BCO-DMO NSF OCE: Biological and Chemical Oceanography

This project will comply with the NSF OCE Data and Sample Policy

Description of Data Types

(1) **Observational**: Chemical and metadata from Expedition 360: Time of collection, date, location, depth, potential temperature (from borehole logging), lithology (from shipboard core descriptions), total carbon and total organic carbon, total nitrogen, δ13C and δ15N (from CHNS analyses done post cruise), are available for all 71 rock core samples to be examined in this project. To maximize access to these data, data will be stored in BCO-DMO and in PANGAEA (http://www.pangaea.de/about/). PANGAEA is a member of the ICSU World Data System, created by the International Council for Science, is jointly hosted by the Alfred Wegener Institute and the University of Bremen, and is established as a repository for the long-term availability and accessibility of archived data and metadata in secure and machine readable formats. These data will also be uploaded to Interdisciplinary Earth Data Alliance (IEDA) at LDEO: http://www.ldeo.columbia.edu/research/marine-geology-geophysics/integrated-earth-data-applications.

(2) **Experimental**:

a. Enzyme activity data: ATP activity measurements were collected on a Turner Design TD-20/20 Luminometer and are available for every sample. Exoenzyme (alkaline phosphatase, leucine aminopeptidase, arginine aminopeptidase) activity rates were collected using a Tecan Spark 10M multimode microplate reader and are available for 17 samples. These data were collected during Expedition 360 or immediately following. Data will be deposited as for (1).

b. Sequence data: MiSeq PE 300bp iTAG data will be generated for all 3 domains of life by end of Yr 1 from all samples in our collection. These data and marker gene data for cultured strains will be deposited to GenBank and made publicly available by the end of the project period. Draft genome data for cultured strains and metagenome data (Illumina NextSeq) for selected samples will be generated by end of Yr. 2 and deposited to IMG Metagenome Expert Review (IMG/MER) and Metagenomics-RAST (MG-RAST), and transcriptome data (Illumina NextSeq) for selected strains and metatranscriptome data for selected samples will be generated in years 2 and 3 and will be archived on the NCBI –Gene Expression Omnibus (GEO) database and made publicly available by the end of the project period. Links to all data will be provided through BCO-DMO.

c. Lipid biomarker data will be collected using triple quadrupole mass spectrometry and confocal Raman spectra for unique cultures and selected rock samples will be collected using a Horiba LabRam HR confocal Raman system. These data will be uploaded to IEDA. IEDA is a community-based, NSF-funded facility that provides data services for the Ocean, Earth, and Polar Sciences to support, sustain, and advance the geosciences. IEDA data systems serve as primary community data collections for global geochemistry and marine Geoscience research, supporting the preservation, discovery, retrieval, and analysis of observational field and analytical data types, enabling data to be discovered and reused by a diverse community now and in the future. Links to data will be provided through BCO-DMO.

d. Cultures of environmentally relevant microbial strains: Microbial strains that have complete taxonomic descriptions and that have been fully characterized will be deposited in the American Type Culture Collection (ATCC) and the DSMZ (Deutsche Sammlung von Mikroorganismen und Zellkulturen). All strains, including the above will be cryopreserved in 15% glycerol stocks and/or 5% dimethyl sulfoxide in -80°C freezers in the PI laboratories

(on emergency power backup) and will be available upon request.

e. Experimental protocols: DNA, RNA and lipid extraction protocols, protocols specific to working with our cultures under pressure, and confocal Raman spectral analysis protocols will be made publicly available through protocols.io by the end of the project period.

f. Educational outreach materials: Lesson plans produced on subsurface microbiology will be posted on the JOIDES Resolution education and outreach website (http://joidesresolution.org/node/3002) and on Amazon Inspire, the new education resource for teachers dedicated to the free search, discovery and sharing of digital education resources (http://www.amazoninspire.com).

g. All data will be shared through peer reviewed publications

Data and Metadata Formats and Standards

(1) Observational data will be submitted as spreadsheet files (.xlsx), (2) Genetic data formatting will follow standards of NCBI, IMG, and MG-RAST and links will be provided through BCO-DMO, (3) Lipid biomarker data as spreadsheet files (.xlsx), (4) Raman spectra as plain text files, (5) Enzyme data will be submitted to BCO-DMO as spreadsheet files (.xlsx).

Data Storage and Access During the Project

All data will be shared electronically among all collaborators by posting as data are generated to a dedicated project page on Basecamp (https://basecamp.com). All experimental data generated at WHOI will be stored on our laboratory computers and associated Network Accessed Storage devices that are backed up daily by WHOI's automated backup service. All experimental data generated at TAMU will be stored on laboratory computers and in Google Drive, for which TAMU faculty have unlimited file storage. Sylvan maintains data storage currently for all files/data generated in his lab on Google Drive. Data generated in collaborator Schubotz' laboratory (lipid biomarker profiles) will be stored on local laboratory computers connected to automated backup devices, and will be shared immediately via our Basecamp website, and downloaded to WHOI computers for backup on the automated system.

Mechanisms and Policies for Access, Sharing, Re-Use, and Re-Distribution

Our project data will be publicly available by the end of the project period through NCBI GenBank/ Gene Expression Omnibus (GEO) database, BCO-DMO, IEDA, MG-RAST, and IMG. Cultures will be publicly available by request to the PIs. Our data will not involve security concerns, nor does access need to restricted in any way. The IODP Expedition 360 data embargo period will expire on 30 January 2017, before the start of this project.

Data Sharing via BCO-DMO and Archiving

BCO-DMO staff will work with us to manage our project data, and data generated during the proposed research project will be contributed to the BCO-DMO system. Links to data repositories for our genomic, lipid biomarker, and Raman spectral data will be provided through BCO-DMO (see above). After data contributed to BCO-DMO are online and fully documented, BCO-DMO ensures that the data are archived properly at the appropriate National Data Center (e.g. NODC) for long-term archive preservation.

Roles and Responsibilities Lead PI Edgcomb will ensure compliance with this Data Management Plan. She will be responsible for deposition of all genetic data generated at WHOI. TAMU PI Sylvan will be responsible for deposition of all genetic data generated at TAMU and ATP and exoenzyme data generated during Expedition 360. Co-I Klein will be responsible for depositing Raman spectral data and lipid biomarker data. Edgcomb and Collaborator Burgaud will be responsible for archiving fungal cultures, and Sylvan will be responsible for archiving bacterial and archaeal cultures.