Data Management Plan

This project will generate oceanographic data (CTD data) and biologically-focused data from water column samples collected in the field in coastal Denmark, during an NSF-sponsored cruise in the North Atlantic, and during German-funded cruises in the Pacific.

The following table summarizes the data to be collected and the form in which the data will be archived. We archive raw data (chromatograms, spreadsheets with calculations) as well as processed data on Sakai, a password-secured web platform maintained by UNC that can be used to post and share data among registered users. We will establish a separate folder for this project, to which members of the Arnosti lab will have access. This will enable us to keep our data set in a central, centrally-located and backed up location, and will ensure that all data remains up to date. We generally archive data according to experiment/cruise (i.e., separate folders for cruises and for incubation experiments). Due to data volume, the sequencing data will additionally be archived separately (see below).

Metadata from our cruises (CTD casts: T, S, O_2, chl fluorescence) will be made available immediately via BCO-DMO; the German CTD-related cruise data is typically uploaded on Pangaea. We will add additional data to BCO-DMO as soon as analyses are completed (enzyme activities; bacterial productivity; cell counts).

16S rRNA and 16S rDNA sequence documentation (Balmonte lab): sequence documentation will follow the standards defined as Minimal Information about a Metagenomic Sequence (MIMS) and Minimal Information about a Marker Sequence (MIMARKS), curated standard format layers for the acquisition and display of information associated with sample acquisition, processing, handling, sequencing, and analysis. These are community standards, agreed using consensus and updated where necessary by routine annual meetings of the Genomic Standards Consortium (www.gensc.org). In addition, these standards are recognized by the INSDC and reported by a keyword (GSC) for compliant sequences. We will rigorously adhere to both standards for sequencing data generated using this proposal.

Type of data	**Brief description of measurement/archive**
Extracellular enzymatic activities of water column communities, measured under atmospheric, in situ, and elevated pressure conditions.	Measured via change in substrate molecular weight (fluorescently-labeled (FLA-) polysaccharides); data recorded on laptop computers, exported as csv files, processed in R. For our planned development of *in situ* measurements, activities will be measured via a change in fluorescence signal, using fiber optics. For peptides and α/β-glucose, hydrolysis is measured via an increase in fluorescence in a plate reader; data are exported to a computer as a csv file. Data are processed via R scripts, which will be deposited on the lab GitHub site. Data will be archived on Sakai as soon as it is generated. Data will be deposited in BCO-DMO concurrently with annual reports.
Bacterial heterotrophic production	Measured via ^{3}H leucine incorporation. Data collected from scintillation counter, processed in Excel spreadsheet. Archived on Sakai as the data is generated. Data will be deposited in BCO-DMO concurrently with annual reports.
Microbial cell counts	Counted via flow cytometry; processed as Excel spreadsheet. Data will be deposited in BCO-DMO concurrently with annual reports.

16S rRNA gene & transcript sequences	Samples to be collected and frozen at -80 °C in the field, and processed in the Balmonte lab in preparation for DNA and RNA sequencing. Sequences will be stored at the Lehigh High Performance Computing cluster. Data will be submitted to NCBI's Sequence Read Archive (SRA) promptly after analysis is complete.
Selfish bacteria analysis; FISH counts	Samples incubated and collected in the field, stored at -20°C, processed at the MPI. Images and data stored on computers at the MPI, which are regularly backed up. Microscopy data will be submitted to Pangaea.

Data Availability: Basic data relating to the stations sampled aboard U.S. ships (CTD data) will be made available immediately in accordance with UNOLS policy**.** A similar policy applies to the R/V *Sonne* (German research vessel); data are submitted to Pangaea. Data collected as described in the table above will be deposited in BCO-DMO, either as acquired, or concurrently with submission of annual reports.

Data Archives

The Sakai website is maintained and backed up by the Information Technology Office at UNC. The GPC/HPLC data (chromatograms, export files) from hydrolysis of polysaccharide and plankton extracts is stored on the associated computer, backed up on an external hard drive, and uploaded to the secure Sakai website. The sequencing data will be stored at the Lehigh High Performance Computing cluster, and will be submitted to the NCBI's Sequence Read Archive prior to manuscript submission (https://www.ncbi.nlm.nih.gov/sra). Lab books relating to the project (log books from cruises, basic experimental data, etc.) are stored indefinitely in the Arnosti lab, and scanned backups are stored on Sakai.

Computers at the Max Planck Institute (MPI) for Marine Microbiology, including the computers used for microscopy, are centrally networked and backed up on a daily basis. Data (photos; cell counts) are permanently archived at the MPI, and the cell counts (from selfish bacteria; total cells; FISH) will be submitted to Pangaea.

Sample archives

Balmonte's lab will archive portions of filters from water column analyses. We anticipate that this material could be used in the future (e.g., to search for specific genes; conduct quantitative PCR on genes for various enzymes, etc). The sample archive could also be used for future analyses of microbial community composition using new approaches/new databases that will doubtless be developed over the next decade. We will continue this practice with the current project, storing all such samples at -80 °C in Balmonte's lab. We would be willing to share such materials on a collaborative basis upon request by colleagues who have a compelling need to use it. These -80 °C freezers are backed by Lehigh's emergency power generators.