

DATA MANAGEMENT PLAN.

1. Products of Research.

As a result of the multi-disciplinary nature of the proposed research, a large array of data would be generated by multiple methodologies and from field and laboratory samples including, but not limited to, basic hydrographic data, biogeochemical concentrations and rate data, transcriptomic and proteomic data, and Nuclear Magnetic Resonance spectra. The data generated by this type of study does not lend itself to easy assimilation into one dataset that can be stored at one location. To overcome this situation, project participants would seek to use multiple data-specific repositories, such as the public databases available through JCVI or GenBank for transcriptomic data, or the Tranche repository found at the Proteome Commons website (<https://proteomecommons.org/tranche/>) for proteomics data (primarily in accordance with any requirements for peer-reviewed journals).

In order to provide a wider oceanographic community with information regarding the types of data generated by the project, the project would be registered with the Biological and Chemical Oceanography Data Management Office (BCO-DMO) as for any project funded by NSF Division of Ocean Sciences, and metadata files outlining the data generated provided to BCO-DMO by the project's principal investigators. Consequently, locations of the repositories (and the data itself) can be linked to the central data facility, such as BCO-DMO, through metadata.

2. Data Format, Data QC, and Metadata.

Data in a range of formats would be generated during the proposal. In many cases, particularly with the “-omics” technologies being employed, the data formats will be platform/technique specific. Raw data generated by the different techniques will be held at the home institutions of the various collaborators and transformed to easily disseminated formats as deemed appropriate by the appropriate PI. Similarly, quality control procedures for the different datasets will be the responsibility of the individual collaborators. This step will allow the initial data collection from each of the components of this study to be managed according to the standard practices and quality control procedures for the contributing laboratories (and their home institutions). This aspect of the data management would also be conducted in consultation with BCO-DMO to ensure adequately detailed metadata files were supplied to BCO-DMO.

3. Access to Data and Data Sharing Policies.

All aspects of the data management for this project (processing, submission, sharing and archiving) will be handled in accordance with current NSF and Division of Ocean Sciences data sharing policies. The data generated during this project would be made publicly available

immediately after publication of the data in peer-reviewed journals or two years after the completion of the project, whichever deadline occurs first. Data included as part of a peer-reviewed publication would be formatted as per the journal's editorial policy and made available as supplementary material where applicable. Requests for unpublished data presented at conferences prior to being made publically available would be handled on a case-by-case basis by the lead PI, in consultation with the project collaborators. Standard oceanographic and biogeochemical data obtained during the research expedition, such as temperature, salinity and oxygen profiles, would be posted to BCO-DMO. Any oceanographic data obtained from nearby NOAA data buoys (i.e. NBDC Stations 41004, 41002 and 41008) would automatically be available through the National Data Buoy Center.

4. Policies for Re-use, Re-distribution and Production of Derivatives.

We anticipate that re-use or re-distribution of any data published in scholarly journals, or submitted as supplementary materials to a publication, would be governed by the copyright policies of the publishing journal. Any unpublished data submitted to a publically-accessible data facility would be automatically available for re-use, redistribution and the production of derivatives.

5. Archiving of data

(i) Short term storage

During the proposed project and for the first two years after the completion of the project, all raw data and data products would be stored (and backed-up) by the individual collaborators at their home institutions in accordance with their lab's standard operating procedures and home institutions' continuance of operation plans. An additional copy of the data would be held by the lead PI on the servers at his home institution.

(ii) Long-term storage

At the completion of the project, all raw data and data products would continue to be stored by the individual PIs. Two year after the completion of the project, all processed electronic data would be transferred to appropriate public-access data repositories in accordance with NSF data management policies. In addition, we would explore the possibility of delivering semi-processed data in a compact form (as Excel spreadsheets) as part of the 'Supplemental Data' for any journal articles. This approach would utilize existing, long-lived publishers' archives as one way of satisfying the requirement for public access to the data. In general, the supplemental data associated with a journal article is 'free' for download.