

Data Management Plan

Data Products: Multiple data types will be generated in this project from field surveys, and field and laboratory experiments including biogeochemistry (e.g., total alkalinity, dissolved inorganic carbon, pH, temperature salinity, nitrate + nitrite, ammonium, phosphate, dissolved organic nitrogen, fluorescent dissolved organic matter, and radon), environmental time-series (e.g., temperature, salinity, and pH), experimental physical and biological data (e.g., coral growth, *Symbiodiniaceae* counts, macroalgae standing biomass, bioerosion rates, microbial community composition), and electric resistivity imaging. All data will be stored as .csv files or .png (imaging), and R scripts. Microbial community analyses will comprise two cohesive DNA sequencing datasets with corresponding metadata submitted to the NCBI Sequence Read Archive as we have done previously. Metadata on survey and experimental designs will be crucial to interpreting the data, and our management strategy ensures consistent formatting and archiving of linked data products. Metadata, biogeochemical data, and biological data will be incorporated into a relational database facilitated by Silbiger. This will enhance accessibility of the data to all PIs providing a searchable platform to promote interpretation and connectivity of all datasets.

Biogeochemical and biological data accessibility: We will immediately contact the Biological and Chemical Oceanography Data Management Office (BCO-DMO) and register our project by submitting project metadata. All field and experimental data and metadata will be submitted to the BCO-DMO for archiving and public dissemination. The data will be contributed to BCO-DMO within 2 years of their production to comply with NSF OCE data dissemination and archiving policy. The data will be accompanied by all relevant activity logs (mesocosm set-up, perturbations etc.). The PIs are already in contact with Cynthia Chandler, the BCO-DMO data manager, regarding required metadata and metadata standards optimum format for data submission, as they have been submitting similar data collected during prior NSF-funded projects (e.g., BCO-DMO ORCID: [0000-0003-2525-3496](https://orcid.org/0000-0003-2525-3496)). Ancillary data products from other programs, including oceanographic data, reef habitat data, and meteorological data, will also be important in interpretation and will be linked through open-access data repositories.

Short term data storage and organization: Data will be collected and added to lab databases that will be mirrored on laptop computers, the CSUN Biology departmental server, a GitHub repository, the C-MORE in-house network servers, and in the Google Drive computing cloud, with weekly backup on local hard drives. As data are generated, they will be collated and stored on shared internal databases and will be accessible by all persons involved in the project. Data will be archived in the original data format to facilitate future data usage.

Metadata Formatting and Archiving: Metadata on experimental designs will be crucial to interpreting the data, and our management strategy ensures consistent formatting and archiving of associated data products. Metadata forms in Ecological Metadata Language (EML) are utilized to create and organize the overall project database for local use, and upon publication for increased ease of data dissemination (see “Publication” section below). Marker gene amplicon metadata will be carefully formatted according to the Genomic Standards Consortium recommendations (MIMARKS). Metadata will follow BCO-DMO guidelines, including variable names, derived units, experimental set ups, analysis methods, descriptions of synthesis or calibration procedures where appropriate, data location, season, and quality control information. Variable names, keywords and metadata standards will follow guidelines available from the Marine Metadata Interoperability Project (marinemetadata.org) and the vocabulary and open access data management best practices. Ocean acidification data will follow EPOCA’s *Guide to best practices for ocean acidification research and*

data reporting. Adherence to these standards will allow metadata to be shared and to be searchable between different databases.

Methodological documentation: Documentation for this project will include the formation of written methodologies for sample collection and processing, made openly available on our respective lab websites as formal Standard Operating Procedures (SOPs), and published as peer-reviewed methodologies where applicable. Quality control will be conducted at each stage of the data acquisition, processing and analyses, including the development of metadata forms detailing the outline of the project, instrumentation used, format of data, QA/QC standards and controls, and funding source amongst other details. All participants are responsible for data collection, quality control, internal database management/curation, and data publication as applicable to their research responsibilities within the project. The graduate and undergraduate students will be trained in and involved in data collection and quality control across the process from collection to publication.

Policies for Data sharing and Public Access: The PIs of this project all have strong track records in data sharing and accessibility, including extensive experience in the use of open-access repositories for data, methods, code (i.e. GitHub), and open access publications. All of the data generated in this study will be made publicly available upon publication in a peer-reviewed journal and/or via online repositories within 2 years of the completion of the project. In addition to publication in peer-reviewed journals, with all relevant attempts to provide original datasets in open access publication repositories, the data generated from this project will be made publicly available online wherever possible. To ensure accuracy and data tracking, the project will have a specific data use policy including:

- User requests require current and valid contact information that will be used by the PI for tracking and documenting data usage.
- Users are required to cite the project publications and acknowledge the NSF as the original funding source.
- PIs and project participants will conduct quality control on the primary data and ensure accuracy of the primary data to the best of their abilities.
- The PIs and project participants will not release any private or confidential information to the public, and in-house databases will be password protected.
- The PIs and project participants will retain intellectual property rights, except where explicitly required to be released for publication and documentation.

Presentations: Our results will be disseminated in presentations at scientific meetings. Data and products from this project will be used in courses at CSUN, UH, and CSULB and course syllabi will be posted on the PIs websites as well as on the project website. The URL for the project website, KNB, and BCO-DMO DOI for each sample / study will be provided in all publications generated by the proposed work.