DATA MANAGEMENT PLAN

Metal and Nutrient Data

This project will produce nutrient concentrations, metal concentrations (ICPMS) and iron isotope (MC-ICPMS) data for a range of dust samples, leaching experiments and incubations. We intend to follow GEOTRACES clean protocols and intercompare techniques with other investigators, including use of the Arizona Test Dust, seawater and isotope reference standards. All data and intercomparison will be published and made available online, and submitted to the Biological and Chemical Oceanography Data Management Office (BCO-DMO). Where appropriate, we will also submit data to the GEOTRACES Data Assembly Center (GDAC), based at the British Oceanographic Data Center (BODC) in Liverpool, UK, and anticipate this data would be made publicly available via GDAC and possible inclusion in GEOTRACES data products and electronic atlases.

Data products:

Raw data consists of collections of recorded mass spectra of blanks, quality control standards (universal to all analyses), analytical standards (project specific), and samples. Metadata will be stored along with raw data, and includes descriptions of the samples, preparation protocols, and other relative information including experimental conditions, sample quantity/concentration, solvents, and analytical conditions (including internal standardization, chromatographic conditions, mass spectrometry methods).

Data formats:

Raw data is electronically generated and stored in Thermo Fisher Scientific *.raw files. Metadata and the products generated by research vary depending on sample and application type. Information provided by users including sample description and preparation, along with relevant metadata will be recorded and stored in various electronic file formats (e.g. *.xlsx, *.docx). Processed mass spectrometry data will be stored as *.mzML open source ms data formats and extracted/aligned peak lists stored as *.csv files.

Data archive:

We will archive all raw data and meta-data containing sample/analysis information on our dedicated metabolomic data server housed within the environmental computing center (ECC) of the College of Earth, Ocean, and Atmospheric Sciences at Oregon State University. This center contains modern, high-throughput computing-class infrastructure and large back-ed up memory systems, and is funded through college support and research computing fees paid by all faculty and staff. Data storage at the ECC is maintained with regular back-ups by a team of administrators and computer specialists. Raw and processed data will be further backed up on cloud storage on Box using Oregon State University's license for unlimited secure storage when applicable.

Data processing:

Data processing will begin with quality control checks conducted by Boiteau lab staff. All sample batches begin and end with analytical blank and quality control standard analysis. Samples are spiked with internal standards (generally metalated complexes or biotin) immediately before analysis to monitor instrument performance. Pooled samples are run

periodically throughout sample batches. These precautions help us identify issues early on and ensure that only high-quality data is analyzed. Data processing will be conducted on our dedicated high performance mass spectrometry informatics computer using CoreMS and custom Python-based scripts generated in-house. Custom code used for publications will be made openly available for download on Github.com by the broader scientific community and cited in publications.

Data sharing and access:

Data and associated metadata will be made available via BCO-DMO after embargo periods or within two years of collection, with information about the repository and data set included in publications.