

DATA MANAGEMENT PLAN

Scientific Data component

Data and metadata collected under the project, “Collaborative Research: Diel Physiological and Vertical Migratory Rhythms in a Tropical Ocean Copepod” will be served and archived through a number of repositories, including the Biological and Chemical Oceanography Data Management Office (BCO-DMO) data collection, the PRIDE archive and the NCBI GenBank.

We will use the pre-cruise planning of sampling strategies and the BCO-DMO guidance on best practices for cruise data management (e.g., generation of cruise reports and sampling event logs). We also will work with BCO-DMO members to publish cruise and experimental data in a timely fashion, compliant with the requirements of the Division of Ocean Sciences Data and Sample Policy. Transcriptomic data and metadata will be separately stored the NCBI GenBank while proteomics data will be housed on the Chorus Project website (see below). These datasets will be cross-listed on BCO-DMO so that all outcomes of the project have a centralized repository.

Maas and Blanco Bercial maintain continuous backups of their personal computers on a business class dropbox drive which also allows for easy sharing among collaborators at different institutions. Data backups will further be maintained throughout the project on external hard drives in each PIs labs and archived on BIOS’s 96TB enterprise class storage array at the end of the project.

Specific plans for particular data types include:

1. **Hydrographic Data:** Metadata for chemical and physical data collected during hydrographic casts made concurrent to sampling will be supplied to BCO-DMO.
2. **Metabolic experiments:** Metadata for the metabolic experiments will similarly be stored in BCO-DMO and a DOI created for the final dataset.
3. **Gene Expression Data:** All original sequence read data files will be uploaded to the NCBI Sequence Read Archive (SRA) database. All metadata and data associated with the differential expression analysis will be submitted to the NCBI Gene Expression Omnibus (GEO) database. The project code capturing this BioProject will be included in associated publications and the link will be posted to BCO-DMO.
4. **Proteomics Data:** Proteomics data (protein and peptide identification and supporting spectral evidence) will be stored by Dr. Nunn’s lab, which uploads published and unpublished mass spectrometry data within 1 week of collection onto Chorusproject.org for free public download, viewing, and analysis under the “NUNN LAB”. The Chorus Project was created for the sole purpose of providing research scientists and developers with the ability to store, analyze, and share their mass spectrometry data. The Chorus application is available on the cloud to anyone and can be accessed for free by the scientific or general public, allowing the dissemination and sharing of data no matter the original raw file format or mass spectrometry platform used to generate the data. Associated location and preparation details will be provided with the analysis files. A link to this dataset will be included in associated publications and will be posted to BCO-DMO.
5. **Peer-reviewed publications:** When possible, results will be published in open access journals, using funds allocated in the budget. In addition, we will follow the most current guidelines, recommendations and requirements from NSF regarding access to publications (“Today’s Data, Tomorrow’s Discoveries”) including archiving within the PAGES (Public Access Gateway for Energy and Science) repository within twelve months of publication.

Educational Data component:

The curriculum elements to be developed in collaboration with the Ocean Academy will be distributed and shared with the marine education community through the Ocean Academy resources webpage. We expect publication of the resulting lesson plan in either the Journal of Marine Education or Science Scope geared towards Middle Schools.

Archiving of samples:

Transcriptomic samples will be stored in RNAlater, with spare individuals archived in a -80°C freezer for up to 7 years at Woods Hole Oceanographic Institution. Left over extracted RNA from pooled samples will similarly be archived until the publication of all datasets to provide backup validation of findings.

Proteomics samples will be digested and analyzed using mass spectrometry. Remaining peptides will be stored in a -80°C freezer for up to 7 years at U. Washington and available for further analysis if requested by outside labs. Spare individuals will similarly be stored in case of need of further analyses.