

DATA MANAGEMENT PLAN

1. *Overview of data produced by this project:* Three types of data will be generated during this proposed work: 1) nucleic acid sequence information (genomic, transcriptomic, TN-seq, and metagenomic), 2) sediment geochemical data, including voltammetry data, and 3) data on microbial physiology of various isolates (including growth and bioelectrochemical datasets). Genetic mutants and an array of plasmids for different microbial strains will be constructed as part of this work and made publicly available. Protocols will also likely be generated in this work and will be shared on Protocols.io.
2. *Standards used for data types anticipated:* All samples including cultures generated during this work will be stored appropriately (i.e., -20 to -80°C) before molecular analyses. Field collected data will be appropriately documented in weather proof lab notebooks and electronic copies will be generated for analysis in the lab. All lab generated data will be appropriately documented in personnel lab notebooks with references to dates, protocols and saved raw data files. Raw data files are backed up on all lab computers using the lab's network storage device (NAS), cloud based storage through UC (OneDrive) and saved on individual work stations. Computed data will be generated and stored using Rowe lab computing resources (NAS, and OneDrive available through UC). All electrochemical data will be stored similarly, with raw (i.e., plain text, csv) maintained in triplicate (as noted above, lab computer, Network storage and UC OneDrive). Processed data (xml) will also be stored along with the appropriate metadata using lab resources. Molecular data will be stored redundantly on Rowe lab computing resources (NAS) as well as on the Ohio Super Computer (OSC) that will be used by the Rowe lab for both data analysis and storage of biological sequence data.
3. *Management of data collection:* Dr. Annette Rowe will be responsible for handling, storing, and uploading nucleic acid information to public databases and linking NCBI data with Biological and Chemical Oceanography Data Management Office portal. All preliminary data for this work is stored in the project "Uncovering novel mechanisms of extracellular electron uptake in subsurface-relevant marine bacterial isolates" which is part of the C-DEBI program. She will also be responsible for archiving field collected geochemical data and uploading information to a new project in the BCO-DMO Data Portal. Data collection and storage will also be performed by the Postdoc and Graduate student funded by this project as advised by the PI.
4. *Dissemination of Data:* Molecular datasets will be permanently archived on the NCBI Genbank database under a unique Bioproject and Biosample number for each microbial strain, and/or field expedition. Raw data for short read sequences, used for genome assemble, transcription analysis, high-throughput genetic screens (Tn-seq) along with long read sequences from the MinION will be uploaded to the sequence read archive (SRA). Assembled genomes, and metagenomes will be uploaded individually and annotated through the NCBI framework before linking to Bioproject resources. The molecular information will be linked to all geochemical information and metadata using the BCO-DMO data portal. Key portions of the data generated from this project will be shared at the time of reporting via supplementary material associated with publications. These will be the most pertinent data for most users of our results. We will make our data available to any others upon request by email or to our institutions, and we will note this in a statement to appear in all of our publications. We will perform the transfer of files using cloud-based file-transfer methods whenever possible. If we are unable to fulfill

a request through direct electronic transfer of files, then we will copy the requested data to a solid- state storage device and deliver it by postal mail (with the costs of shipping and the storage device to be paid by the requestor)

5. *Policies for data sharing:* Post publication, all data generated in this work will become public domain and available upon request. All Nucleic acid data will be made available in the form of raw data (Sequence Read Archive, NCBI) and genome/metagenome assemblies (GenBank, NCBI), however specific data files not publicly archived, will be made available upon request and/or shared through cloud-based resources. Characterized cultures will be deposited to international culture collection or made available upon request. Similarly, modified strains will be available upon request post publication.
6. *Long term preservation of data:* Data files will be stored on a cloud-based archive (OneDrive) and redundantly stored on the Rowe lab network assisted storage (NAS). These resources will ensure availability of this data for the foreseeable future. Access to data will be shared amongst other researcher upon request, but much of the data will be made publicly available on archived databases post publication of research result or at the end of the granting period (NCBI, BCO-DMO etc.). Long term storage (>10 years) of data will be done utilizing the Rowe lab NAS and/or UC computing resources.