Data Management Plan

The overall COBRA data management goal is to ensure that all data and products generated in connection to COBRA are archived in stable repositories and publicly accessible for the long term. The data and products generated by the COBRA community will include a wide variety of geophysical, geological, geochemical, and biological information, in addition to physical samples, education materials, scientific publications, technical documents, reports, software, policy briefs, and methodological information. We consider that there are four main data "categories" that will be generated in association with the COBRA project: (1) data resulting from **expeditions-of-opportunity** that COBRA leverages for the leadership training programs, (2) data resulting from early career **researcher exchanges** within the COBRA network, (3) **educational materials** from the leadership training program implementation and COBRA meetings, and (4) data for the **evaluation of COBRA**. Each of these categories will have specific data management needs, as described below.

To facilitate and enhance discoverability of these four data categories, COBRA will establish a web-based search portal linked to a SQL server that points to all data types deposited in appropriate internationally accessible data repositories. The design of this portal will follow a model successfully established within the NSF-funded C-DEBI Science and Technology Center. Currently, much of the data generated by the deep-sea community related to the crustal ocean biosphere is publicly available but difficult to discover and/or use. Disparate data types are deposited in various locales and often decoupled from each other in the process. Though cruise reports generate pointers to the various data types, not all cruise reports are themselves public. Furthermore, different countries use different repositories, and parse data in different formats. While the COBRA web portal will not address data conversions across formats, it will endeavor to become a "one-stop-shop" for links to archival data. As such, this site will enable data discoverability and collaboration among the community, addressing a need of the NOMEC strategy (Ocean Science and Technology Subcommittee of the Ocean Policy Committee, 2020).

The COBRA data portal will include links to datasets, protocols, and publications. Dataset information in the portal will include a description of the dataset, description of acquisition and processing methods, a list of relevant instruments, information on file types and a list of parameters, names of dataset maintainers, and a link to the URL where the data can be found on the primary host repository. Protocol entries will include links to new wet lab protocols uploaded to protocols.io or data processing pipelines described in repositories such as github. Publication information will include a link to the manuscript as well as a COBRA contribution number, to aid in reporting. A form will be available on the website to invite edits and new additions, such that scientists, institutions, etc., will be able to easily request website changes and suggest the addition of relevant new data. Experts from Bigelow's IT department will facilitate establishing the portal within the COBRA website, entries will be maintained by the Managing Director, and co-PI McManus will oversee compliance of reporting. If COBRA sunsets, a static version of the COBRA web portal will be archived in perpetuity on the Bigelow Laboratory website, and we will explore options for transitioning the web portal to an open-access wiki to enable continued use beyond the program.

Data Category #1: Expeditions-of-opportunity that COBRA leverages for the leadership training program will have established data management plans according to the guidelines of the entities that have funded those programs, such as the NSF, NOAA, other national agencies, private foundations, etc. For example, U.S.-based scientists conducting deep-sea research with funding from the NSF's Division of Ocean Sciences in the Geosciences Directorate (i.e., OCE) must conform to the OCE's policy on disseminating and sharing research results (data, samples, collections, other materials), including placing datasets in publicly accessible repositories. Likewise, international scientists participating in the International Ocean Discovery Program (IODP) are required to follow the policies on data and sample sharing as agreed and enforced by the research community and

associated agencies. *Thus, COBRA will not have primary responsibility for making these datasets (i.e., underway data, cruise reports, other primary data generated from the cruises) publicly available, and cannot dictate how, when, or where the PIs of those expeditions make data publicly accessible.* However, COBRA will link to the data, once available, via the web-based portal. Further, COBRA will encourage PIs of the expeditions-of-opportunity to "brand" their primary datasets as affiliated with COBRA and to submit COBRA website data portal entries for inclusion in the searchable web-based system.

Data Category #2: Early career researchers that participate in the leadership training program and/or research exchanges within the COBRA network will generate and/or process a variety of datasets. **To facilitate archival and discovery of these new datasets, COBRA will establish a new "program" within the Biological and Chemical Oceanography Data Management Office (BCO-DMO) database, which is a data repository supported by NSF's Division of Ocean Sciences in the Geosciences Directorate.** All COBRA members who receive funding support will be required to populate new "projects" that link to the COBRA program in BCO-DMO. Depending on the dataset type, the project entry may contain links to primary data files or stable links to where primary data are housed on other public data repositories (for example, nucleic acid sequence data uploaded to the Sequence Read Archive of GenBank of the National Center for Biotechnology Information in the USA, or multibeam data submitted to the Marine Geoscience Data System hosted at the Lamont-Doherty Earth Observatory with funding from NSF). Each of these submissions will be given distinct Digital Object Identifiers (DOIs) by BCO-DMO for attributable reference. This model will allow anyone in the world to access datasets generated with (partial) funding from COBRA. As a stipulation of funding support, researchers that receive funding from COBRA will be required to (i) agree to make data publicly available as soon as feasible, and no later than within two years of data generation or upon publication in primary research literature, whichever is sooner, and to (ii) report information on these data types to COBRA in the form of both an end-of-project report and requests for updates for compiling annual reports.

Data category #3: COBRA will make all educational materials resulting from the leadership training program, such as training modules, lecture recordings, and worksheets, publicly available. Where feasible, the programs will be described in the primary literature, such as the journal for marine education (*Current*), the *Journal of Geoscience Education*, the *Journal of Ocean Technology,* or the *Oceanography* magazine, with the educational materials made available as supplemental materials. These publications and materials will also be linked to the COBRA data portal for discoverability. Redundant copies of, or links to, these materials will also be archived on the COBRA data portal under a section for educational resources, as well as on partner websites where appropriate (for example, materials relevant to E/V *Nautilus* posted on OET' website and to R/V *Falkor* to SOI's website). COBRA co-PIs have experience with publishing such materials in these fora (Orcutt et al., 2011; Orcutt et al., 2013; Wheat et al., 2018). In addition, links to presentations, recordings, and reports from COBRA workshops and symposia will also be made publicly accessible on the COBRA website.

Data category #4: Raw data collected by our CREDITS program partners and the independent external evaluator, to evaluate the effectiveness of our team science trainings and the overall performance of the program, will not be made publicly available in accordance with human subject privacy regulations. The evaluators will provide summarized results of these evaluations to COBRA leadership, with no data linked to personal attribution. Executive summaries of these evaluations will be made publicly available on the COBRA website, with the detailed reports provided as appendices with the COBRA annual reports to NSF. Our CREDITS partners may aggregate some of the team science evaluation data with results from their other programming, to publish reports on the effectiveness of Team Science training. COBRA will link to any such reporting in the COBRA data portal.