

Data Management Plan

We will gather many samples and a broad array of geochemical, geophysical, and microbiological data in the course of the proposed research. While each category of samples or data will be of interest in its own right, their true power will lie in the fact that the datasets are closely related to each other and share a coherent sampling and analytical plan. Consequently, it is important that the core material, generated data, and metadata be available to future researchers who may be interested in the material and/or data from a variety of perspectives. To ensure this availability, the proponents will meet all NSF reporting requirements.

Sample archives. Solid material, in the form of cores, sample splits, and -80°C frozen samples, will be curated and retained after the expedition and made available to other investigators that wish to use them for other means. The cores will be curated and archived at the NSF-funded URI Marine Geological Samples Laboratory (MGSL). Within a year of their arrival at the MGSL, they will be made available for sampling to outside investigators. The MGSL will deposit all metadata for the cores, including location, water depth and sample type, will be included in the Index to Marine Geological and Lacustrine Samples (IMGLS) database operated by NOAA (<https://www.ngdc.noaa.gov/geosamples/index.jsp>). Samples dedicated for microbial community investigations (sediment and water) will be stored in archives of -80°C samples, curated, and maintained by the PIs at URI for post-cruise access to the broader community. Most of the core material will be intact and other researchers will be able to resample it directly. The PIs have a long record of supplying unused splits and frozen samples to other researchers through the years.

Data standards. To ensure broad access, we will follow the metadata guides and references produced by the Marine Metadata Interoperability Project. All the omics data and metadata will conform to the current standard set by the community, such as standards set by the National Microbiome Data Collaborative (NMDC). We will adapt FAIR (Findable, Accessible, Interoperable, and Reusable) Principles to manage the 16S amplicon sequencing, metagenomics, and metatranscriptomics data. In reporting analytical metadata, we will follow the principles articulated by Goldstein *et al.* (2003). Calibration standards, standard reference materials, internal standards used to quantify analytical precision, and other items analyzed as part of good analytical practice will be fully described in relevant publications.

Release of protocols and workflows for molecular preparation and data analysis. Protocols for DNA/RNA extraction, 16S rRNA gene amplification, and sequencing library preparation will be publicly released to *protocols.io* (<https://www.protocols.io/>). Analysis of the sequencing data will follow standardized pipelines using open-source software. We will use the URI high-performance computing (HPC) facility to perform the molecular sequencing data analysis prior to deposition in public libraries. The Zhang laboratory owns a 150TB storage system on the HPC system that can be directly accessed by the regular and high-memory computing nodes to be used for the proposed research. Additionally, the HPC system provides a 655TB scratch storage system that is shared among all researchers at the URI community. New computational programs, whenever developed for this project, will be released as open-source repositories at GitHub. Analytical pipeline and computational workflows will also be publicly released on *protocols.io*.

Data availability and data release. We are committed to ensuring widespread data availability and utility. All published data, as well as related unpublished data (where appropriate), will be archived in appropriate internationally accessible databases, Shipboard survey results, including all underway data, will be submitted to the ship operator and appropriate national databases (e.g., BCO-DMO, NCEI or MGDS) within a year of cruise completion. Significant research findings will be submitted for publication beginning in the second year of the funded performance period. All raw sequencing reads data will be deposited into the NCBI Sequence Read Archive (SRA, <https://www.ncbi.nlm.nih.gov/sra>), co-assembled metagenomic contigs will be deposited to the NCBI Whole Genome Shotgun (WGS) submissions, and co-assembled metatranscriptomes will be deposited to the NCBI Transcriptome Shotgun Assembly (TSA) database. Functional

annotations of the metagenomes and metatranscriptomes will be deposited to the Integrated Microbial Genomes & Microbiomes data management system (IMG/M, <https://img.jgi.doe.gov/m/>), which will be linked to the deposit of metadata in the Genomes OnLine Database (GOLD). Other data will be deposited at the Biological and Chemical Oceanography Data Management Office (BCO-DMO), the National Center for Environmental Information (NCEI), or the NSF-supported Earthchem database (<http://www.earthchem.org>), as appropriate. As generally dictated by journal policies and data volume, data associated with specific publications will be placed in publisher-hosted archives as supplementary online materials.

As evidence of our past data-sharing practices, we note the following: (i) we have routinely submitted shipboard data from our past expeditions to the ship operator(s) and appropriate national databases within a year of expedition completion, (ii) we have routinely submitted the full spectrum of laboratory data (geochemical data, sequencing data, etc.) to appropriate international databases at or before the time of publication in peer-reviewed journals, (iii) the cores from our previous piston-coring expeditions are curated, archived and available for sampling at the NSF-funded URI Rock and Core Repository, (iv) the cores from our previous drilling expeditions are curated, archived and available for sampling on request at the IODP Core Repository (College Station, TX), and (v) -80°C microbiological samples (sediment and water) are also archived and curated at either URI (piston-coring expeditions) or Texas A&M University (drilling expeditions) for post-cruise access by the broader community.