Data Management- van Woesik 2020
Thermal stress and differential recovery of coral reefs

All data used in this study will be submitted and thereby managed by the Biological and Chemical Oceanography Data Management Office (BCO-DMO) at Woods Hole (http://bco-dmo.org/data/). For the active NSF grant, "Adjustment of western Pacific Ocean coral reefs to sea-level rise and ocean warming", award OCE-1657633 (finishing at the end of 2020), data are located at: https://www.bco-dmo.org/node/709534/. Additionally, the data for "Identifying coral reef 'bright spots' from the global 2015-2017 thermal-stress event", award OCE-1829393, is in the process of being deposited at: https://www.bco-dmo.org/award/762951. The bright spots project has dedicated considerable effort into generating a flexible and comprehensive coral-bleaching database. The new database for **Coral Recovery** will leverage what we learned from establishing the coral-bleaching database (see further information below).

The amount of data that will be collected by this current coral-recovery project will be considerable. Successful completion of objectives and comparisons relies on quality control and data management starting early and continuing throughout the program. Microsoft Access will be used to create the relational, Coral-recovery database. Using a comprehensive relational database provides a flexible platform. This database will allow external researchers to query specific information based on shared attributes, thus creating new spreadsheets ready for new analyses. The key to successful queries lies in the proper linkage of related data tables. Data for the current proposal will be also made available to BCO-DMO by late 2022, in the 2nd year of the project (which will end at the end of 2023). As in previous awards, van Woesik will make available all the data-analysis code (all the R code) by the end on 2023. There will be line-by-line annotations within the body of the code (for example see: https://github.com/InstituteForGlobalEcology). R. van Woesik has a policy that all newly written code that comes for his lab should be shared so that the field moves rapidly forward, but also that there is an archived record which is repeatable, annotated, and sharable (see also all the shared projects and R code at: https://github.com/rvanwoesik).

The project will employ a dedicated technician for two years to ensure that the data has gone through quality assurance, and are properly curated and accessible. The technician, working with the PIs, Post-doc and PhD student, will be responsible for the database and ensuring that is it fully functional and shared widely with the coral reef community. This person will also be instrumental in our outreach to the community of reef conservationists and managers.

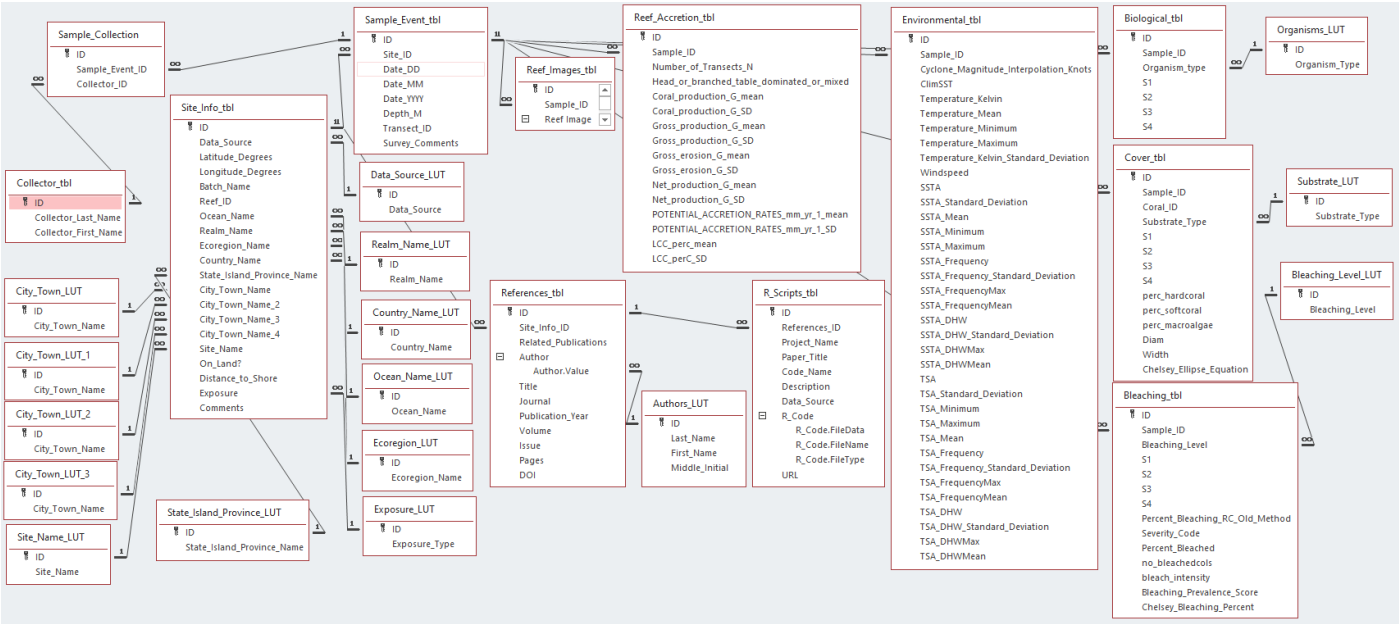

Fig. 1. The relationships of data tables from the coral bleaching database. The lines drawn between tables connect fields that are related.

Data in the coral-bleaching database are stored in several, related tables. The static location data (latitudinal and longitudinal coordinates, distance to shore, and exposure) are stored in the table "Site_Info_tbl," and the dynamic data are stored in separate tables. A site can have multiple sampling events (i.e. multiple depths and/or multiple dates sampled), so these data are stored separately in the table "Sample_Event_tbl." Data collected during these sampling events are stored in multiple related tables: reef accretion data, environmental data (CoRTAD, and cyclone data), biological data (organism types), coral cover data (percent hard coral cover), and bleaching data (percent bleaching). Published works and any R code available are also stored in tables connected to the sampling event. Tables with enumerated lists are used to ensure integrity in naming conventions. These tables are denoted with "LUT."