

DATA MANAGEMENT PLAN

I. Types of data, samples, and materials to be produced: This research project will collect a variety of data and samples from the fieldwork, as well as the laboratory analyses of those samples. These will include: 1) biological specimens (= sponges and their microbial symbionts), 2) images of these specimens and the metadata associated with them [*e.g.*, date and location of collection, field characteristics, *etc.*], 3) phylogenetic (taxon-by-character) matrices, 4) genetic sequences, 5) biochemical profiles [*i.e.*, proximate biochemical composition, secondary metabolites, stable isotopes], 6) functional attributes [*e.g.*, trophic dynamics, feeding deterrence, *etc.*], and 7) coral reef biodiversity of reefs in the Caribbean basin, as well as the oceanographic [*e.g.*, temperature, light, nutrients, *etc.*] conditions at all sites. The PIs will coordinate, archive and manage all data generated by the project participants.

II. Standards for data and metadata: Field notes will be collected on underwater paper *in situ*, transcribed into notebooks upon return to the laboratory, and ultimately converted into electronic formats archived on redundant hard drives. Digital photographs taken in the field will be downloaded, labeled appropriately, and saved to computers and external hard drives at each institution. Results from laboratory analyses will be recorded in laboratory notebooks, converted into electronic formats, and stored in secure locations in the PIs' offices and laboratories. Computers within the PIs' laboratories will be backed up on external hard drives daily. All electronic data will be archived in formats described below on university computers, and backed up within the University of New Hampshire and the University of Mississippi's Supercomputing Center in a sector dedicated to the PIs for management of their data. In compliance with NSF guidelines, hard copy versions of the data will be retained for at least three years following the award period.

For specimen-based data, we will adhere to Darwin Core standards and maintain compatibility with DiGir and TDWG formats. Images will be stored in JPEG and/or TIFF formats. Phylogenetic data will be stored in TreeBASE (treebase.org) formats, intercompatible with NEXUS and NeXML formats and the PhyloWS API. Gene sequence data will be stored in GenBank XML format and in Dryad for maximum interoperability. Tab-delimited, plain text files will be used for raw data storage as they can be easily imported into various analysis programs such as Microsoft Excel, PRIMER, R, and ArcGIS. Biochemical profiling, functional attributes, and coral reef surveys will be archived in Microsoft Excel spreadsheet with the appropriate metadata in Microsoft Word files. Physical oceanography data will be processed and archived with appropriate metadata.

III. Policy for data access and sharing: Data access and sharing will comply with all NSF policies. The data from this study will be made available through the institutional web pages of the PIs at UNH and UM as well as specific established data portals (BCO-DMO). The PI websites will be an important primary archive for reports and publications, as well as survey data, metadata, and photographs. The sequencing data from the biodiversity surveys will be deposited for broader dissemination at established data portals (*e.g.*, NCBI, Dryad). All oceanographic data from our water column work (*i.e.*, temperature, salinity, irradiance, spectral absorption, nutrients) will be archived in Microsoft Excel spreadsheets with the appropriate metadata in Microsoft Word files, and maintained at UNH and UM.

Biological samples, including DNA, RNA, and cDNA, will be shared freely upon request via a Material Transfer Agreement (MTA; see IV). Sequence data will be made publically available in a timely manner, and next generation sequence data will be posted to NCBI GenBank, Dryad and other archival databases described above, prior to publication. We will also provide the necessary information for our NGS methodologies and validation protocols. Processed or annotated data will be made available as soon as possible right after publication. We already have scripts in place to generate and submit sequence data to the NCBI Sequence Read Archive (SRA) in an automated process. Scripts written during the course of developing our bioinformatics pipeline will be made open source and available upon validation.

IV. Provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements: While the primary goal of this project, to assess the sponge biodiversity of the Caribbean, does not represent a significant intellectual-property issue, these specimens do have potential biotechnological value. When we apply for scientific collecting permits we will specifically identify this potential issue, and gauge interest in a possible collaboration relative to intellectual property. Should either country wish to enter into an agreement, it will likely represent a screening arrangement between UM's National Center for Natural Products Research (NCNPR) and the sovereign nation that will be defined up-front based on prior agreement practices of the interested parties. Subsequent usage will require an MTA and assurances that the original terms of the agreement are respected. In the event that discoveries or inventions are made in direct connection with these data, access to the data will be granted upon request once appropriate invention disclosures and/or provisional patent filings are made. We do not anticipate any privacy, security or ethical considerations associated with samples collected during this research project.

V. Policy for re-use and/or re-distribution of materials produced: Physical samples will be stored in public repositories (Smithsonian Institution National Museum of Natural History Invertebrate Zoology Collection [NMNH]), as well as the laboratories of the PIs, where they will be accessible according to the facilities' policies and/or requests to the PIs. Samples will be deposited prior to publication and voucher data or specimen numbers will be included in relevant publications. Data will be disseminated through appropriate peer-reviewed scientific publications, and shared via open-access public databases (including NMNH, TreeBase, and Dryad) or as supplemental material to publications where no databases exist.

Electronic data and/or publication access will be posted at the PIs website approximately six months after sample processing, and will be available upon publication for re-analysis or for new analyses. For data that cannot be immediately published (*e.g.*, intellectual property data: IV), data will be embargoed for up to two years through the Biological and Chemical Oceanography Data Management Office (BCO-DMO; www.bcodmo.org/), or until publication.

VI. Data and sample archiving and preservation: Physical archival samples will be photographed *in situ* and preserved in RNAlater, DNA buffer and 70% EtOH for future molecular work and for taxonomic vouchers, respectively. Metadata for each sample will include: sample code, taxonomic identification, collection date/time, site name/depth, and GPS coordinates. Physical samples will be deposited with the digital metadata that will be available to Internet-based searches (*e.g.*, collections.nmnh.si.edu/search/iz).

All DNA, RNA, and cDNA resources will be stored at -80°C at UNH for a minimum of five years after the publication of results from this project. Information on the individual samples, including metadata and other pertinent information (*e.g.*, sequence protocols, publications, *etc.*), will be digitized using Microsoft Excel, and at least two backups will be stored independent from the master copy. NCBI storage is also publically available, and backed up daily. Secondary metabolites isolated from feeding deterrent sponges will be maintained at UM's NCNPR in -80°C freezers.

Following NSF guidelines, all project data will be deposited at the BCO-DMO for broad dissemination and redundancy with the PI's digital files; utilization of this service ensures that the raw data sets will be available in useful formats in perpetuity. Oceanographic time series and hydrographic data will also be deposited with and shared via the National Ocean Data Center (NODC; www.nodc.noaa.gov/).