Plan Overview

A Data Management Plan created using DMPTool-Stage

Title: Data Management Plan for CAREER: Examining Kelp Forest Ecosystems Response to Interactions Between Local Disturbances and Climate Change from Local to Global Scales

Creator: Jarrett Byrnes

Affiliation: Non Partner Institution

Principal Investigator: Jarrett Byrnes

Data Manager: Jarrett Byrnes

Funder: National Science Foundation (nsf.gov)

Funding opportunity number: NSF 17-537

Template: BCO-DMO NSF OCE: Biological and Chemical Oceanography

Last modified: 07-01-2024

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customize it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Data Management Plan for CAREER: Examining Kelp Forest Ecosystems Response to Interactions Between Local Disturbances and Climate Change from Local to Global Scales

Data Policy Compliance

Identify any published data policies with which the project will comply, including the NSF OCE Data and Sample Policy as well as other policies that may be relevant if the project is part of a large coordinated research program (e.g. GEOTRACES).

The project investigators will comply with the data management and dissemination policies described in the *NSF Award and Administration Guide* (AAG, Chapter VI.D.4) and the *NSF Division of Ocean Sciences Sample and Data Policy*. This data management plan is written in accord with policies of the Biological and Chemical Oceanography Data Management Office. All data and code products of this research will be provided to the Biological and Chemical Oceanography Data Management Office, National Centers for Environmental Information, and the Ocean Biogeographic Information Service with accompanying metadata for full release either during the lifetime of the project or within two years of completion, project objective depending. Further, this data management plan follows the practices laid out by the Kelp Ecosystem Ecology Network as part of their network methods (http://kelpecosystems.org).

Pre-Cruise Planning

If the proposed project involves a research cruise, describe the cruise plans. (Skip this section if it is not relevant to your proposal.) Consider the following questions: (1) How will pre-cruise planning be coordinated? (e.g. email, teleconference, workshop) (2) What types of sampling instruments will be deployed on the cruise? (3) How will the cruise event log be recorded? (e.g. the Rolling Deck to Repository (R2R) event logger application, an Excel spreadsheet, or paper logs) (4) Will you prepare a cruise report?

Pre-event planning will be done via a planning workshop and the data management plan for all activities will be detailed in a standardized handbook of methods. This handbook contains detailed methods for each sampling event, data sheets customized by region, a field data sheet management protocol, and a data entry protocol including a chain of custody log. The actual sampling events will be recorded on paper logs which are photographed post-dive in the field to ensure no lost data (e.g., high winds or gear malfunctions on following dives) and when returned to shore are scanned into PDF documents before being archived. Data from benthic temperature loggers will be downloaded immediately following the cruise and archived using cloud data sharing services.

Description of Data Types

Provide a description of the types of data to be produced during the project. Identify the types of data, samples, physical collections, software, derived models, curriculum materials, and other materials to be produced in the course of the project. Include a description of the location of collection, collection methods and instruments, expected

dates or duration of collection. If you will be using existing datasets, state this and include how you will obtain them.

The project will produce several observational and experimental datasets, described in the list below. In addition to the datasets described below, the project will produce data on temperature and wave heights. All data and the code from the data processing pipeline will initially be made available via Github for version control. The data include

1) Site description, logger deployment records, and diver sampled abundance and point count data from surveys before and after kelp manipulations from before and after experimental kelp manipulations.

2) Temperature data during the year after the kelp removal data from HOBO temperature loggers.

3) Modeled wave height data from the Global Ocean Wave model of Reguero et al. 2012 for experimental removal sites.

3) Site description, logger deployment records, and diver sampled abundance and point count data from observational surveys.

4) Temperature and wave pressure measurements from sensors deployed at each observational survey site.

All diver data sheets will be scanned and entered data will be archived in CSV format before cleaning and processing. Data from observational survey sites should span the five years of the project. Data from experiments will be collected throughout the duration of the five years of the experiment, with data from different regions coming in at different times.

