

GENERAL: N.R. Bates and R.J. Johnson (BIOS). Our data management plan is based on guidelines established by the National Science Board and the National Science Foundation and covers dissemination and sharing of materials and data that are expected to be collected as part of the research described in the above-named proposal. The data management plan has five categories:

OBSERVATIONAL DATA:

(1) Types of data, samples, physical collections. Johnson and BATS/Hydrostation S programmers are continually enhancing programming scripts to streamline processing of all Hydrostation S related data. The current data system based on structured array Matlab files for each cruise for both CTD processing and bottle data (salinity and dissolved oxygen), is efficient, robust and scalable for future data efforts. Each Matlab structured array for each cruise contains all levels of CTD data, bottle data, meta data and calibration information. For QC purposes, all CTD and bottle data streams have quality masks consisting of matrices of the same dimensions as the datafiles. The WOCE quality flag criteria are used with all Hydrostation S data.

CTD data processing is typically completed within one week of the cruise following analyses of salinity and dissolved oxygen samples, and subsequently QC'd prior to the next cruise to help ensure no propagation of CTD related problems (for methods see Knap et al., 1997). The first level of QC for a cruise starts with CTD Data processing which consists of two stages: (1) CTD signal conversion and dynamic sensor correction; and (2) static drift corrections and empirical field calibrations. Once the CTD data are QC'd the next step is to assess the quality of the discrete salinity and dissolved oxygen samples as compared against the CTD profiles. If the data appear errant (salinity differences  $> 0.003$ ; dissolved oxygen differences  $> 3 \mu\text{mol/kg}$ ) at a specific depth, then a decision is made to flag the OTE bottle. Following QC at an individual cruise level, data are quality controlled 'globally' using time-series methods with all previous Hydrostation 'S' data. If data appear suspect then it is assigned a flag = 3 (questionable) or 4 (bad), whereby for the latter, disseminated data will be replaced by a nominal bad value = -999.

Following each cruise, underway data will be contributed by the vessel operator to the UNOLS central data repository managed by the Rolling-Deck-to-Repository (R2R) project, which will ensure that the underway measurements are archived at the National Oceanographic Data Center

(2) Standards to be used for data and metadata formatting and content. All primary CTD data will be archived in standard Sea-Bird ASCII and HEX files. Processed CTD data, discrete sample data, and meta data will reside internally at BIOS in Matlab (\*.mat) files and ASCII. For dissemination purposes data will be provided in ASCII, Matlab, Excel and compressed tar files.

(3) Mechanisms for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements. Hydrostation S data access will be through two national based web portals (Biological and Chemical Oceanography Data Management Office, BCO-DMO and the Clivar and Carbon Hydrographic Data Office, CCHDO), and a cloud based FTP web site maintained by BIOS. Data updates will be provided on a semi-annual basis in June and December of each year and distribution of data to BCO-DMO will be as a compressed tar file via a shared Dropbox folder hosted at BIOS. As per previous conversations with BCO-DMO staff and their policy, they will submit all data to the National Oceanographic Data Center for long-term archive when the

timeline is triggered and further will assign a Digital Object Identifier (DOI) to the submitted data with a unique version number for each new data submission. Staff (Steve Diggs) at the CCHDO will routinely pull data from the BIOS Hydrostation S FTP site and convert to WHP-Exchange and WHP-NetCDF formats for hosting on their web portal and the OceanSITES site.

Members of the oceanographic community can always contact the Hydrostation PI's ([nick.bates@bios.edu](mailto:nick.bates@bios.edu) and [rod.johnson@bios.edu](mailto:rod.johnson@bios.edu)) for more recent datasets, other data products or data queries. We will also continue to ask the oceanographic community at large to bring to our attention any problems with the data or web site so that these problems can be corrected in a timely manner. At the time of writing this proposal, Hydrostation S CTD and bottle data can be extracted through the FTP site for data through July 2020, while preliminary CTD data; plots, and cruise event logs for cruises through January 2021, are also available at this location.

(4) Policies and provisions for re-use, re-distribution, and the production of derivatives. Data from this work will be submitted to BCO-DMO where they will be publically available two years after collection (although we aim for within six months of collection). Peer-reviewed papers will be written about these data, along with details of their collection and data interpretation. All data generated as a result of research performed at BIOS shall be the intellectual property of BIOS, as well as the investigator or co-investigator overseeing the research. BIOS will retain an implied copyright for these data.

(5) Plans for archiving data, samples, and other research products, and for preservation of access. Each investigator is responsible for maintaining data associated with their own research group's activities, and in accordance with any institutional requirements of the PI's home institution. Basic practices for key areas are outlined. **Lab notebooks.** All information connected with initial data collection, analysis, and results shall be kept in a lab notebook and in cases digital notebooks may be used. Regardless of media, these notebooks shall be stored securely. Hard-copy notebooks shall be situated to enable institutional retrieval. Digital notebooks will follow short-term and long-term policies of each laboratory outlined below. All research notebooks of each investigator are property of their home institution. **Short-term storage and data management.** Data storage shall be appropriately and redundantly stored using computer hardware and software that is available to each laboratory, and may include in-house and/or off-site resources. The PI and Co-Is are responsible for due diligence with respect to short-term storage of data. Additionally, all data shall be retrievable from primary media or back-ups, as well as reasonably protected from accidental loss due to corruption, power loss, or failure of computer hardware. **Data security.** Data shall be stored on either off-network mobile devices (external hard drives) or off-site cloud resources. Password protection will be utilized. This data must be made available to senior institution officials in the case that any institutional liability issues should arise. **Data backup.** All associated project data at BIOS will reside on primary and secondary Linux workstations with both servers supporting a RAID 1 data storage system which in turn is backed up daily using a Crashplan near real-time replication service. These systems are adequate and provide scalability for new data collected from this project through 2025. **Deposit and long-term preservation.** Long-term strategy. After the project has been completed, arrangements shall be made to transfer data at each home institution from short-term storage to a long-term archival system. **Length of archival.** Data will be kept in long-term storage for at least five years, or until it has been successfully uploaded to and made publically available through a nationally or internationally funded database specific to that data