Data Management Plan

This project will generate data from samples collected in the field during one NSF-sponsored cruise in the North Atlantic. We will additionally have data from incubation experiments (mesocosm experiments, including substrate addition experiments) and from investigations of archived samples, as well as data from cruises of opportunity.

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; UNC has contracted with Microsoft OneDrive (a password-secured web platform) 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 (and Andreas Teske) 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 next generation sequencing (NGS) 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. We will add additional chemical metadata to BCO-DMO (e.g. DOC; C/N; total carbohydrates) as soon as analyses are completed. Further data (enzyme activities; bacterial productivity; cell counts) will be added to BCO-DMO as it is processed and checked. Data from the incubation experiments will also be submitted to BCO-DMO. Beginning with Hoarfrost et al. (2017), we also deposit scripts used to process and analyze the data, and to generate figures, in github repositories.

NGS Sequence documentation (Teske 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; enzyme activities measured in incubation experiments.	Measured via change in substrate molecular weight (fluorescently-labeled (FLA-) polysaccharides and plankton extracts); increase in fluorescence (MCA-labeled peptide substrates); data recorded on laptop computers, exported as csv files, processed in R. Data will be archived on OneDrive as soon as it is generated. Data will be deposited in BCO-DMO after processing and checking.
Bacterial heterotrophic production	Measured via [3H]-leucine incorporation. Data collected from scintillation counter, processed in Excel spreadsheet. Archived on OneDrive as the data is generated. Data will be deposited in BCO-DMO after processing and checking.
Microbial cell counts	Direct counting via microscopy; processed as Excel spreadsheet; archived on OneDrive. Data will be deposited in BCO-DMO after processing and checking.

C/N analysis, DOC analysis, POC analysis	Computer-generated (from analytical instrument) output will be stored in master Excel sheet. Archived on OneDrive as the data is generated. Submitted to BCO-DMO as data is generated.
Peptides and total carbohydrates (dissolved and particulate)	Measured analytically (see Methods); data to master Excel spreadsheet. Data will be deposited in BCO-DMO after data is processed and checked.
NGS of microbial community composition	Samples to be collected and frozen at -80C in the field; processed in the Teske lab and analyzed in the facility maintained by UNC's Department of Genetics. Sequences will be stored on computers in the Teske and Arnosti labs. Data to be submitted to GenBank as manuscripts are submitted for review.

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**.** Data collected as described above will be deposited in BCO-DMO, either as acquired, or as manuscripts are submitted.

Data Archives
The OneDrive website is backed up by Microsoft. 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 OneDrive website. The NGS data is stored on computers in the Teske lab, and will be submitted to the NCBI's Short Read Archive prior to manuscript submission. Lab books relating to the project (log books from cruises, basic experimental data, information relating to chemical analyses such as total carbohydrate measurements, etc.) are stored indefinitely in the Arnosti lab. We also scan our lab books and store the PDFs on OneDrive as an additional backup. NGS sequences will be submitted to the NCBI GenBank short read archive: https://www.ncbi.nlm.nih.gov/genbank/submit

Sample archives
The policy for Arnosti/Teske lab collaborations over the last several years is to archive portions of filters (from water column analyses) and extracted DNA for future analysis. We anticipate that this material could be used to search for genes related to specific microbial functions. 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 one of the Teske lab's freezers. 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 UNC's emergency power generators.