and Chemical Oceanography Data Management Facility where data will be available online through http://bco-dmo.org/data/. Data to accompany the GO-SHIP samples, that have already been collected, will be available through the CLIVAR program but we will make our subset available as well. Standard operating procedures for these analyses will be linked to data files and also available through publications. The pre-cruise plan will include station locations and sample collection protocols while actual sampling locations and measurements will be recorded on paper logs (and photographed following each station to be converted to electronic files) and digital logs.
2. All raw instrument data files (NMR free induction decays (FIDs), GC/MS chromatograms and mass spectra, raw isotope traces), method files, and sequence files will be maintained in at least two places: on the instrument computers where they were collected, and on a central backup drive at the host institution (SIO, UCSD, WHOI-NOSAMS). For example, after data acquisition from experimental runs, all raw mass spectra will be stored on an online SAN connected storage cluster that is owned and maintained by UCSD. We anticipate being able to archive data indefinitely in this fashion. These data will be freely shared upon request following publication of the relevant results. However, these files use proprietary formats specific to instrument software that are not accessible to everyone. For this reason, the field does not commonly share data at this level.
3. All handwritten notes and field data will be recorded in laboratory notebooks, which will be scanned and archived on the lab backup drives.
4. All ‘processed’ data, which includes thermographs, Fourier Transformed NMR FIDs, GC×GC classification scheme data, calibrated GC-flame ionization detection (FID) abundances, user-generated MS libraries and $\delta^{13}C$ values, $\Delta^{14}C$ values, and sample metadata (location, depth, volume) will either be published as supplementary electronic files accompanying peer-reviewed manuscripts or be available from the PIs upon request following publication of the data. We will also include all of the raw ramped temperature oxidation data and corresponding $\delta^{13}C$ and Fraction Modern data as Electronic Annex or Supplementary Information files (MS Excel format). NOSAMS accession numbers are searchable (by NOSAMS) and will accompany all isotope data. These data along with any user-submitted information are stored at NOSAMS indefinitely. We will strive to make our complete datasets accessible to the public either via the publisher’s website or upon request in a format used by commonly available software.
5. The most "sharable" data that will result this investigation are the radiocarbon data and thermograms generated from the ramped pyrolysis oxidation (RPO). The publications have all of our isotope data, thermograms and derived oxidation kinetics included as supplementary information. These datasets will also have their metadata (which is limited to CTD information) reported in the same table. We will aslos include our RPO data through the RPO database maintained on GitHub by Jordon

Hemingway (http://github.com/FluvialSeds/RPO_Database). This site is a repository for several aquatic RPO datasets. Our samples were analyzed at NOSAMS using nearly identical parameters, and so, we expect them to be useful as a comparative dataset.

6. Data from analyses in listed above and sample metadata (location, volume) included in publications will be linked to online data repositories in (1) or provided as supplementary electronic files accompanying peer-reviewed manuscripts. When possible the metadata will be linked to the mass spectrometry (MS) datasets (below).

7. The major data types generated during carotenoid degradation experiments include UV visible spectra and liquid chromatography-mass spectrometry data. These data are primarily derived from the oxidation of model compounds under artificial sunlight and as such, represent fairly unique datasets. For example, the lamp conditions, temperature control etc., are investigation-specific. All of these data will be provided in our papers. A compilation of all compounds detected by MS, their unique identfiers, and their bulk characteristics will be provided in a supplementary table with a link to their Mass Spectrometry Interactive Virtual Environment (MassIVE) dataset (e.g. for fucoxanthin: https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?task=399b88674f514cdaac1a1d6d79839874 and Global Natural Product Social Molecular Networking repository (Fucoxathin: https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=eed955f447fa4bc9ad0116b8e24c2303). Each carotenoid has its own unique identifier and all datasets will be publicly available. These databases will now serve as additional compound libraries for the LC-MS community, and as a living site, compounds in our dataset that we were unable to identify can be annotated by other users if they match to a known compound in their dataset.

   Sample archive. The extraction, workup, and analyses of described in this proposal will consume most of the sample, thus we expect there to be little material to archive.