Data and Metadata Formats and Standards

Identify the formats and standards to be used for data and metadata formatting and content. Where existing standards are absent or deemed inadequate, these formats and contents should be documented along with any proposed solutions or remedies. Consider the following questions: (1) Which file formats will be used to store your data? (2) What type of contextual details (metadata) will you document and how? (3) Are there specific data or metadata standards that you will be adhering to? (4) Will you be using or creating a data dictionary, code list, or glossary? (5) What types of quality control will be used? How will data quality be assessed and flagged?

Field dive sampled data will be photographed in the field and scanned immediately on return from a day of sampling. Photos and scans will be saved on a network-wide Dropbox (http://dropbox.com) account in folders labeled with protocol, network member name, site, and date. Data will be entered using Excel using standardized already developed templates with imposed data validation to minimize data entry error. Data will be quality control-checked using read-back methodology in which sheets are read aloud while a member validates each entry. Data from sensors is all saved in long CSV formats with each measurement date-time stamped.

Upon completion of the quality control process, data is sent to the Byrnes lab for centralized data management. Data sets are run through scripted quality control checks ensuring completeness, scans

for data entry outliers, and taxonomic standardization of species lists against the World Registrity of Marine Species (http://www.marinespecies.org/). Any failures are turned back to the network member for correction. Temperature and pressure sensor data are similarly analyzed for outliers indicitive of sensors not being deployed during member specified dates.

Data is then merged into a common set of long-format CSV files suitable for futher use and analysis. The entirity of the data cleaning pipeline, including code, raw data, and cleaned merged data, is hosted on the version control archive GitHub (e.g., https://github.com/kelpecosystems/observational_data).

Upon archiving (see below), the PI will generate metadata using the ISO19115 compliant Marine Community Profile 2.0 metadata schema (https://marinemetadata.org/references/marineprofile19115) via a user interface at TemperateReefBase.

Data Storage and Access During the Project

Describe how project data will be stored, accessed, and shared among project participants during the course of the project. Consider the following: (1) How will data be shared among project participants during the data collection and analysis phases? (e.g. web page, shared network drive) (2) How/where will data be stored and backed-up? (3) If data volumes will be significant, what is the estimated total file size?

The investigators (PI and network members) will store original project data (including spreadsheets, CSV and ASCII files, images, and scans of data sheets) on laboratory computers. In the Byrnes lab, all computers backup both to independent hard drives regulated by Apple Time Machine and to the cloud via a lab Dropbox (http://dropbox.com). Scans of data sheets will be sent to the Byrnes lab upon completion of each site for immediate redundancy and backed up in Dropbox. Once each member completes data entry and quality control, Excel or CSV data files will be sent to the Byrnes lab and immediately 1) be placed into a cloud dropbox and 2) be added to the version controlled raw data portion of the project's public (for observational data) and private (for experimental data) Github account.

Mechanisms and Policies for Access, Sharing, Re-Use, and Re-Distribution

---

Describe mechanisms for data access and sharing, and describe any related policies and provisions for re-use, re-distribution, and the production of derivatives. Include provisions for appropriate protections of privacy, confidentiality, security, intellectual property, or other rights or requirements. Consider the following: (1) When will data be made publicly available and how? Identify the data repositories you plan to use to make data available. (2) Are the data sensitive in nature (e.g. endangered species concerns, potential patentability)? If so, is public access inappropriate and how will access be provided? (e.g. formal consent agreements, restricted access) (3) Will any permission restrictions (such as an embargo period) need to be placed on the data? If so, what are the reasons and what is the duration of the embargo? (4) Who holds intellectual property rights to the data and how might this affect data access? (5) Who is likely to be interested in re-using the data? What are the foreseeable re-uses of the data?

Pre-publication Data Management

Observational data sets are immediately available within and outside of the network upon deposition via the public Github archive (or other version control repository, should something happen with Github). Once quality controlled the Each member's dataset is available under CC-BY attribution for immediate use.

Experimental data is available within the network immediately via a private Github Archive with the understanding that members are to be included as co-authors on the first publication using this data. After publication, or two years from the termination of this grant, this archive shall be made entirely public.

Data Sharing Policies

Broadly, KEEN is interested in open and transparent data sharing while ensuring that members get proper credit. Our data sharing policies are as follows.

