Data Management Plan

Data Policy Compliance

The project investigators will comply with the data management and dissemination policies described in the NSF Award and Administration Guide and the NSF Division of Ocean Sciences Sample and Data Policy. The research project will be registered with the BCO-DMO repository (www.bco-dmo.org), and this infrastructure will be used to share and preserve data for longterm access.

Data Collection and Data Types

Proper data collection and processing is part of all operating protocols used for this project. Analytical equipment will be calibrated and checked regularly as part of our standard quality assurance. The PI and Co-I have extensive experience in collecting the physical and digital data produced from this research.

A. Physical samples: The physical samples this project will generate comprise artificial seawater nitrate standards as well as purified nitrate fractions from artificial mixtures and marine samples.

B. Digital data: Digital data generated in this project will comprise data on seawater ion concentrations and nitrate fractions collected by ion chromatography, as well as mass spectrometric data collected by ESI-Orbitrap. All raw analytical data will be stored in their original file formats to preserve maximal information content: Chromeleon .cmbx (50MB/analysis) files for ion chromatography; Xcalibur .raw (5-500MB/analysis) files for ESI-Orbitrap; and open XML spreadsheets (.xlsx, herafter XLSX) for all manually collected data (sample and experimental meta-data, 50kB-2MB/file). All raw data files will be stored together with meta-data in XLSX format (50kB-2MB/file), including unique sample identifiers (UIDs), processing methods, and analytical conditions. All fully processed data will also be stored in XLSX format (50-500kB/file). All data reduction and processing scripts as well as analysis and visualization code will be stored in their original file formats (.Rmd for RMarkdown, .ipynb for Jupyter notebooks, 10kB to 5MB/file) together with all necessary auxiliary information for reproducible data reduction in XLSX format (10kB-1MB). Data volumes for the different file formats and data types are best estimates from average file sizes previously recorded for similar data.

Data Storage, Preservation, and Responsibilities

A. Physical samples: All nitrate and seawater samples that are not completely consumed by the analytical procedures will be archived frozen at CU Boulder for a minimum of 5 years after completion of this project and will be available for further study upon request.

B. Digital data: Our data archiving and access preservation strategy follows the rule of 3-2-1 for each data file: at least 3-copies at 2-physical sites per 1-file. During the project, data will be stored in commonly readable formats on digital media (copy 1, site 1) and kept secure at the University of Colorado Boulder using commonly accepted digital security practices (https://www.dataone.org/best-practices). Data will also be locally backed up to secure external drives and servers belonging to the PI (copy 2, site 1). Finally, the team will use the Open Science Framework (OSF) repository to archive raw data, metadata, protocols, analytical results, and final processed data at all levels for a minimum of 50 years (copy 3, site 2). The OSF is a free, cloud-based repository for researchers, which allows easy and rapid electronic access to the entire project team as well as full data sharing with outside parties.

In addition, the BCO-DMO data repository and its infrastructure will be used to share and preserve data for longterm access. Data sets produced will be made available through the BCO-DMO data system within two-years from the date of collection. The project investigators will work with BCO-DMO data

managers to make project data available online in compliance with the NSF OCE Sample and Data Policy. Metadata will be prepared in accordance with BCO-DMO conventions (i.e. using the BCO-DMO metadata forms) and will include detailed descriptions of collection and analysis procedures.

Data Distribution and Public Access

As outlined above, project information and data will be deposited and maintained primarily on OSF and the BCO-DMO data center, which allow easy and rapid electronic access to the entire project team as well as full data sharing with outside parties and the public. Sharing of data beyond the project team before publication will require permission of the PI, Co-I and participating collaborators, but the intent is to make all data broadly available to the community as quickly as possible. Both PI and Co-I have a track record of efficient and transparent data distribution in publications, and a record of participating in community data-sharing activities. In all cases, data and code will be made publicly available at the latest by 2 years after their collection or production in this project. Processed data and other results will be presented at professional meetings and reported in peer-reviewed journal articles. The data sets for each publication will become permanently available as online supplementary material in the formats supported by the individual journals, through the persistent identifiers (DOIs) provided by the OSF and BCO-DMO, and through dataset specific repositories. All code developed for this project will be archived and made accessible through the open-source online code repository GitHub as previously demonstrated in publications by the PI. All team members will have an OrcID (http://orcid.org/), so that their publications and data sets are easily and permanently connected to them.