Data Management Plan
Project: The influence of plant functional types on ecosystem responses to altered rainfall
PI: Elsa Cleland (UCSD), Co-PI: David Lipson (SDSU), Co-PI: John Kim (SDSU)

Training
At the start of the funding period the PIs, senior personnel, technician and students on the project will convene a dedicated data management meeting. At this time the PIs will set out naming, processing and storage conventions for all data collected at the experimental and observational sites, as well as conduct training in annotating datasets with necessary metadata. All participants will be trained in data management best practices (*e.g.* Borer *et al.* 2009). This training will be reiterated at a yearly data management and analysis meeting, to remind participants of the conventions and train any new participants.

Collection
Datasets to be collected in the experimental and observational components of the research are described in the Project Description. Most datasets will be collected 1-3 times per year (i.e. production and decomposition, ecophysiological functional traits, soil extractable nutrients and mineralization rates, soil microbial community composition and function). Temperature, light availability and soil moisture at multiple depths in the experiment will be logged every 15 minutes, these data will be stored on local data loggers and downloaded every two weeks.

Processing
Data originally recorded on physical paper datasheets will be transferred each day into spreadsheets using non-proprietary software (*e.g.* open office platforms stored as ASCI files, .txt or .csv formats). Data will immediately be checked for outliers in the R statistical program, and any outliers will be checked against the paper copies for transcription errors. Paper copies will be kept on file for at least 10 years. All field-collected data will be stored on the Division of Biological Sciences Twinlake server which is backed up daily and is duplicated on servers in two separate buildings on campus. Data collected by project staff based at SDSU will be stored on the Field Stations Program Ringtail data server, which features an expandable RAID system configured to store data with redundancy across multiple drives. Furthermore, for added redundancy, coordination, and availability all project data will be shared (copied) periodically between UCSD and SDSU, in a manner consistent with university policy.

Analysis
Data analysis of field collected data will be scripted and extensively annotated in the R statistical program and the scripts will be stored on the server along with the datafiles. DGVM simulation runs will be performed on a high performance parallel computing platform, a 96-node Linux cluster, maintained jointly by USFS Pacific Northwest Research Station and Oregon State University. Simulation output is in NetCDF format, a data format popular in climate research, readable with many free software programs. DGVM output will be analyzed and displayed with the ESRI ArcGIS software suite.

Documentation
All datasets will be annotated with meta-data. As data are generated they will be entered into Morpho, a free resource for associating Ecological Metadata Language (EML) with archived

datasets. It will be the responsibility of each researcher to annotate their data with metadata, and it will be the responsibility of the PIs to check weekly (during the field season, monthly otherwise) with all participants to assure data is being properly processed, documented, and stored.

Data Products, Curation and Data-Use Policy

All raw data will be made freely available by the time of publication or the end of the funding period, consistent with NSF policy. When data are associated with a publication, the raw data and associated analysis R scripts will be archived in a source such as Ecological Archives or Dryad. DGVM outputs from the project will be archived in a recognized national data center, such as the ORNL DAAC, ESRI DataBasin, or UNH EOS-Webster, making data freely available for download. At the end of the funding period all datasets will be entered together as a data package into the KNB database (Knowledge Network for Biodiversity) and/or the DataOne database, with the goal of archiving the data together for perpetuity. Prior to the end of the funding period data will be made available by request with the stipulation that if the data are used in publication then the researchers that collected the data need to be informed of the planned use and be offered authorship as appropriate.

Past activities

All senior personnel have been active in efforts to efficiently document and synthesize ecological data. As a postdoctoral researcher at the National Center for Ecological Analysis and Synthesis (NCEAS) PI Cleland participated in the SEEK-BEAM working group (Science Environment for Ecological Knowledge - Biodiversity and Ecological Modeling and Analysis), an ecoinformatics effort to test the implementation of Morpho and other ecoinformatics tools developed at NCEAS. She also led an NCEAS Distributed Graduate Seminar to teach data management and synthesis techniques in the field of Functional Ecology. Collaborator John Kim has worked as data manager for SDSU Field Stations Program and for SDSU Global Change Research Group. He has also taught data management workshops sponsored by NSF: for postdocs and new faculty through SEEK, and for field stations staff at RCN Resource Discovery Initiative for Field Stations Training. Finally, both PIs have a demonstrated commitment to publishing data and associated meta-data. PI Cleland published all data associated with her synthetic work at NCEAS (Cleland et al. 2008), and CoPI Lipson submitted all data from a recently completed NSF grant (OPP 0421588) to the ARCSS site at EOL (http://www.eol.ucar.edu/projects/arcss/).

Note: For references please see main references document