Data Management plan

1. Types of data

High-resolution digital images: This includes *in situ* images of pocilloporid colonies and quadrats used for analysis of colony size and abundance (~500 GB).

Genetic samples: Genetic samples (preserved tissue and extracted DNA) will be deposited into -80ºC and -20ºC chest freezers at Florida State University.

DNA sequences: Aligned forward and reverse DNA sequence data (several hundred bp) of the mitochondrial marker ORF for ~5,000 pocilloporid corals of different sizes around Moorea will be deposited in Dryad Digital Repository or Genomic Observatories Metadatabase (GeOMe).

Field data: The fieldwork will generate a large quantity of population-level data on the growth and performance of pocilloporid corals around Moorea. These data would take the form of field censuses for abundance, recruitment, growth rates, and survivorship of spatially-referenced individual colonies.

Computer code: This includes R code and model outputs from the Integral Projection Model portion of the study.

2. Standards for data and metadata format and content

This proposal will build on the strong history of sharing data through Biological & Chemical Oceanographic Data Management Office (BCO-DMO), Moorea Coral Reef Long-term Ecological Research (MCR-LTER), National Centre for Ecological Analysis and Synthesis (NCEAS), European Project on OCean Acidification (EPOCA), and United State Geological Survey (USGS; through the Powell Center). Through these programs, we have considerable experience in developing databases (with metadata) for sharing a diversity of data types including physical data, biological data, and images (photoquadrats).

Metadata associated with this research will include information on sites, quadrat locations, sample collection date, time, location, experimental treatments, and environmental variables. All raw sequence data will include necessary and detailed metadata, including a description of the sample, library, and sequencing method.

Modeling work will be done in R. We intend to follow the Best Practices for Scientific Computing[1]. This entails writing annotated code that other people can later understand, using a version control system, embedding documentation, and reviewing code with collaborators. R Code will be integrated with a version control system such as GitHub, and stored on GitHub. Files will also be backed up to university servers.

3. Policies for access to and sharing data

All digital images and data will be stored permanently and backed up on a solid state hard drives that Burgess and Edmunds have at Florida State University and California State University

[1] Wilson G, Aruliah DA, Brown CT, Chue Hong NP, Davis M, Guy RT, et al. (2014) Best Practices for Scientific Computing. PLoS Biol 12(1): e1001745. doi:10.1371/journal.pbio.1001745

Northridge, respectively. Following NSF policies, we will make all project data, including metadata files, genetic sequence data, R code and model outputs, available on publically accessible servers within two years of data collection. The exceptions will be data related to graduate thesis projects, which will not be made available until 12 months following graduation. We intend to publish our code as software on a data archive such as Dryad, regardless of whether it is required by the journal.

4. Polices and provisions for re-use, redistribution, and production of derivatives

All users will have open and free access to our data within two years of collection, unless otherwise embargoed to meet the needs of graduate thesis preparation. Although not required for access, we will encourage all users to agree to acknowledgement access to our data and make contact with the lead PIs in the spirit of effective collaboration.

5. Archiving and access to data

Over the last 12 years, the Edmunds lab has had a strong and well-developed policy of sharing data through web-accessible systems. We have spent considerable time working with BCO-DMO to make data available from on-going awards, and are making data sets from published manuscripts available within months of publication; we are working to close this temporal gap by assigning DOI values as manuscripts are submitted. We already are ensuring that the raw data as well as data sets in support of specific publications are made available.

Additionally, we have worked to leverage the Information Management support inherent in the Moorea Coral Reef LTER (on which Edmunds is a co-PI) to support aspects of our data management from other projects located in Moorea and St. John. The objective of this effort is not to duplicate the products delivered by BCO-DMO, but rather to provide wider access to thematically cohesive data (for example, data from coral reefs in Moorea), and to leverage access and exposure to coral reef time series data at the MCR-LTER site (which manages data locally rather than at BCO-DMO as per NSF support). The objective of creating positive synergy between our efforts with BCO-DMO and MCR-LTER will be assisted by the hiring of our Independent IM Contractor (Hannah Ake) as a data manager at BCO-DMO. In this new role, she will continue to be based in Los Angeles and will work to facilitate data management of our pocilloporid. A key role of our local data management is to provide wider access to data and deliverables from our research that exceeds the services provided by BCO-DMO. We are particularly interested is serving project deliverables to end users with a diversity of skill levels including graduate students and resource managers.