CAREER: Integrating Seascapes and Energy Flow: learning and teaching about energy, biodiversity, and ecosystem function on the frontlines of climate change.

Data Management Plan

Types of data, samples, and physical collections

1. Individual species names and measurements of nekton and invertebrates as a comma separated file.
2. Stable isotope values of nekton and plant tissues as a comma separated file.
3. Physical measurements of water height, temperature, salinity, and PH as a comma separated file.
4. Images, orthomosaics, gps coordinates, and classified images will be stored in their appropriate file type on hard drivers.

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

Training

At the start of the funding period the PI, Co-PIs, technicians, and students on the project will convene a dedicated data management meeting. At this time the PI will set out naming, processing, and storage conventions for all data collected at the experimental and observational sites, as well as conduct training in annotating datasets with necessary metadata. All participants will be trained in data management best practices (Borer et al. 2009). This training will be reiterated at a yearly data management and analysis meeting, to remind participants of the conventions and train any new participants.

Data and Metadata Standards

Our metadata standards will follow lab conventions including standardization of ID labels based on sampling site and date.

All biological samples will be labelled with standardize system

Site name, Sampling date, sample number e.g. BA8216-01

All digital and drone based samples will be saved with descriptive filenames to indicate the location and date the data was taken.

Species data will be a datasheet providing ecological and taxonomic information for species in the sample including species id as well as Phylum/Class/Order/Family/Genus/Species standardized to World Registry of Marine Species taxonomy.

Stable isotope data will be a datasheet including**:** species id, Phylum/Class/Order/Family/Genus/Species, Carbon isotope value ($\delta^{13}C$)**,** Nitrogen isotope value ($\delta^{15}N$)**,** Sulfur isotope value ($\delta^{34}S$)**,** Carbon Content (%C)**,** Nitrogen Content (%N)**,** and Sulfur Content (%S)

Drone Based Imagery data will be a file folder including raw, processed, and classified image data.

Analytical Code

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 UL. UL operates onsite 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. For the biological data I anticipate generating less than 5GB of data in the course of this project, mostly in text-based formats. These data are backed up daily to offsite storage and servers. Temporary working documents, data sets, and code with version control will be kept GitHub that also allows for version recovery. All image data will be housed on solid hard drives stored in the Nelson lab at UL.

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 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.

Upon publication, data, metadata, and R-scripts associated with individual publications will be made available through Dryad). 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.