DATA MANAGEMENT PLAN

Data and information management for this project will be performed in close collaboration with the Moorea Coral Reef (MCR) LTER program at UC Santa Barbara, which has agreed to provide data documentation as well as storage and web access services to this project (facilitated by involvement of MCR LTER PI Schmitt). The MCR LTER Information Management System (IMS) will facilitate the archival cataloging of our data for long term preservation, enable the discovery of our data, and enhance their suitability for synthesis by us and others.

The MCR Information Management System meets or exceeds all current LTER criteria for Information Management as set forth by the LTER Information Managers Committee and LTER Network Information Systems Advisory Committee. With respect to data documentation, metadata documents in EML 2.1 dynamically derived from Metabase will be produced and maintained for our data (see below), which will make them readily adaptable to future enhancements in metadata documentation. Data accessibility will be through the MCR LTER website (and BCO-DMO) which meets or exceeds the Guidelines for LTER Website Design and Content. The website uses hierarchical navigation to provide access to the data catalog, publications and research foci.

Types of Data, Samples, and Other Materials: Multiple data types will be generated by the PIs and participants in this project including: 1) Ecological data such as species abundance, diversity, and composition of fishes and invertebrates in experimental corals stored as jpg, csv, and txt; 2) Coral condition data including coral size and growth, lipid, protein and carbohydrate content, tissue biomass, *Symbiodinium* density, chlorophyll content stored as jpg, csv, and txt; 3) Photographs of reef habitat stored as jpg, and 4) Simulation data from modeling efforts including reproducible R scripts and data outputs stored as pdf, csv, and txt.

Data and Metadata Standards: Data packages will conform to the most recent version (V2.1.0) of "EML Best Practices for LTER Sites" (August 2011). Metadata features will include embedded or online links to methods and protocols, full temporal, spatial, and taxonomic coverage, keywords from the MCR vocabulary, the NBII thesaurus, and/or the LTER Controlled Vocabulary, and units registered in the LTER Unit Dictionary. All data tables will be subject to EML Congruency Checker. Beyond the required elements, some datasets may provide explicit indexing keys and table-joining keys to facilitate cross-dataset synthesis.

Policies for Data Access, Distribution, Re-Use, Re-Distribution, and Production of Derivatives: Our databases and model codes will be subjected to the same data access policy and data distribution schema as the MCR. MCR data use policy and data release policy conform to LTER Network policies and use "Type I versus Type II" terminology. With some exceptions that are consistent with LTER data policies, MCR data are Type I (publicly available). There is no delay in releasing MCR data to the public once Quality Assurance has been verified. In accordance with LTER policy, graduate student thesis data are archived and cataloged as Type II (not released until thesis publication). Physical samples will be made available upon request where not consumed by analyses.

Policies for access and sharing will include provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements. To ensure accuracy and data tracking, the project will have a specific data use and re-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.

- Users have the final responsibility for any errors in their external and secondary analyses, while the PI 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 PI and project participants will not release any private or confidential information to the public, and in-house databases will be password protected.
- The PI and project participants will retain intellectual property rights, except where explicitly released for publication and documentation.

Active Data and Sample Storage: UCSB's Marine Science Institute data storage and access will be provided through infrastructure and IT resources, including a web server, database server, and filesystem server with a Storage Area Network (SAN) supporting snapshots. The archived data are supported by purposeful redundancy in backup systems provides for disaster recovery with off-site copies stored in a separate location, and allows for file restoration from more frequent on-site backups. The entire data catalog inventory is cached annually as a DVD archive off campus. Upon acquisition from the facilities, data will be quality controlled and added to the project server. All data on the project server will be accessible by all persons involved in the project.

We will use computers at UCSB, UGA, and Tulane for primary storage, statistical analysis, and modeling, with backup locally and to Dropbox (for each of inter-campus file sharing). We will also keep, store, and copy all handwritten and electronic field notes. Both during fieldwork and once we have returned to our universities, we will back up all data files on the MCR, UGA, or Tulane servers to ensure against damage or loss of the laptop and/or external hard drive. All quantitative data will be permanently archived in fully documented MCR databases and other pertinent materials will have metadata documentation that also is accessible on the MCR website.

To ensure long term usefulness of project data we will work to provide metadata of the highest quality. Data reach the MCR Information Management office in different stages of maturity depending on their type. Quality control (QC) will be performed automatically within the database, flagging further QC for human inspection. The MCR Information Manager will work with us regarding naming conventions, guidance for file organization and format, and designation of space on the server to back up raw files remotely. All PIs will be responsible for ensuring data and model codes are added to the catalog in a timely fashion.

Data Preservation and Archival: Data will be archived permanently in the original data format and also in more common, non-proprietary formats (e.g., tiff, csv, txt, fasta, etc.) to facilitate future data usage. Ecological, physiological, physical, and modeling data generated by the research and associated metadata will be stored long term on the UCSB servers, and archived at the NSF funded Biological and Chemical Oceanography Data Management Office (BCO-DMO). All data sets, derived data products, and associated metadata will be submitted and archived to the indicated databases and repositories within 2 years of collection.

Participant Roles in Data Management: The PIs are responsible for supervising all data management in cooperation with the project participants. 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 postdoc, graduate students, and undergraduate students will be trained in and involved in data collection, quality control, data storage, and data archival process. Given that all data are open to all project participants and will be made publicly available, no data loss or confusion would occur from departure/transition of key personnel during or after the project period.