Data Management Plan

I.   Products of Research

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.

II.   Data Storage and Preservation

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.

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.

III.   Data Formats and Metadata

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 on 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.

IV. Data Dissemination & Policies for Data Sharing and Public Access

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. In order to gain access to CAISN data, 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 the 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 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.

V. Roles and Responsibilities

The PI will head the implementation and monitoring of the data management plan procedures.