

DATA MANAGEMENT PLAN

General:

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:

(1) Types of data and samples

Data will be collected over a period of two years and include (i) data directly measured by moored sensors in-situ, (ii) data gained through measurement of water samples in the laboratory and (iii) processed data (e.g., metabolic rates). The data directly measured in-situ is first stored on the logging device, which will be downloaded and stored after each deployment period. This includes dissolved oxygen, light intensity (PAR), temperature, salinity, pressure (depth) and pH. Data gained from water sample analysis include total alkalinity (TA), dissolved inorganic carbon (DIC), inorganic nutrients (nitrite, nitrate, phosphate, ammonium) and chlorophyll-a. Processed data include net photosynthesis rates, respiration rates and calcification rates. All data sets will be accompanied by metadata, including date, time, experiment ID, replicate ID, unique sample ID, and geolocation where relevant. The software Microsoft Excel, R and MatLab will be used to manage and analyze the data.

(2) Standards to be used for data and metadata formatting and content

Primary data derived from the SeaBird sensors and sensors connected to the incubation chambers will be archived in standard instrument format, such as ASCII, hexadecimal, .csv or .txt, data formats that are readable by most data processing software. Metadata will be stored as text files (.txt). Processed data and data of discrete water samples will be provided as .txt or .xlsx formats.

(3) Mechanisms for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements

All data produced as part of this grant is intended for publication in peer reviewed journals, conference presentation, teaching, and outreach activities and will be made available to the public with the permission of PI for that data. Metadata will provide name and business contact info of the individual responsible for data archival. Privacy is not expected to be an issue. For sharing data files amongst the researchers involved in the project we will use password-protected Dropbox software or Google Drive.

(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 as soon as possible after collection and validation (we aim for less than one year after collection). Peer-reviewed papers will be written about these data, along with details of their collection and data interpretation, and references to the BCO-DMO datasets.

(5) Plans for archiving data, samples, and other research products, and for preservation of access

All data will be retained for at least three years beyond the award period, as required by NSF guidelines. For short- and long-term storage, raw data and their derivative products from downstream analysis will be backed up in lab repositories of the data generator. Lab notebooks - All information connected with initial data collection, analyses, and results shall be kept in a lab notebook. In cases in which data are too plentiful to record by hand in a paper notebook, digital notebooks may be used. Regardless of media, these notebooks shall be stored, as well. If hard-copy notebooks are used, storage and archival shall be situated to enable institutional retrieval. If digital, storage and archival shall follow short-term and long-term policies of the laboratory outlined below. All research notebooks of each investigator are property of the home institution.

Short-term storage and data management

Data volumes: Data storage will be appropriately and redundantly stored using computer hardware and software that is available to BIOS and AMNH staff, and may include in-house and/or off-site resources (e.g., Dropbox or Google Drive). The PI is 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 will 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: Data will be backed-up immediately after download from the instruments or weekly in case data is a product of laboratory measurements. Sawall, Grundle and Goodkin use external drive backups and cloud backups supported by BIOS or AMNH IT staff.

Deposit and long-term preservation

Long-term strategy: Within each project year, data will be transferred to BCO-DMO for public access and long-term storage. 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.

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.