

DATA MANAGEMENT PLAN

1. Types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project. The study will produce a range of data types from the sampling and experiments conducted. The data types include raw and corrected instrumental data of isotope ratio measurements, derived nitrogen fixation rates, gene sequences, and gene copy abundances. Ancillary hydrographic data will also be collected on particulate C and N and chlorophyll a. Ancillary hydrographic and chemical measurements will also be collected at each sampling site (temperature, salinity, photosynthetically available radiation, nutrient concentrations), in addition to total organic carbon and nitrogen concentrations and chlorophyll-a concentrations.

2. Standards to be used for data and metadata format and content. All data will be collected using certified reference materials (CRMs) where available, including CRMs for nutrient analyses (NRC Institute for National Measurement Standards, <http://www.nrc-cnrc.gc.ca/eng/services/inms/reference-materials.html>), and chlorophyll among others (http://www.turnerdesigns.com/t2/doc/appnotes/998_0058.html). Where CRMs are not available, other best practices will be used to include internal standards or appropriate controls. Similarly, stable isotope ratio analyses will be standardized vs. internationally recognized isotope standards from the IAEA and USGS.

During the submission process of the data to depositories, the required metadata are provided following the data depository guidelines – see 5) below. As data are generated the data and metadata will be archived primarily on Excel spread sheets. Spreadsheets with data will be backed up at least weekly using TimeMachine or a similar program. Lab notebooks containing the raw data generated by this project will be kept. Detailed descriptions of procedures and methodological approaches, deviations from protocols, specific equipment, and chemical reagents utilized for this project will be catalogued by all project participants. Physical notebooks and additional password protected electronic notebook resources (electronic notebooks shared with PIs) serve as the complete notebook record. Notebooks and digital copies thereof will serve as permanent records of the project and will be available upon request from NSF program managers.

3. Policies for access and sharing, including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements For sharing files between the laboratories we will use in-house or cloud based storage which are password protected. Students and personnel working in the project will contribute to these datasets, sharing their results with the team, and will have access to data from the entire team. The data distribution regarding publications between different team members will be decided collectively with the entire team. The inclusion as an author in manuscripts will be determined based on criteria by the International Committee of Medical Journal Editors (<http://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html#two>). Distribution of data to Biological and Chemical Data Management Office (BCO-DMO <http://www.bco-dmo.org/>) will be done within two years of their collection via a password-secure FTP. All data generated as a result of research performed at the UMassD, UConn, or OSU, shall be the intellectual property of thereof and the PIs overseeing the research. As such, the Universities will retain an implied copyright for these data.

4. Policies and provisions for re-use, re-distribution, and the production of derivatives Data from these laboratory experiments will be submitted to BCO-DMO where they will be available to the public. Once submitted and made publically available via the archival databases described above, all data and research products will be available for unrestricted re-use or re-

distribution. Acknowledgement of the original source of the data will be requested, but not required for their use. Archived and frozen (-80°C) sub-samples of backup samples will be available to interested parties for their independent analysis following publication of results.

5. Plans for archiving data, samples, and other research products, and for preservation of access to them. All aspects of our research will be documented in both physical (Lab Notebooks) and digital formats. Logs will be kept by PIs and project participants for all samples collected and analyses performed. The PIs will collaborate closely to develop the NSF annual/final report. The PIs will be in charge of archiving the raw data from analytical measurements conducted in the respective laboratories. Moisaner will archive the sequence and qPCR data and will maintain records of instrument raw and analyzed data files for all the sequence analyses. Sanger sequencing data will be submitted to GenBank (<http://www.ncbi.nlm.nih.gov/>), and the amplicon and metagenomic/metatranscriptomic data to NCBI sequence read archive (<http://www.ncbi.nlm.nih.gov/sra>), respectively. Sequence data will be submitted with logistical information, with all required metadata, as well as ancillary data as much as where they are available. Cruise data will be submitted by the PI to the BCO-DMO. We are familiar with these data archival systems and have previously submitted to them. Any other relevant data relevant that does not conform to the above archival systems, including metadata supporting common sampling and experimental activities will be deposited in the Dryad database, (<http://datadryad.org/>).

5.1. Short-term storage and data management. Data storage shall be appropriately and redundantly stored using the computer hardware and software available to the PIs, and may include both in-house and/or offsite resources. The PIs are responsible for due diligence with respect to short-term storage of data. Additionally, all data shall be retrievable from primary media or back-ups, as well as reasonably protected from accidental loss due to corruption, power loss, or failure of computer hardware. Data shall be stored on either off-network mobile devices (external hard drives) or offsite cloud resources, using password protection. These data (and associated passwords) must be made available to senior institution officials in the case that any institutional liability issues should arise. Stored data shall be backed-up weekly on redundant in-house data backup systems.

5.2. Deposit and long-term preservation. Within two years of data collection, data will be transferred to BCO-DMO for public access and long-term storage. After the project has been completed, arrangements shall be made to transfer data from short-term storage to a long-term archival systems. Data will be kept in long-term storage for at least five years, or until it has been successfully uploaded to and made publically available through a nationally or internationally funded database specific to that data.