An example Data Management Plan

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

Author: Carly Strasser (*strasser@email.com*)

Project description

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.

Method description

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 sub-sampled 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.

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.

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 alongside the data file.

Access and Use Policies

We are required to share our data with the Canadian Aquatic Invasive Species Network (CAISN), 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.

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 Knowledge Network for Biocomplexity (KNB) data repository 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.

Long-term storage

The data set will be submitted to KNB for long-term preservation and storage. The authors will submit metadata in Ecological Metadata Language (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.

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.