

## Data Management Plan

PIs, co-PIs and senior personnel from four institutions will be involved in this research, thus coordination and project management will be important for the success of the project. All members of our research team will be responsible for preserving data quality through standardized field and laboratory practices. Dr. Paul will coordinate the team's efforts to ensure accuracy of manual data entry, data file uploads and subsequent proofing. Drs. Paul, Aeby and Häse will also supervise data quality measures, file naming, data documentation and metadata creation. All team members will meet at the beginning of the project at the Smithsonian Marine Station, either in person or via video conference to plan and coordinate data management activities. We will interact regularly during the time that Drs. Aeby and Ushijima are in Florida to conduct field work, and then continue with monthly conference calls after that to ensure that all project activities are on schedule and all team members are kept informed.

According to Smithsonian, OSU, University of Florida and NSF guidelines, all project participants will undergo Responsible Conduct of Research (RCR) training. RCR Training is provided through the Collaborative Institutional Training Initiative (CITI) on-line, and it is free to Smithsonian, OSU and affiliated project personnel. The CITI RCR Training covers the following essential RCR topics: • Research Misconduct • Data Acquisition and Management • Responsible Authorship • Peer Review • Laboratory Animal Welfare • Human Subjects Protections • Mentoring • Conflicts of Interest • Collaborative Research

Data will be used for the generation of annual and final reports to NSF as well as peer-reviewed articles and presentations. Implementation, status, and any changes to original data management plan will be reported in annual and final reports to NSF. All project personnel will follow the Division of Ocean Sciences Sample and Data Policy and utilize the Biological and Chemical Oceanography Data Management Office (BCO-DMO) for all data management and data and sample storage for this project. Additionally, the Smithsonian IT and physical infrastructure will be used to support data management, data sharing and archiving through SIDORA. SIDORA is a community-based open-source project to create a freely available software environment that is designed to capture the full output from research projects, including the intellectual context that organizes the digital artifacts created in the research process.

1. *Types of data, samples, physical collections, and other materials to be produced in the course of the project:* This project will generate several kinds of data from various platforms (16S rRNA gene sequence data for microbial isolates and amplicon sequence data through Illumina Miseq sequencing, discrete sampling, and laboratory-based results). Matching datasets of microbial communities and bacterial isolates from healthy and diseased corals will be collected and all corals will be photographed in the field. Results of transmission experiments will also be maintained in spreadsheet format and digital photographs. Permits for collection of samples and coral fragments in the Florida Keys National Marine Sanctuary (FKNMS) will be obtained from the NOAA Florida Keys National Marine Sanctuary. A State of Florida Special Activity License is already available through our partnership with Florida Fish and Wildlife Conservation Commission.

2. *Standards to be used for data and metadata format and content:* The majority of our data will be created and stored as plain text files as they are easily portable between data processing software programs (e.g. Matlab, Excel, etc), can be read on any computer platform, and are suitable for uploading to web-based databases. Metadata that we need to support our field and laboratory observations include variables such as time and day of year, units of measure, accuracy and precision of measurements, methods of measurement and sampling, investigator, and data processing protocols. Metadata will be entered into a computer spreadsheet program and stored as ASCII text. We will use the metadata standards outlined in the document titled "Data Management and Guidelines Manual" produced by the

NSF-funded Biological and Chemical Oceanography Data Management Office (BCO-DMO). See <http://bcodmo.org/resources>. The metadata standards used here are community accepted standards that have been developed and implemented with support from the NSF by the BCO-DMO. Metadata will describe and document our datasets. All biological measures will be specified and units designated within metadata files. Metadata and data will also contain descriptions of the experimental manipulations that we impose and digital images of the sampled corals. All methods used for data collection, processing, QC/QA, and analysis will be described, as well as the title, creator, file format, file structure, geographic location, and timestamp. Access rights and access information regarding where the public can access data will be included as part of metadata. Excel files and appropriate project metadata and dataset metadata forms will be submitted to the BCO-DMO. The dataset from this project will not be copyrighted so there should be no limitations or restrictions on data sharing.

*3. Policies for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements:* Interpreted data products will be incorporated primarily into peer-reviewed journal articles, and in Supplemental Information to journal articles, but will also appear in conference presentations. In the case of journal articles, pdf versions of the article will be posted in full compliance with publisher. The Smithsonian has a fully accessible online repository for published data resulting from federal funding. All raw experimental and biological data and metadata from this project will be archived in the BCO-DMO and in SiDora. Data products will be shared according to established policies and archived for long term storage. All raw data and data products will be made available for public use according to NSF guidelines. Investigators intend to share with other researchers the primary data, coral and microbiological samples, and other supporting materials created or gathered in the course of this project. Access to the data will be free of charge and access to samples will only involve costs of shipping and handling. The quality-filtered sequencing reads for 16S rRNA gene sequence data will be submitted to NCBI's Sequence Read Archive and links to these datasets will be established in the BCO-DMO project information. Sanger sequencing of 16S rRNA genes used to identify bacterial isolates will be submitted to GenBank. The isolated organisms will be submitted to publicly available culture sample collections such as the American Tissue and Culture Collection or the National Center for Marine Algae and Microbiota at the Bigelow Laboratory for Marine Sciences. Culture collection accession numbers will be linked to the BCO-DMO project information. Cultures will also be maintained frozen in glycerol at -80 °C at the Smithsonian Marine Station.

*4. Policies and provisions for re-use, re-distribution, and the production of derivatives:* There will not be any permission restrictions placed on the data. The end-users of these data are likely to be resource managers or microbiologists interested in coral diseases, marine pathogens and related topics. There are no reasons not to share or re-use the data (i.e. ethical, non-disclosure, etc.).

*5. Plans for archiving data, samples, and other research products, and for preservation of access:* All members of our research team will be responsible for preserving data quality through standardized field and lab practices; however, Dr. Paul will coordinate the team's efforts to ensure accuracy of manual data entry, data file uploads and subsequent proofing. Drs. Paul, Aeby and Häse will also supervise data quality measures; file naming, data documentation and metadata creation. During the course of the project, all data will be fully backed up on the research data storage facilities at the Smithsonian, UF, and OSU through use of the automated network back up (incremental and full backups). Illumina sequencing data at UF is backed up by the sequencing facility, and Dr. Meyer has a subscription for data storage on UF's HiPerGator 2.0 high performance processors. As discussed above, the data will need to go through basic quality control steps before being uploaded, stored, and shared with the general public. Anonymization is not necessary for these data. We will be following metadata protocols outlined in the BCO-DMO Data Management Guidelines Manual to create appropriate metadata and documentation for our dataset.