

## Data Management Plan

### 1. Products of Research

The proposed research will include four kinds of data: (1) biological specimens that will be deposited into museum collections with associated metadata (e.g., date and location of collection, images, and morphological features measured in the laboratory or field); (2) genetic sequences (including 28S and 16S barcodes) and associated metadata (species description, experiment, collection location); (3) biochemical data and associated metadata (e.g., stable isotope concentrations for each experiment and each species, nutrient concentrations); and (4) survey data from questionnaires as part of the broader impacts.

The above data and metadata will be collected and stored in standard file formats to ensure that data can be shared and uploaded to a variety of Internet-based, open-access databases. 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. Gene sequence data will be stored in GenBank XML format for maximum interoperability. All next-generation sequencing datasets will be stored as raw data in the GenBank short read archive and as processed data in NEXUS format. Biochemical data will be stored as tab-delimited, plain text files and archived as supplemental materials to publications and deposited in BCO-DMO. Analytical workflows (including R and Python scripts) will be implemented in Arbor and archived at the Arbor GitHub website; workflows will also be made available as supplemental materials to publications.

### 2. Data Storage and Preservation

Specimens freshly collected during this project will be deposited into museum collections at the National Museum of Natural History (NMNH). Metadata and images related to specimens deposited in museum collections will be digitally stored by the museum collection and will be available to Internet-based searches.

Phylogenies used in published analyses will also be deposited in Open Tree of Life ([opentreeoflife.org](http://opentreeoflife.org)). Genetic sequences will be deposited in GenBank ([ncbi.nlm.nih.gov/genbank](http://ncbi.nlm.nih.gov/genbank)) with appropriate metadata attached to each sequence.

All DNA resources will be stored at -80°C at Nova Southeastern University 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 at Nova Southeastern University and Stony Brook University. NCBI storage is also publically available, and backed up daily.

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 datasets will be available in useful formats in perpetuity. In addition, all publication files, specimen records, images, and outreach survey data will also be stored on an open-access project website maintained on GitHub.

### **3. Dissemination of Data and Metadata to Others**

Data from this project will be disseminated through scientific publications and EOL, as well as through the open-access repositories described above, including NCBI, and Open Tree of Life. We will maintain an open-access project website. In addition, we will provide links to all data repositories at the Porifera Tree of Life website, maintained by PI Thacker.

### **4. Policies for Data Sharing and Public Access**

None of the samples or data involved are proprietary, high-security, or sensitive. Physical samples will be stored in public repositories (museum collections) where they are accessible according to the facilities' policies. Samples will be deposited prior to publication and voucher data or specimen numbers will be included in publications. Data will be shared via appropriate open-access public databases (e.g., NCBI and BCO-DMO).

As the data and samples will be archived in public institutions and published, no permission is needed from the PIs for re-use. Samples and data will be made available upon publication for re-analysis or for new analyses.

### **5. Roles and Responsibilities**

The PIs of this proposal will be responsible for ensuring that the specimens and data collected during this project will be deposited in the appropriate public repositories described above.