Data Management Plan

Products of Research: Data obtained from this project will include benthic cover of algae and invertebrates, coral distributions, fish diversity and biomass, and genetic data for *Montipora* (described below). Archived material not directly used for this project will likely be used to generate additional data in collaboration with others to produce new research findings. These data can also provide material for graduate students (one already planning to use it), as well as undergraduate projects. All remaining data will be archived and made available for collaboration. Genomic data from larval and adult corals are described under genetic data below.

Data Format, Storage, and Preservation: Newly produced data, as well as our pertinent baseline data gathered in previous years, will be saved in excel spreadsheets, as comma delimited DAT files. DAT files can be imported to and used with many database and software programs, making that a flexible format. All data will be stored using standards developed for widespread usage and will include date (month, day, yr., hour, sec), species/characteristic, and other relevant details. Data will initially be stored on a WB backup drive in Bernardi's lab and on UCSC's Department of Ecology and Evolutionary Biology secure server. Co-PIs Crane, Nelson and Paddack will be responsible for maintaining and preparing field data. For long-term storage, the data will archived with the Biological and Chemical Oceanography Data Management Office (BCO-DMO) (http://bcodmo.org/). All NSF's OCE sample and data sharing policies will be followed. The Data Management plan will also be published and available at BCO-DMO.

Genetic data**:** Genomic datasets generated by this project will be archived with a standardized meta-data table outlining original data generator(s), associated publications, and details about sampling sites, experimental history, etc. As the input file formats for each of the analyses are often different, they will also be linked to each raw dataset. PI Bernardi will be primarily responsible for creating and maintaining our genetic data storage. There is currently no universal standard format for storage of genetic data, but the Small Read Archive (SRA) at NCBI is the best option for persistent storage of raw Illumina reads.

Data Dissemination & Policies for Data Sharing and Public Access: All data will be made publicly available through the BCO-DMO site and directly upon request following publication and after any embargo period required for respective journal publications. Data and reports will also be available from the PIs, and through our project website which has a link to project resources at http://ulithimarineconservation.ucsc.edu/news/ After the project is completed all data will be made publicly available on the open access BCO-DMO repository immediately upon publication, or within 2 years of collection, whichever is sooner. All GenBank data will be publicly available as soon as we will have deposited them there, using that option while uploading them.

Specifically, benthic habitat data (quadrats and band transects) and fish abundance and biomass data (band transects) gathered during years 2012, 2013, and 2014 will be made publicly available through BCO-DMO. We will also make available DNA sequence data for *Montipora sp*.

We will then make publicly available through BCO-DMO newly acquired data on benthic habitat, fish abundance and biomass, as well as coral recruitment, and RAD sequence data of *Montipora* that we will obtain in 2015 and 2016.

Roles and Responsibilities: The data management will be the responsibility of PI Giacomo Bernardi who will work with BCO-DMO to ensure that the data submission and archiving plan is followed. Data will be made openly available to all collaborators and responsibility for the data will be transferred to BCO-DMO following completion of the project. Bernardi will also be responsible for the archiving and preserving of any unused sample material and making them available for collaboration. Bernardi will also be responsible for management, archiving and sharing of the genetic data.