Data Management Plan

Data QA/QC

Calibration protocols for all instruments will follow the manufacturer's recommendations with respect to frequency and procedure. Complete calibration histories will be documented via paper records, with metadata files used to record the most recent and applicable calibration date for each record. Raw data downloaded from instruments will immediately be archived on internal servers; post-processing will be performed with copies of the original. Hand-collected data will be entered into customized data templates and immediately transcribed into corresponding computer spreadsheets or relational databases. Paper copies will be archived in notebooks. All electronic records will be backed up to remote systems on a daily basis.

Depending on the data source, all incoming data will be evaluated by either senior project personnel or computer algorithms. Suspect data (e.g., questionable time stamp, out of sequence, out of range for an instrument, outside of normal range) will be marked with a flag set in an additional field. The interpretation of suspect data flags will be described in metadata files. Philosophically, we believe in retaining all data and labeling suspicious values, which may either be recoverable (e.g., data corresponding to an erroneous pressure value in a CTD data file can still be assigned to a depth range by using the two flanking values) or have a useful meaning in another context (e.g., accelerometers that collect wave data generally respond to the dominant wave form, until the waves exceed a threshold – data above that threshold no longer reflect the dominant wave, but may be useful to somebody decomposing complex mixtures of different wave forms). The basis for excluding data from subsequent analysis will be clearly stated in all publications.

Data Archive

We are committed to making our environmental data freely available in as timely a fashion as possible, subject to the limitations of existing data repositories. The following plan is based on our understanding of existing services, but data archives are expanding and new repositories are being developed. Consequently, we may modify our plans if more appropriate repositories become available during the course of the project. Proposed archiving and access plans are outlined by data type.

1) Hydrographic and current profile data from CTD and ADCP surveys. All CTD and ADCP data will be archived with the NSF-funded Biological and Chemical Oceanography Data Management Office (BCO-DMO) within 2 years of collection, with metadata posted within 90 days. We will work with BCO-DMO to assure that the data and metadata will be submitted in a format acceptable to BCO-DMO.

2) Larval data. Barnacle larval densities from the vertical sampling will be similarly archived with BCO-DMO on a similar time schedule. We reserve the right to embargo the larval data for a reasonable period of time to permit publication of the results. Physical samples will be archived within our labs until the results are published, at which time samples will be discarded. There does not appear to be sufficient justification to incur the costs of archiving these specimens long term.

3) Intertidal settlement densities, reproductive output, and community experimental data. We will also make these data available to BCO/DMO. These data types are for benthic phases and some of the measured variables (e.g., settlement, reproduction) integrate over longer time periods than any site-specific environmental data that we will collect, while experimental data only have meaning within the experimental context. Consequently, these data appear to have little value to the community outside of the planned sampling or experimental contexts, and we propose to primarily make data available post-publication, with peer-reviewed journal articles serving to alert the community to the existence of those data. However, we can also make these data widely available by publishing "data papers" in the Ecological Society of America's *Ecological Archives* and portals like DRYAD. Moreover, if other appropriate data portals develop during the course of this project, we will submit the pertinent data to that portal.

4) Elemental analysis data. Elemental data along with all metadata will be submitted to BCO-DMO shortly after they are collected and analyzed. We reserve the right to embargo the elemental data for a reasonable time period of time to permit publication of the results, but all should be submitted by the completion of the project. In addition, we will make data available to the community post-publication through peer-reviewed journal articles and portals like DRYAD, serving to alert the community to the existence of the data. If additional appropriate data portals emerge, we will submit the pertinent data to that archive as well.

5) Predicted larval trajectories from models. Model output including gridded sea level, temperature, salinity, velocity, and mixing coefficients as well as predicted larval trajectories, derived settlement density and connectivity matrices, all in NetCDF format and OPeNDAP compliant, will be submitted to BCO-DMO. BCO-DMO will archive all the data they manage at the appropriate national archive facility, such as NODC and NGDC. In addition, model results will be archived at the project website and served via a THREDDS server at http://rocky.umeoce.maine.edu/thredds/catalog.html.

All classes of data as well as metadata associated with this proposed research, including information on sites, experiments, and data collected (*e.g.*, date, time, location, experimental treatments and maintenance, and environmental variables measured) will be archived and made accessible to researchers and the public on the Northeastern University Marine Science Center website (http://www.marinescience.neu.edu/). Data formats will be comparable to those used by the national repositories. Any embargoed data will be made available to project participants only via a password protection system until the embargo is lifted. This venue offers the advantage of quicker turn-around when posting new data (i.e., fewer administration layers involved in processing), but at the expense of long-term (beyond the termination of the project) stability. Hence the redundancy of the two systems maximizes both speed of availability and stability.