**Data Management Plan**

This project will generate data from samples collected in the field during three cruises (two NSF-sponsored cruises, one cruise of the Alfred Wegener Institute). We will also obtain data from experiments conducted as part of other cruises, including the experiments conducted by Prof. Rudi Amann during the recent AMT22 cruise: (http://www.bodc.ac.uk/projects/uk/amt/), and experiments to be conducted during the upcoming DeepDOM cruise (Monteviedo, Uruguay to Bridgetown, Barbados; Dr. Liz Kujawinski, chief scientist). We will generate data (sample analysis will take place in the lab) from all of these cruises. The incubation experiments carried out aboard ship for our two NSF-sponsored cruises and the Alfred Wegener Institute sponsored cruise will also generate data on microbial community responses to carbon addition as well as data on carbon degradation from these experiments.

The following table summarizes the data to be collected and the form in which the data will be archived. For most of the data, we will establish an Excel master spreadsheet (managed by Arnosti) in which to track and archive data, arranged by cruise, location, and experiment. We will also archive the raw data (chromatograms, spreadsheets with calculations.) UNC has established Sakai (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. Because of the quantity of data to be generated, separate spreadsheets will be generated for the extracellular enzymatic activity data. Due to data volume, the pyrosequencing data will additionally be archived separately (see below).

| **Type of data** | **Brief description of measurement/archive** |
|---|---|
| Extracellular enzymatic activities of water column communities; enzyme activities measured in incubation experiments. | Measured via increase in fluorescence (MCA-labeled peptide substrates) and change in substrate molecular weight (fluorescently-labeled (FLA-) polysaccharides and plankton extracts); data recorded in lab books (MCA) and processed as Excel spreadsheets. For FLA-substrates, chromatograms are exported as Excel files from chromatography system for further processing. |
| Bacterial heterotrophic production | Measured via $^{14}$-C leucine incorporation. Data collected from scintillation counter, processed in Excel spreadsheet. |
| Microbial cell counts | Direct counting via microscopy; processed as Excel spreadsheet; data to master Excel spreadsheet. |
| C/N analysis, DOC analysis, POC analysis | Computer-generated (from analytical instrument) output will be stored in master Excel sheet. |
| nitrate, ammonium, total dissolved nitrogen, orthophosphate, chlorophyll | Measured analytically (see Methods); data to master Excel spreadsheet. |
| Peptides and total carbohydrates (dissolved and particulate) | Measured analytically (see Methods); data to master Excel spreadsheet. |
| pyrosequencing of microbial community composition | Samples to be collected and frozen on dry ice or 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 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 also be made available to other researchers, upon request, after publication. Pyrosequencing data will be submitted to NCBI (the Short Read Archive) after analysis.

**Data Archives**
The Sakai website will be maintained and backed up by the Information Technology Office at UNC. In addition, the master Excel spreadsheet will be stored indefinitely on Arnosti's computer (backed up three times weekly on two different external hard drives). 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 pyrosequencing data is stored on computers in the Teske lab, and will be submitted to the NCBI's Short Read Archive once analyzed. Lab books relating to the project (basic experimental data, worksheets for cell counts, MUF/MCA activities, leucine incorporation, and chemical analyses) are stored indefinitely in the Arnosti lab. We have recently also begun to scan our lab books and store the PDFs on our computers as an additional backup.

**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 in the future 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.