

## Data Management Plan

### 1. The types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project.

- (a) Physical and chemical data. Time, date, location, depth, salinity, temperature, and general nutrients have been collected for every sample in the field. Physical data has been jointly recorded by all members of the HADES expeditions and the remaining data (general nutrients) will be shared when available among all PIs.
- (b) Lander video and still imagery. Video and image data from lander deployments from the Kermadec and Sirena Deep cruises were collected by HADES collaborators at the University of Aberdeen and the Schmidt Ocean Institute. Video and still images were collected on the Mariana Trench Challenger Deep cruise run by the PI of this proposal.
- (c) Viable counts, microbial strains and enrichments. Viable counts of microbes incubated under differential conditions will be collected, and subsequently unique culturable isolated microbial strains will be collected and identified.
- (d) Sequence data. Sequence data will be generated during 16S rRNA identification of isolated strains, genomic characterization of isolates and single cells, and V3-V4 analysis of microbial communities.
- (e) Flow cytometric and microscopic data. Viral and microbial counts using flow cytometry and images identifying active microbial cells via epifluorescence microscopy will be generated.
- (f) Experimental data and protocols. Experimental protocols that may be specific to these samples (e.g. working with specific microbes at high pressure, identifying and/or sorting active cells, collecting samples under *in situ* pressure conditions) may be modified during the course of this proposal.

### 2. The standards to be used for data and metadata format and content (where existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies).

All genome sequence submissions will adhere to the standards promoted by the Genomic Standards Consortium. Sequence data that will be available for future analyses on publicly available sites, such as GenBank, the Integrated Microbial Genomes (IMG), and Rapid Annotation using Subsystem Technology (RAST) databases will follow the appropriate guidelines and provide contextual and experimental metadata upon submission. Microbial strains deposited in culture collections will follow the appropriate preservation procedures depending on the culture collection. Confirmation of the correct strain by the laboratory after being submitted and cultured by the culture collection will be conducted as appropriate. Biological materials will be shipped according to UN transportation guidelines.

### 3. Policies for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements.

- (a) Lander video/still imagery. Highlights of the video data of the lander deployments from the Kermadec and Sirena Deep cruises have been already made available through public outreach on Youtube through the University of Aberdeen Oceanlab channel (<https://www.youtube.com/channel/UcamJAVJXDDLgzdTmTLRkFZQ>). The video collected on the Mariana Trench Challenger Deep cruise run by the PI of this proposal will be shared on the Scripps Institution of Oceanography Youtube channel (<https://www.youtube.com/user/scrippsceanography>) and the PIs laboratory website when available. Microscopic images generated as described in 1(e) will be available in publications or upon request.
- (b) New microbial strains. New microbial strains that have complete taxonomic descriptions will be deposited in

the American Type Culture Collection (ATCC), the Japan Collection of Microorganisms (JCM), and the German Collection of Microorganisms and Cell Cultures (DSMZ). Data from taxonomic characterization will be shared as a Google document among members of the lab. All strains will be available upon request and those that do not have high pressure cultivation equipment may visit the PIs laboratory for sample preparation.

(c) Sequence data. All phylogenetic (full length 16S rRNA genes and I-tag V3-V4 sequences) and genomic data (single cell and pure culture) will be deposited in GenBank. Single cell and isolate genomic data will be publicly available on the IMG Expert Review (IMG/ER) and the RAST databases for future comparative analyses. Metagenomic data will be publicly available on the IMG Metagenome Expert Review (IMG/MER) and Metagenomics-RAST (MG-RAST) for future work. All raw and unassembled data will be available on the PIs laboratory website.

(d) Experimental protocols and data. Experimental protocols that may be specific to these samples (e.g. working with specific microbes at high pressure, identifying and/or sorting active cells, collecting samples under *in situ* pressure conditions) will be integrated into publications and be available upon request. Key figures, images, and videos will be uploaded to the websites of the PI and to public access sites such as FigShare and ResearchGate to increase visibility of the work.

#### 4. Policies and provisions for re-use, re-distribution, and the production of derivatives

The findings from this study will be published in peer-reviewed journals. Publication in open access journals will be done when appropriate to help disseminate the data to a larger audience. Public lectures and presentations by both the PI and the graduate student, such as at the American Society for Microbiology, American Geophysical Union, American Society for Limnology and Oceanography, and Deep-Sea Biology conferences, will further help in dissemination. Sequence data will be archived at Genbank, RAST, and/or IMG as described in 3(c). Images and video will be shared upon collection as appropriate. Supplementary data will be provided in coordination with published manuscripts and will be accessed on the publisher's website or will be distributed by the PI upon request. Data will be shared among other members of the HADES program when necessary to facilitate integration of data into a larger understanding of trench ecology. All data will be made available at the time of publication or at the end of the grant period consistent with NSF policy.

#### 5. Plans for archiving data, samples, and other research products, and for preservation of access to them

(a) Storage of strains, cultures, and other physical materials. All active samples and cultures will be maintained in a cold room at 4°C. A backup generator is employed in case of failure. New microbial strains isolated from this work will be cryopreserved as duplicate stocks in 15% glycerol and/or 11% dimethyl sulfoxide in an ultralow -80°C freezer. In cases where detailed analyses have been performed, samples will be deposited in open culture collections and made available upon request, as described in 3(b). All extracted and amplified DNA and filters used for microscopy will be archived and stored at either -80°C or -20°C indefinitely.

(b) Sequence data. Sequence data will be principally stored on external hard drives and at the San Diego Super Computing Center (SDSC) via the Triton Shared Computer Cluster (TSCC). Smaller sized experimental data will be stored as Google documents. Sequence data will be additionally preserved on publicly available databases and publicly available websites as previously described in 3(c).

(c) Experimental data. This includes viable counts, microbial growth characterization, direct viral and microbial counts, activity measurements, and computational analyses including phylogenetic and taxonomic data. This data will be saved on a laboratory computer and made available for all team members using Google documents, including other HADES team members that will integrate this data into their own analyses and publications. Access will be provided via the PI's laboratory website.