This is an example Data Management Plan created by participants of a DataONE best Practices Workshop.

Data Management Plan

I. Products of Research

We will collect insects annually from the 30 experimental plots at each of the eight sites (see body of proposal for sampling details). Samples will be immediately deposited in sealable containers labeled with the date, site code (already existing), block, plot, and subsample. An associated record of any observations or notes will be entered in a field tablet computer and labeled with the same information. We will also record environmental information including temperature and general observations. Labeled samples will be transported back to the laboratory, where they will be sorted and identified using a dissecting microscope. We will identify and count the arthropods to the classification of order, with the exception of members of the order Auchenorrhynca, which will be identified to species or morphospecies. Identifications will be reviewed by multiple researchers associated with the project and verified with the assistance of Stuart McKamey of the Systematic Entomology Laboratory of the USDA Agricultural Research Service. Representatives of the identified species and morphospecies will be vouchered to the Bell Museum of Natural History at the University of Minnesota (U of M).

Abundance for each group will be recorded by hand in a laboratory notebook during sorting. These data will be transcribed into an Excel spreadsheet as each sample is completed. The spreadsheets will be stored on a controlled-access U of M server directory that is backed up offsite nightly. Files will be named according to the format site_mmddyyyy_plot.csv using existing unique site codes. Lind will be responsible for the data during and after data collection until publication.

After identification, Arthropod samples from each experimental plot will be subsampled and sent to the University of St. Thomas Kay lab for stoichiometric analysis. We will receive a spreadsheet of data after processing is complete. This spreadsheet will include the insect identification (including site code, date, year, plot, and arthropod identification) and percent by mass of carbon, phosphorus and nitrogen. These files will be saved as csv files in the previously described server directory.

Our data set will be used in combination with the existing Nutrient Network (nutnet.unm.edu) data on plant responses to nutrient manipulation. The NutNet data is currently stored and managed in a MySQL relational database housed at the Minnesota Supercomputing Institute and accessed through a secure internet connection. We will add our data and metadata to the NutNet relational database. The existing csv files will be read into temporary tables in the MySQL database, and then inserted into permanent data tables using insert query statements. The existing database schema links tables of data observations to a “plot” table describing the experimental unit. New tables will be created for each of the arthropod data types (abundance and stoichiometry) containing the unique plot identifier. Multiple tables may be necessary for efficient data storage and management; for example, an “Arthropod” table holding scientific names for use can be used to constrain the labels of abundance records to acceptable possibilities.

II. Data Storage and Preservation

The writer software will be preserved by the HDF Group for the life of the HDF libraries. The HDF Group uses industry-standard best practices to ensure the integrity of their software and systems. Once the map writer has been used to generate maps for every HDF file in existence, the continued existence of the writer software is not required. The reader software will be preserved at SourceForge.org for as long as there is community interest. The collection of HDF files will be preserved at NSIDC as long as utility is deemed high.

III. Data Formats and Metadata

The maps that will be produced are self-describing and require no documentation. Once created, the maps will become part of the data product, and will be included as one component of the metadata for the HDF product.

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

This proposed project will create ancillary metadata that will enable future users to read all HDF (Hierarchical Data Format) formatted data. The reader software is open source (licensed under GPLv3) and is freely available. The HDF data at NSIDC is freely available with no restrictions. The collection of HDF files has a suggested citation, which is described in the metadata.

The user community for the map writer is expected to be any repository with HDF data. The user community for the map reader is expected to any user of HDF data in the future who wishes to re-use or adapt map-reading software to their own purposes. The user community for the collection of HDF files is expected be the community of tool developers who need a wide range of data types with which to test their product.

V. Roles and Responsibilities

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