

Data Management Plan

Data will be managed within a framework designed to integrate measurements, disseminate results, and enable hypothesis testing. We feel that dissemination and integration of the above components is critical, for both understanding key physiological processes at the cellular level, determining the influence of environmental variables on transcription, and resolving differences in community structure in response to environmental variables. Our overall goal is to optimize the availability and utility of data to all members of the Center, facilitating communication and ultimately the publication of results. We also note that we will start managing our data at the beginning of the project.

The Data Manager, will take the responsibility for acquiring, storing and sharing data following the “best practices” outlined by the Vice President for Research at Bowling Green State University. The Director will supervise the Data Manager in this capacity. Whereas metadata (nutrient and toxin data) will be immediately available to the public, the ‘omics data will initially be shared only among the Center PIs, technicians, and graduate and undergraduate students directly involved in Center activities. As a routine exception, if appropriate to advance scientific collaboration and policy outcomes, data will be shared with other OHH Centers, federal researchers and stakeholders upon request. Decisions to release such data reside with the Internal Steering Committee. However, we reserve the right to safeguard and protect graduate student projects and delay data sharing by a modest time lag to ensure the students’ right to publish original data. As stated below, at the conclusion of the project, all internally shared data will be made freely available via NCBI, BCO-DMO, journal electronic annexes, or posted to a BGSU website.

Data and samples will come in several forms. For biological sample management, a dedicated -80 °C facility at BGSU is used for sample storage (typically filters, DNAs, RNAs) for ready access and cross referencing to metadata. For biological data, integration will start with the development of a web-driven repository for investigators that will feed into an all-project database on a password protected FTP server.

Our objective is to publish all data in peer-reviewed journals and submit all data to that journal to be published in freely-accessible electronic annexes. The analytical procedures will also be made available to other researchers who wish to replicate all or part of the experiments for their own projects. The Final Report submitted to the NIEHS and NSF will include a summary of all data generated.

Specifics on where data will be posted. With respect to distribution to the scientific community, sequence data will be deposited in the National Center for Biotechnology Information (NCBI) as well as within the *Metagenomics Analysis Server* (MG-RAST) which will allow us to couple molecular and site specific data for use by other researchers. Transcriptome data will be deposited in NCBI Transcriptome Shotgun Assemblies (TSA). Raw metatranscriptome data will be stored at BGSU, with backup data on backup drives maintained at the University of Tennessee. Nutrient, toxin and chlorophyll data from all field samples will be curated and stored at BGSU and freely available to all. Raw data (uncurated) files at their source within the Facilities Core (SUNY-ESF and OSU Stone Lab) will be maintained as a backup. Submitted data will confirm with relevant data and terminology standards.

Cultures and strains that may arise from this study will be made publicly available. BGSU maintains a culture collection of bloom-forming cyanobacteria and picocyanobacteria in that has been shared with users upon request. Viable stocks are maintained in DMSO in the -80 °C facility. Additionally, any new strains will also be shared through culture facilities (e.g., the National Center for Marine Algae) which hosts both marine and freshwater cultures.

Finally, all biological, physical and chemical environmental metadata and NCBI accession numbers will be deposited in the Biological and Chemical Oceanography Data Management Office (BCO-DMO; www.bco-dmo.org). We agree to deposit genetic outcome data into TSA and BCO-DMO repositories as soon as possible but no later than within one year of the completion of the funded project period (six years) for the Center, or upon acceptance of the data for publication, or public disclosure of a submitted patent application, whichever is earliest.

Statement regarding access to the data. We agree that data will be deposited and made available through TSA, which is an NIH-funded repository, and BCO-DMO, and that these data will be shared with investigators working under an institution with a Federal Wide Assurance (FWA) and could be used for secondary study purposes such as identifying pathways for biosynthesis of novel therapeutics (*e.g.* see letter from Biosortia Pharmaceuticals). I agree that the names and Institutions of persons either given or denied access to the internal data, and the bases for such decisions, will be summarized in the annual progress report. In all publications and presentations we will identify where the data will be available and how to access the data, as well as acknowledge the repository and funding source in any publications and presentations. As we will be using TSA, which is an NIH-funded repository, this repository has policies and procedures in place that will provide data access to qualified researchers, fully consistent with NIH data sharing policies and applicable laws and regulations.