

Data Management Plan

This project will generate data primarily consisting of the concentrations and $\delta^{34}\text{S}$ values of bulk DOS, or of individual organosulfur compounds. This will be a small (in terms of file size) dataset that does not require any elaborate schemes for management. In general, the data will be curated as follows:

1. All raw instrument data files (HPLC traces, EA/IRMS chromatograms), method files, and sequence files will be maintained in at least two places: on the instrument computers where they were collected, and on the Sessions' group central backup drive. The latter is currently implemented as a 6 Tb network-isolated RAID hosted on a server in our labs. Hard disk sizes are increasing much more rapidly than we are accumulating data, so we anticipate being able to archive data indefinitely in this fashion. The network drive serves both as a common data repository for the group, and as a redundant (mirrored-drive) backup. These data will be freely shared upon request following publication of the relevant results. However, the 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 (see #3 below).
2. All handwritten notes will be recorded in laboratory notebooks, which will be scanned and archived on the lab backup drive.
3. All 'processed' data, which for this project means mostly concentrations and $\delta^{34}\text{S}$ values, will be published as supplementary electronic files accompanying peer-reviewed manuscripts, commonly in CSV or Excel format. This approach means that our complete dataset is accessible to the public via the publisher's website in a format used by commonly available software.
4. All processed data will be submitted to appropriate public databases following publication. Because we are working primarily on previously collected samples, we deem it preferable to associate our data with those samples in the archives where they were previously deposited. This means the National Center for Environmental Information (NCEI, formerly National Oceans Data Center) for samples from US collaborators Aluwihare and McCarthy, and the German Oceanographic Data Center (DOD) for samples from collaborator Dittmar. If it is feasible to deposit our entire dataset into both repositories, then we will do so. Data for new samples collected from HOT and BATS will be archived in the repository used by those programs, i.e. NCEI.