Examples of Data Management Plans

Stephanie Hampton
National Center for Ecological Analysis and Synthesis
University of California, Santa Barbara
Experimental lab data

**Background**

*Project name*: Effects of temperature and salinity on population growth of the estuarine copepod, *Eurytemora affinis*

**Description of project aims and purpose:**

We will rear populations of *E. affinis* in the laboratory at three temperatures and three salinities (9 treatments total). We will document the population from hatching to death, noting the proportion of individuals in each stage over time. The data collected will be used to parameterize population models of *E. affinis*. We will build a model of population growth as a function of temperature and salinity. This will be useful for studies of invasive copepod populations in the Northeast Pacific.
Experimental lab data

1. Information about data

Every two days, we will subsample *E. affinis* populations growing at our treatment conditions. We will use a microscope to identify the stage and sex of the subsampled individuals. We will document the information first in a laboratory notebook, then copy the data into an Excel spreadsheet. For quality control, values will be entered separately by two different people to ensure accuracy. The Excel spreadsheet will be saved as a comma-separated value (.csv) file daily and backed up to a server. After all data are collected, the Excel spreadsheet will be saved as a .csv file and imported into the program R for statistical analysis. Strasser will be responsible for all data management during and after data collection.
Experimental lab data

1. Information about data

Our short-term data storage plan, which will be used during the experiment, will be to save copies of 1) the .txt metadata file and 2) the Excel spreadsheet as .csv files to an external drive, and to take the external drive off site nightly. We will use the Subversion version control system to update our data and metadata files daily on the University of Alberta Mathematics Department server. We will also have the laboratory notebook as a hard copy backup.
Experimental lab data

2. Metadata format and content

We will first document our metadata by taking careful notes in the laboratory notebook that refer to specific data files and describe all columns, units, abbreviations, and missing value identifiers. These notes will be transcribed into a .txt document that will be stored with the data file. After all of the data are collected, we will then use EML (Ecological Metadata Language) to digitize our metadata. EML is one of the accepted formats used in Ecology, and works well for the type of data we will be producing. We will create these metadata using Morpho software, available through the Knowledge Network for Biocomplexity (KNB). The documentation and metadata will describe the data files and the context of the measurements.
Experimental lab data

3. Policies for access, sharing, and reuse

We are required to share our data with the CAISN network, after all data have been collected and metadata have been generated. This should be no more than 6 months after the experiments are completed. Interested parties must contact the CAISN data manager (data@caisn.ca) or the authors and explain their intended use. Data requests will be approved by authors after review of the proposed use.
Experimental lab data

3. Policies for access, sharing, and reuse

The authors will retain rights to the data until the resulting publication is produced, within two years of data production. After publication (or after two years, whichever is first), the authors will open data to public use. After publication, we will submit our data to the KNB allowing discovery and use by the wider scientific community. Interested parties will be able to download the data directly from KNB without contacting the authors, but will still be required to give credit to the authors for the data used by citing a KNB accession number either in the publication’s text or in the references list.
Experimental lab data

4. Long-term storage and data management (archiving)

The data set will be submitted to KNB for long-term preservation and storage. The authors will submit metadata in EML format along with the data to facilitate its reuse. Strasser will be responsible for updating metadata and data author contact information in the KNB.
Experimental lab data

5. Budget

We will use a tablet computer for data collection, which will cost approximately $500. We anticipate that data documentation and preparation for reuse and storage will require approximately one month of salary for two technicians. These technicians will be responsible for data entry, quality control and assurance, and metadata generation. These costs are included in the budget in Lines 10-15.

Example from Carly Strasser
Image data

Background

This project generates time- and location-stamped image files of natural resources in Delaware County, PA. The images serve as a record of the occurrence of creatures, natural artifacts, and conditions at specific places and times during the period 2003 through 2011. For many of the photos taxonomic information is also available. The occurrence data are observational and qualitative, and in representing the presence of various taxa will be used for posters, books, and a web site promotion of the natural diversity of the County and the preservation of its remaining natural resources.
Image data

1. Information about data

Data will be captured with a digital camera capable of creating images with sufficient taxonomic detail to allow identification to the species level. Images are stored as JPG files with embedded EXIF and IPTC information describing the exposure, camera type, lens, and metering mode, along with photographer-generated cataloging tags that may include taxonomic identity of the organism(s) in the photograph. Images will be stored in a date-hierarchical file structure (year, date) on redundant disk drives and managed using the commercial iMatch software. Images will be maintained with unique file names. Quality flags are used to differentiate images of varying quality. J. Smith will be responsible for the data management during the course of the project.
2. Metadata format and content

Metadata about timing and exposure of individual images is automatically generated by the camera at the time the photo is taken. GPS locations are subsequently added by post-processing GPS track data based on shared time stamps. GPS data stored in image files depicting rare or locally sensitive species will be obfuscated in the file metadata but can be made available for appropriate, approved uses. Metadata for the image dataset as a whole will be generated by the image management software (iMatch) and will include time ranges, locations, and a taxon list. Those metadata will be translated into Ecological Metadata Language (EML), created using the Morpho software tool, and will include location and taxonomic summaries.
Experimental lab data

3. Policies for access, sharing, and reuse

The image collection will be made available beginning in 2015. They will be available as digital photographs viewable on the web in a restricted form that prevents downloading. Summaries of temporal and spatial distributions of taxonomic groups will be also be made available on Smith’s website for use without permission, but users will need permission to access the original high resolution photographs from which these distribution data were derived. Smith will retain copyright and the originals will be licensed using a Creative Commons license (Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License). Data can be cited by referring to the image website.
Experimental lab data

4. Long-term storage and data management (archiving)

Long term storage will involve exporting the metadata stored within the JPG files (location, EXIF, and tag information) into text files to facilitate access to that metadata and as a safeguard should the specifics of JPG metadata structure change. The image files will be stored in a single zip file containing the date-hierarchical file structure. As noted above, EML metadata will be created and the resulting dataset will be submitted to the Knowledge Network for Biocomplexity (KNB) archive.