

## Data Management Plan

1. The following types of data will be collected and produced in the course of the project:

- a) A variety of physical data (temperature, salinity, dissolved oxygen, fluorescence, scatter) using a Sea Bird Electronics (SBE) conductivity, temperature, and density (CTD) profiler (from experiments to occur in Bergen, Norway). We will also collect *in situ* light levels using a Biospherical PAR meter.
- b) Fluorescence data (FIRE fluorometer) using software developed by Satlantic (Halifax, CAN). Cell abundance data will be generated using a Guava easyCyte HT (Millipore) flow cytometer, using software native to the cytometer. These data will then be exported in Excel or Matlab.
- c) RNA-seq data and 16S amplicon sequencing from metagenomics data sets will include raw image files (e.g. CASAVA for BaseCalls to FASTQ files from Illumina), Fasta format (Raw and quality-trimmed sequence reads, consensus sequence of assembled contigs), and gene annotation and assemblies.
- d) Gene expression profiles and melt curve profiles generated on the StepOnePlus real-time PCR system will be exported to Excel.
- e) Cell physiology measurements by confocal microscopy will be saved as individual images or z-stacks as \*.tif or \*.raw format. Cellular fluorescence measurements on the BD FACS Aria II flow cytometer will be displayed in plots that are gated to define populations of interest and analyzed using BD FACSDiva software. Biochemical measurements from microplate assays obtained from the SpectraMax M2 plate reader will be analyzed in SoftMaxPro Software. Results will be exported in Excel. Protein expression data obtained by Western blot analysis will be imaged via the FluorChem Imaging Workstation and transilluminator by Alpha Innotech, and analyzed via the FluorChem Digital Darkroom User Interface and AlphaView image analysis software and exported as \*.fcz or \*.png files.
- f) Mass spectrometry data will be collected in .wiff format, to be collected and analyzed using the AB SCIEX Analyst and MultiQuant programs.

2. Standards to be used for data and metadata format and content:

- a) All CTD data will be acquired as \*.dat or \*.hex for use in Seasave v7. The data will be converted to a \*.cnv file for use in SBE Data Processing's Sea Plot or converted to ASCII for use in Excel.
- b) FIRE fluorometer data is collected using Satlantic software FIREView, FIREPro and FIRECONT, all of which run in a DOS OS. Data files (\*.dat) will be exported and imported into Excel for formatting. Flow cytometry data will be acquired using the native software of the flow cytometer, producing files in FSC 3.0 format, and allowing for import into other programs such as Excel.
- c) We will ensure our molecular datasets conform to the Minimal Information about a Metagenomic Sequence (MIMS) and Minimal Information about a Marker Sequence (MIMARKS) as described by the Genomic Standards Consortium (GSC). RNA-seq data and analysis will conform to the standards as described in the Encyclopedia of DNA Elements (ENCODE) Consortium.
- d) Mass spectrometry data will be quality controlled with several pre-processing steps using both the native instrument software and freeware.

3. Policies for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights requirements:

- a) All data sets will be available online within two years of sample collection and analysis. Chemical, physiological, microscopy, and oceanographic data will be made available at the Biological and Chemical Oceanography Data Management Office (BCO-DMO; <http://www.bco-dmo.org/terms-use>) data repository.

- b) Raw sequencing data and alignment information from Illumina RNA-seq and metagenomics experiments will be made available within two years of final data analysis on NCBI's Sequence Read Archive (SRA) (<http://www.ncbi.nlm.nih.gov/sra>).

#### 4. Policies for data use:

- a) We will generally apply the least restrictive data and program licenses to generate the highest impact for collaborative efforts. Data will be freely available for commercial and non-commercial re-use after publications in peer-reviewed journals with the provision that the NSF grant be properly acknowledged in publications and presentations.
- b) When possible, open-access journals will be used to publish work. Publication of data shall occur during the project, if appropriate, or at the end of the project, consistent with normal scientific practices. Research data which documents, supports and validates research findings will be made available after the main findings from the final research data set have been accepted for publication.
- c) Materials generated under the project will be disseminated in accordance with participating institutional and NSF policies. Materials may be transferred to others under the terms of a material transfer agreement.

#### 5. Plans for archiving data/samples:

- a) Data will be submitted BCO-DMO as described above and published, both in print and online as journal articles or supplementary material. In addition, all data files will be retained at least five years after the end of the project, as well as stored at the local home institutes.
- b) At Haverford College, raw data will be stored on an encrypted network housed in the Whalen Laboratory and backed-up on institution-supported storage server which is redundantly spread across multiple san servers for easy backup. Tape archives of the data are also archived, and the college maintains both local and off-site copies of these tapes as a normal function of disaster preparedness. Within the TriCollege Libraries Institutional Repository (Haverford, Bryn Mawr, Swarthmore), Haverford College runs a DSpace repository called Triceratops (<http://triceratops.brynmawr.edu/dspace/>) to collect, describe, provide access to, and preserve digital materials associated with research and scholarship. Raw data pertaining to molecular analysis and physiological measurements will be maintained in this Repository and made publically available at no charge to the user. Electronic copies of senior theses are maintained by the Science Librarian via a Digital Thesis Archive at Haverford College and can be accessed upon request.