BCO-DMO NSF OCE: Biological and Chemical Oceanography National Science Foundation

---

Data Policy Compliance

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

Description of Data Types

(1) Observational: Water column physicochemical data (time of collection, date, location, depth, temperature, salinity, optical transmissivity, dissolved oxygen, nitrate, nitrite, and ammonia concentrations, phosphate, particulate organic carbon, nitrogen and phosphorus, dissolved organic carbon, nitrogen and phosphorus, hydrogen sulfide, and methane) will be deposited in BCO-DMO and in PANGAEA (http://www.pangaea.de/about/) at the end of Y2 by Co-PI Taylor. 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. The cruise data will be deposited to R2R, Rolling deck To Repository (https://www.rvdata.us) by PI Pachiadaki.

(2) Experimental:

a. ***In situ* and onboard rate measurements** in unfiltered and pre-filtered ("particle-free") water samples. All calculated rate measurements will be deposited as for (1) by Co-PI Taylor.

b. **Metatranscriptomic data**. Metatranscriptomes will be produced from water samples filtered and preserved *in situ* corresponding to the same depths where the *in situ* rate measurements will be performed. The raw sequences will be deposited to GenBank and MG-RAST, while the assembled data to will be uploaded in IMG Metagenome Expert Review (IMG/MER). Data made publicly available by the end of the project period or upon publication (which ever happens first) by the postdoc. Links to data and associated water column data will be provided through BCO-DMO.

c. **Single cell genomes**. The generated genomes will be deposited to IMG/MER and to GenBank, and made publicly available by the end of the project period or upon publication by the postdoc. The genomes will be incorporated to publicly-available genomic databases, MiGA and MarDB. Links to data and associated water column data will be provided through BCO-DMO.

d. **Fluorescence in situ hybridization (FISH) counts** for key prokaryotic groups and unicellular eukaryotes. These data will include counts for all probe targets for all water samples collected. In addition, we will count total cells on DAPI stained filters for all samples. Data will be deposited as for (1) by Co-PI Edgcomb.

e. **SIP-Raman-FISH data.** The single cell activity data will be deposited as for (1) by Co-PI Taylor.

f. **Experimental protocols**. RNA extraction protocols, protocols for FISH and SIP-Raman-FISH hybridization with all probes used in the study will be made publicly available through protocols.io by the end of YR2 by the postdoc (RNA extraction), Co-I Taylor (SIP-Raman-FISH) and Edgcomb (FISH).

g. **Data processing scripts**. All bioinformatic scripts used for the processing will be uploaded in github and will be made publicly available by the end of the project period by Pachiadaki.

h. All data will be shared through peer reviewed publications

Data and Metadata Formats and Standards

1. Observational data will be submitted as spreadsheet files.
2. Genetic data formatting will follow the standards of NCBI, IMG/MER, and MG-RAST and links to these data will be provided through BCO-DMO.
3. Cell counts and single cell activity data of key prokaryotic groups will be submitted as spreadsheet files.

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 folder on Google Drive. A Slack workspace will be generated in order to facilitate project discussion among the collaborators and file exchange.

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.

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, BCO-DMO, MG-RAST, and IMG/MER. Our data will not involve security concerns, nor does access need to restricted in any way.

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 and image 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

PI Pachiadaki will ensure compliance with this Data Management Plan. She will upload the scripts and the cruise data, and she will supervise the WHOI postdoc on the deposit of the genomic and transcriptomic data, as well as the protocols and the scripts. Co-PI Edgcomb will deposit the FISH data. Co-PI Taylor will deposit water sample physicochemical, the rate measurements and SIP-Raman-FISH data generated at Stonybrook and the collaborative labs.