Data Management plan

1. Types of data

Visual Surveys. The project will conduct visual surveys of the reefs at various locations using the CUREE robot. A subset of these images will be annotated for training neural networks. The collected images along with the annotations and training metadata, will be archived and also made easily accessible using public machine learning dataset repositories such as zenodo.org. Code will be hosted on public github repositories.

Ecological. The project will generate records of: (a) coral colony abundances and sizes for two species of coral (*Pseudodiploria strigosa a*nd *Dendrogyra cylindrus*) through analyses of legacy data (photoquadrats) and new surveys conducted by CUREE, (b) estimates of demographic rates (recruitment, growth, and mortality) obtained from the literature and through contrast of coral densities among years, and (c) projections of coral abundance over time and estimates of time to extinctions. These data will be managed as .csv files and full metadata will be archived with the standards we have developed for ecological data within the long-term ecological project from this site (see bco-dmo project 835192) and the MCR LTER project on which Edmunds is a co-PI (e.g., data set **knb-lter-mcr.4.39** in the MCR LTER catalogue).

2. Standards for data and metadata format and content

This proposal will build on our strong history of sharing data through BCO-DMO, Moorea Coral Reef LTER, and the Environmental Data Initiative (EDI). EDI offers advantages in this regard, notably in the user interface and the ability to access and reference data sets through unique DOIs. Through these efforts, we have considerable experience in developing relational databases (with metadata) that can be readily utilized by others.

Metadata associated with this research will include information on sampling locations, sample mode, and abundances.

3. Policies for access to and sharing data

All data will be stored permanently and backed up on solid-state hard drives and local servers at WHOI and CSUN. Following NSF policies, we will make all project data, including metadata files, data from field surveys, R code and model outputs, available on publicly 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 computer code 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

We all have maintained strong and well-developed policies of sharing data through web-accessible systems. We will deposit data into the BCO-DMO, to be made publicly accessible within two years. We will work towards assigning DOI values as manuscripts are submitted.