1) Experimental manipulation data is private until publication or two years after the completion of this grant. Members are assured co-authorship on the first publication using their data, at which point it becomes public. Data is made availabe under the CC-BY license with the additional stipulation that members be queried and offered potential co-authorship if their manipulation data is to be used for later publication. However, they may only claim authorship if they are willing to engage in a significant intellectual contribution to the piece, defined by doing at least two of the following collaborating on ideas, analysis, writing, and editing; merely correcting typograpical errors or not offering any feedback is seen as insufficient for authorship.

2) Observational data is made public as soon as it quality controlled data is merged into the network data set. Inividual member data sets will be made citable (see below) and are offered via a Creative Commons-By license. This ensures members get credit for their work. It also ensures that observational data is immediately available and relevant to the scientific and management communities.

3) Non-KEEN datasets that are reshaped into KEEN format will be made available separately and their use must correspond to the original authors data sharing policies. For example, while we will provide scripts and instructions to merge the SBC-LTER data into KEEN data, they have requested that their data not be made available as part of the main dataset so that they might maintain control and monitor access as is required by their funding.

Long-Term Data Sharing and Archiving with TemperateReefBase

The Kelp Ecosystem Ecology Network is one of the partner organizations aiding in the creation of TemperateReefBase (http://temperatereefbase.imas.utas.edu.au/), a data warehouse funded by the Australian Integrated Marine Observing System (IMOS) to host geospatially specific measurements of biotic and abiotic condition of temperate reefs around the globe. Already, TRB hosts the KelpTime kelp timeseries dataset from the Krumhansl et al. 2016 paper, in addition to a wide variety of data sets from other organizations.

TRB requires project scientists to generate metadata using the ISO19115 compliant Marine Community Profile 2.0 metadata schema via an intuitive user interface on intitial data creation. Metadata is maintained through data update cycles. Individual data sets will also recive a DOI for citation.

As quality controlled data is merged into the network dataset, we will archive it at TRB. Upon archiving materials at TRB for the first time, we will generate metadata for all data products.

Data Sharing via BCO-DMO

After deposition with TemperateReefBase and metadata generation, I will work with the BCO-DMO to archive data with the National Centers for Environmental Information. Data, samples, and other information collected under this project can be made publically available without restriction once submitted to the public repositories. Machine readable metadata will have already been generated by TRB, thus making archiving and access provision straightforward. As the Ocean Biogeographic Information Service begins to add abundance data to its list of available data types, I will work with IOBIS to add our records of species abundances from the observational sampling data.

Plans for Archiving

Describe the plans for long-term archiving of data, samples, and other research products, and for preservation of access to them. Consider the following: (1) What is your long-term strategy for maintaining, curating, and archiving the data? (2) What archive(s) have you identified as a place to deposit data and other research products?

The PI will ensure that the original underway measurements are archived permanently at TemperateReefBase, the National Centers for Environmental Information, and/or the Ocean Biogeographic Information Service as appropriate. BCO-DMO will also ensure that project data are submitted to the appropriate national data archive. The PI will work with TRB and BCO-DMO to ensure data are archived appropriately and that proper and complete documentation are archived along with the data.

Roles and Responsibilities

Describe the roles and responsibilities of all parties with respect to the management of the data. Consider the following: (1) If there are multiple investigators involved, what are the data management responsibilities of each person? (2) Who will be the lead or primary person responsible for ultimately ensuring compliance with the Data Management Plan?

The lead PI is ultimately responsible for the compliance with the data management plan. Already, PI Byrnes has provided members with data management plans, data entry templates, chain of custody logs, and data archival structures that are compliant with the DMP.

Members are responsible for management of their initial raw data sets and are asked to send quality controlled data to PI Byrnes as soon as it is complete. PI Byrnes has created an automated data quality control and merging pipeline for the observational data set, and will work the the Postdoc to create one for the experimental data set.

During year one of the project, PI Byrnes will handle all data coordination and management, ensuring that data management pipelines are setup. After year one, the project postdoc will handle data management duties in consultation with PI Byrnes. The postdoc will also be responsible for creating software to translate other data sets to the KEEN format and to manage the dissemination of those data sets where possible to help them gain pracitcal experience in data managament as career development (see Postdoctoral Mentoring Plan).

For long term archiving, the PI with work with the BCO-DMO, TRB, and OBIS to ensure proper data deposition and access.