

## - Data Management Plan -

### Tracking of Multiple Sample Types

This project will produce a comprehensive set of interdisciplinary data associated with every sample including time stamps, lat/long/depth coordinates, digital still and video imagery, basic field data from sensors mounted on the ROV *Jason*, aqueous chemistry of fluids and volatiles, radiocarbon and <sup>13</sup>C isotopic signatures of several inorganic and organic carbon species, carbonate mineral and rock geochemistry, carbon transformation rate measurements from *in situ* incubation experiments, Raman spectroscopy of single cells, and DNA/RNA sequences from communities and single cells. The coordination of this broad array of data begins with rigorous sample identification and tracking. Each sample will be logged and given an identification number (incorporating the ROV *Jason* dive number) upon collection on the seafloor, and each subsample processed shipboard will receive a derivative identification number that is uniform among all investigators. Rock, fluid and biological samples will be registered with SESAR (System for Earth Sample Registration - <http://www.geosamples.org/>). All data associated with each subsample will be tracked and managed in a shared MySQL database hosted on the Brazelton lab server and maintained by a full-time technician (Christopher Thornton) whose salary is funded by Brazelton's laboratory start-up budget. This open, query-supported database will allow each member of the team to monitor each subsample: its location and status, any information and data currently available, the name of the responsible team member, and an agenda of additional planned analyses.

### Data Formats and Standards

The contextual environmental and chemical data collected for each biological sample will far exceed the minimal requirements established by the Genomic Standards Consortium as the Minimal Information about a Metagenomic Sequence (MIMS) and the Minimal Information about a Marker Sequence (MIMARKS). The MIMS and MIMARKS formats will be used to report all contextual data associated with any set of DNA or RNA sequence data submitted to the databases described below. Furthermore, all sensor-generated data and routine sample-based data will be contributed to the Marine Geoscience Data System (MGDS) and EarthChem immediately following the cruise according to their specified formats. All shipboard data will be archived in the R2R NSF-UNOLS supported repository, and E-Log will be utilized to capture metadata associated with all over the side operations.

### Access to Samples and Data

Immediate public access to DNA and RNA sequence data will be provided by submission to an array of public databases: VAMPS ([vamps.mbl.edu](http://vamps.mbl.edu)) for 16S rRNA data, MG-RAST ([metagenomics.anl.gov](http://metagenomics.anl.gov)) and IMG-M ([img.jgi.doe.gov](http://img.jgi.doe.gov)) for metagenomic and metatranscriptomic data, and IMG-M for single-cell genomes. Also, all sequence data will be submitted to the appropriate NCBI databases for public access. Public database submission will occur immediately after quality filtering of the sequence data is completed (< 1 month after data generation). Processed data (e.g. 16S rRNA diversity analyses and genomic assemblies) will be submitted to the relevant databases above prior to manuscript submission. All chemical data and other contextual information (in MIMS and MIMARKS format) associated with each DNA/RNA sequence dataset will be included at the time of database submission with the exception of a few sensitive data products which will be made publically available after publication of the relevant manuscript or else not more than two years after data generation. Still and video images will be made available via the Biological and Chemical Oceanography Data Management Office (BCO-DMO). Complete sample inventories will also be available on BCO-DMO, and requests by other researchers to perform complementary analyses of samples will be considered immediately following the cruise.

### Reuse and Redistribution

Our team has a philosophy of open access and will strive to publish in freely available peer-reviewed journals or journals that provide an open access option. Furthermore, prior to publication, we will post all routine data, metadata, logs, and raw files of still and video imagery to the databases described above, which may be freely used by the public (with attribution). As described above, DNA sequence data will be available in public databases that allow free reuse, redistribution, and unlimited production of derivatives (with attribution).

### Archives for Data and Samples

One of the volunteer berths on the ship will be filled by Christopher Thornton, an IT technician from the Brazelton lab, who will supervise (in collaboration with the ROV *Jason* team) data storage, back-up, and archiving during and after the cruise. All photos, videos, sample logs, and dive reports will be collected on terabyte drives, distributed to all PIs, and stored on backed-up servers in the Brazelton (UU) and Kelley (UW) labs. Also, all imagery and sensor data collected by ROV *Jason* will be archived at the National Deep Submergence Facility at Woods Hole (NDSF). Real-time logging data compiled by the Jason Virtual Van will be made available immediately after it is archived post-cruise at NDSF with open access. In addition to the video materials published by the digital media workshop participants described in Broader Impacts, other highlights and key imagery to support science publications will be stored on the Lost City web site server maintained and backed up by the University of Washington Center for Environmental Visualization (<http://www.cev.washington.edu/>).

Fluid samples (in the form of frozen water, 0.2 um filters, and glycerol-preserved cell suspensions), basement rocks, and carbonate chimney samples that were frozen immediately at sea for microbiological and geochemical analyses will be stored by Brazelton and Lang at their respective laboratories in -80°C and -20°C freezers. The freezers are equipped with alarm systems and emergency back-up power. Rock and chimney samples intended for petrological studies will be stored in sterilized foil with splits distributed between ETH-Zürich and UW. Ampoules for gas extraction will be stored in Lilley's laboratory (UW).