

## Data Management Plan

This plan is presented to comply with NSF (GEO-OCE) policy and directions on the dissemination and sharing of research results as described in Grant Proposal Guide (PAPPG NSF 19-1) and in compliance with Data Management and Data Reporting requirements of NSF Division of Ocean Sciences Sample and Data Policy (NSF 17-037).

*Data Collection and Storage:* Metadata and complete data sets that will be produced during the project include microbial genomic data, images, and geochemical composition and flux data. Digital data, including imagery, sequence, and environmental data, generated during this project will be archived on the redundant and backed up Thurber lab computer system. Metadata and primary data, when appropriate, will be entered into the Biological & Chemical Oceanography Data Management Office (BCO-DMO) Data Coordination Center's Directory Interchange Format as required by the Program, and has been done for all NSF awards that Thurber has been awarded in the past. Proof of submission and (meta)data links will be provided in annual and final NSF project reports.

*Access and sharing of data:* We will employ different NSF-recognized repositories according to the nature of the data and disciplinary convention, ensuring public access to our research (meta)data within current policy guidelines. Appropriate data formats will be used to facilitate efficient communication with these repositories. All primary data will be made available upon quality assurance and within two years or publication, whichever comes first. All data sets will include metadata, calibration information (when appropriate) and sample site coordinates so that collected data may be geo-referenced. Seafloor images and video will be used for outreach purposes and made available for educational use on Thurber's website and through the outreach activities including a short documentary focused on the biogeochemical impacts of coral bleaching. Data will be disseminated through publications in scientific journals and presentations at scientific conferences. Genome sequence and 16S will be uploaded into the GenBank at the time of publication and full Amplicon results will be uploaded into NCBI – Short Read Archive data portal prior to the publication of scientific manuscripts resulting from this project. All other data will be ultimately submitted to the BCO-DMO.

All significant findings from the project will be promptly prepared and submitted for publication, with authorship that accurately reflects the contributions of those involved. We aim at publishing most, if not all, results in open access publications, whilst the data and metadata itself will be freely accessible through repositories as explained above. After publication, any additional data will be made available through the Thurber Lab's website including archival data servers through the Center for Genome Research and Biocomputing at Oregon State University. Additionally, all protocols developed and used during this study will be kept on file (and available through the above servers) so that all analyses can be replicated, and data analysis pipelines will be made available through GitHub, and also be supplied as supplementary material in publications.

*Example of Previous Data Archiving:* <https://www.bco-dmo.org/person/709617>