

## **Data Management Plan**

The NSF team of this international collaborative project will primarily use the Biological and Chemical Oceanography Data Management Office (BCO-DMO) as our primary public data repository. We will register our project with BCO-DMO and submit data with metadata periodically. Data will also be archived on the UNCW ArcShare server which is a dedicated and physically secure server for research data that is backed-up daily. The PI will also make data available through publications, supplemental data and on request. Data will be published whenever possible through open access journals.

### **(1) Types of data, samples, and other products.**

The primary *product* of this research program will be scientific publications that fully present, analyze and discuss the data listed below:

- Nutrient and chemical analysis data
- Cell physiological data- e.g. growth rates, uptake kinetics, photosynthesis
- Western blot, immunocytochemical and other biochemical data
- Genetic sequences and bioinformatics analysis of Si transporters
- DNA and Transcriptome data (UK)
- cDNA and genomic clones of Si transporters
- Microscopy images (SEM, Raman and AFM), spectra, and associated metadata
- Electrophysiological recordings of coccolithophores and *Xenopus* oocytes
- Ecosystem modeling using Si requirements and transport kinetics (e.g. DarwinMIT model) (UK)
- Outreach and video material will be posted to social media platforms and be available on request for educational purposes.

### **(2) Standards for data format and content.**

The project is primarily laboratory based. Modeling data will be derived from existing publicly available oceanographic data sets (e.g. MARDAT, Western Channel Observatory, Darwin-MIT model). The PI will work with BCO-DMO staff to identify the appropriate raw and derived experimental data to be deposited, and to insure compliance of data archiving with associated metadata. Biological resources (e.g. phytoplankton culture strains and *E.coli* containing DNA clones) will be maintained in the PI's laboratory at UNCW and at MBA UK with appropriate digital cataloging and documentation using Excel spreadsheets (.xlsx). Hard copies of experimental protocols and data sheets containing pertinent metadata will be stored in binders and as digitally scanned copies (pdf). A master spreadsheet (.xlsx) containing information on the location and content of all digital files will be kept and regularly updated as experiments progress. Analytical data will be stored as their appropriate file types, with DNA sequence data stored as raw output files (abi, sff), transcriptome (.bci, Fastq) and as final files (.txt in fasta format). Microscopy images will be stored as Tiff and Jpeg format and videos in .avi and Quicktime (MP4) format. All digitized data will be backed up on external hard drives both on-site (PI's computers) and off-site on dedicated UNCW ArcShare server or MBA data server to ensure data preservation and availability.

### **(3) Policies for access and sharing.**

In addition to using BCO-DMO, publications resulting from this project will be readily available through library databases, and will be published open-access when possible. Raw datasets from all project partners will be posted online to international databases, as appendices or supplementary material in publications, on the PIs websites, through BCO-DMO or upon request. DNA sequence data with relevant metadata will be deposited in international databases to provide global, online access to molecular data, including the National Center for

Biotechnology Information's (NCBI) GenBank sequence data archive and NCBI's Sequence Read Archive (SRA) for next generation sequencing projects.

**(4) Policies for re-use and re-distribution.**

The data arising from this research program will primarily be disseminated as published journal articles. Raw data will be deposited with the appropriate public repository as described above, and PIs will maintain copies for provision to other researchers on request. The PIs and collaborators will retain rights of the data until it is published or two years after its collection, whichever comes first. On publication, the data and biological resources will be supplied to interested parties on request. Prior to publication materials and data will be provided on approval with co-authors/collaborators. Digital copies of data will be supplied to requestors through FTP or via cloud based sharing such as Dropbox. The investigators will request appropriate credit to be given when data resources are used in another work.

**(5) Plans for archiving data, samples and other products.**

All NSF-funded persons associated with this research program, will undertake the Responsible and Ethical Conduct of Research (RCR training) in online and face-to-face sessions, as per UNCW policy. This training sets the benchmark in quality assurance for collection, maintenance and archiving of scientific data. Physical copies of experimental protocols and data sheets containing metadata from experiments will be stored in secure lab locations at the institution they were generated (MBA or UNCW). They will also be scanned and digital copies will be made available at both institutions. All digital data will be stored on a combination of internal and external hard drives as the grant progresses and after the grant ends. A backup of all digital data will be stored on secure servers at the respective institutions that are backed up daily.