

Data Management Plan

Types of data, samples, and physical collections

1. Individual organisms preserved in formalin and/or ethanol.
2. Genetic data stored as FASTA format files.
3. Species by sample data (i.e. abundances) stored as comma separated files.
4. Stable isotope compositions of animals and sources of organic matter stored as comma separated fields.
5. Species level body size as mass stored as comma separated files.

These data will be accessible in three forms:

1. Scripts that allow other investigators to recreate our analyses in R.
2. Raw data
3. Summary data/results from analyses

Data and Metadata Standards

The date (deployment and collection), location (latitude, longitude, depth), collector and processor, vehicle information, wood type, wood size (volume, surface area, weight) will be linked to a unique wood fall identifier.

Body size data, isotope ratios, and genetic files will be connected to species data through a species identifier consisting of the genus and species epithet. Size data files will include the following fields: species name, mass (g), data collector, and date of collection. We will use open, non-proprietary file formats, e.g. csv, for the body size datasets. Isotope ratios datasets will include species name, number of individuals pooled for a given isotopic measurement, mass (micrograms) of each individual included in the pooled sample, stable isotope compositions ($\delta^{13}\text{C}$, $\delta^{15}\text{N}$, $\delta^{34}\text{S}$), elemental compositions (% C, % N, % S), habitat (e.g., log interior, log exterior, sediment surface, sediment subsurface), data collector, and date of collection.

Scripts for R will include documentation for the code generator, date of origination and modification, R-package libraries required for script. Scripts will be compatible with current R-packages. Our code will have an open source license allowing for dissemination, usage, and modification if appropriately referenced and used non-commercially. This extensive documentation, openness, and compatibility will allow for reproducibility and increased utilization by others for new analyses.

Data Storage, Management, and Backup

During data assembly and analysis, working datasets will be housed at the National Evolutionary Synthesis Center (NESCent). NESCent operates six Apple Xserve and Linux servers for development and production of databases, web-applications, and infrastructure software. A shared file server provides more than 2TB of RAID storage for large files and databases with enough unused disk slots available to double the capacity. The data center is supplied with supplemental air conditioning, gigabit network connections, UPS for backup power, and a combined file/tape backup system. We anticipate generating less than 5GB of data in the course of this project, mostly in text-based formats.

Access, Sharing, Archiving, and Preservation

All data produced by this project including final ecological and companion physical data and habitat data will be submitted to and managed by the Biological and Chemical Oceanography Data Management Office (BCO-DMO). Upon award PI will coordinate with BCO-DMO (discussion and coordination has already began with Robert Groman) and register project and project metadata. Species data will also be submitted to the Ocean Biogeographic Information System (OBIS) a database of global marine animal and plant distributions. COI sequences and photographs of corresponding taxonomic vouchers will be uploaded to the Encyclopedia of Life and Genebank.

Upon publication, data, metadata, and R-scripts associated with individual publications will be made available through Dryad (<http://datadryad.org/>). Dryad is an international repository of data underlying peer-reviewed articles in the basic and applied biosciences enabling scientists to validate published findings, explore new analysis methodologies, repurpose data for research questions unanticipated by the original authors, and perform synthetic studies. Dryad is governed by a [consortium of journals](#) that collaboratively promote data archiving and ensure the sustainability of the repository.

Biological samples after processing and identification will deposited at the California Academy of Sciences (see budget justification and letter of support).