Data Management Plan

Data Collection and Storage: This project will generate frozen sediment samples, microbial genomic and transcriptomic data, seafloor images and video, 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, which includes an offside NAS system to facilitate automatic copies of all data and products as they are generated. Sediment samples collected for this study will be labeled using archival tape and then stored in the Thurber lab under proper conditions for long-term storage (e.g. -80°C for DNA samples and in DNA/RNA Shield and at -80°C for RNA) in a freezer that is remotely monitored to alert Thurber (and university facilities) if the freezer fails.

Access and sharing of data: Upon award, this project will be registered with the USAP Data Coordination Center so it can be registered in the Antarctic Master Directory. Metadata and primary data will be provided to (or referenced in the case of genomic data) the Biological & Chemical Oceanography Data Management Office (BCO-DMO) in Directory Interchange Format as required by the Program, and has been done for all NSF awards that Thurber has been awarded in the past. All primary data will be made available within two years of collection, at the time of publication, or prior to the end of the grant, whichever comes first. Proof of submission and (meta)data links will be provided in annual and final NSF project reports.

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. Genome and transcriptome sequence and 16S rRNA Sequence will be uploaded into the NCBI – Short Read Archive (SRA) and extensive metadata will be included in the submission to the SRA to allow the greatest utility of those data to other researchers. All other data will be ultimately submitted to the BCO-DMO. Data will be disseminated through publications in scientific journals and presentations at scientific conferences.

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. All protocols developed and used (including bioinformatic pipelines) during this study will be included as supplementary material (when allowable by the journal) and through GitHub so that all analyses can be replicated.