

Data Management Plan

Coccolithophore Mixotrophy
W.M Balch, Bigelow Laboratory, East Boothbay, ME 04544

My 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, samples, physical collections – Year 1: I am proposing to do a series of laboratory experiments on the incorporation of DOC and POC by coccolithophores into their organic carbon and PIC coccoliths (which would challenge the traditional alkalinity pump paradigm). Laboratory experiments will be focused on axenic coccolithophore clones from the Bigelow NCMA and their ability to take up radio-labelled DOC compounds. I will do the following experiments: 1) screen a large array of dissolved organic compounds potentially taken up by cultured coccolithophores, 2) determine the *uptake* kinetics of four compounds for the five coccolithophore clones with the highest affinity for them, 3) determine the *growth* kinetics on these same compounds in grow-out experiments in light and darkness, 4) assess the exchangeability of these four DOC compounds and 5) preliminary flow cytometry experiments at Bigelow using cultures as well as seawater samples brought back to Bigelow from the ongoing sampling program, GNATS, which samples across the Gulf of Maine.

Year 2- Perform a 12d research cruise on a medium-sized research vessel from the Gulf of Maine to Bermuda. The overall plan is to sample deep-dwelling coccolithophores for their ability to take up labelled DOC compounds (as well as bicarbonate) and to phagocytize bacteria-sized particles. The work will involve incubating them with radio-labeled DOC compounds (based on year-1 laboratory observations) and fluorescently-labeled bacteria. Then, post-incubation, concentrate them using centrifugation and use a flow cytometer with multiple lasers to sort the coccolithophores. The sorted cells will be examined for their uptake of the ^{14}C -DOC compounds into POC and PIC. This will be compared to autotrophic production estimates using sorted cells incubated in $^{14}\text{C}\text{-HCO}_3^-$. Finally, the flow-cytometer results will also provide evidence (a) whether coccolithophore cells phagocytized the particles as well as (b) the calculated rate of ingestion.

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

I have discussed my data specifics with Cyndy Chandler (Principal Investigator of BCO-DMO) and she directed me to templates for data inventory. BCO-DMO will ensure that data and metadata are available in standards-compliant formats. Tabular data and metadata from our laboratory experiments in year 1 will be in Excel spreadsheets (results of coccolithophore screening, DOC uptake and growth experiments, preliminary trials with the flow cytometer using cultures and field material brought back to Bigelow Laboratory. Tabular data and metadata from the year-2 cruise will be dealt with as follows: CTD casts will be in Sea Bird formatted files. Other variables sampled from water casts will be kept primarily as tabular data in Excel spreadsheets; SEM microscope images will be stored as TIF files. Data will be made available online via the BCO-DMO data system (<http://bco-dmo.org/data/>) and BCO-DMO will submit all data to the National Oceanographic Data Center for long-term archive. BCO-DMO has the ability to assign Digital Object Identifiers (DOIs) to data sets to promote proper citation of published data.

(3) Mechanisms for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements - Distribution of data to BCO-DMO will be done within two years of their collection via password-secure FTP. All data generated as a result of research performed at Bigelow Laboratory shall be their intellectual property, as well as WMB. As such, each laboratory will retain an implied copyright for these data.

(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. Peer-

reviewed papers will be written by Balch. PI's will present these data, along with details of their collection and data interpretation.

(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 below.

Lab notebooks - All information connected with initial data collection, analysis, 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 each laboratory outlined below. All research notebooks of each investigator are property of their home institution.

Short-term storage and data management

Data volumes - Data storage shall be appropriately and redundantly stored using computer hardware and software that is available to the Balch's laboratory, and may include in-house and/or off-site resources. 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 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 - Balch uses the Bigelow Laboratory high-performance elastic computing system backed by 200TB of high performance storage. This system facilitates the Laboratory's long-term data management strategy as well as short-term hourly backups. Cruise data will be backed up on this system once the cruise is completed and personnel return to Bigelow Laboratory.

Deposit and long-term preservation

Long-term strategy - Within two years of data collection, 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.