

INCREASING THE VALUE OF ECOLOGICAL DATA WITH STRUCTURED METADATA: An introduction to Ecological Metadata Language (EML) and principles for metadata creation

Eric Fegraus, Sandy Andelman
National Center for Ecological Analysis & Synthesis
12 April, 2004

Introduction

The long-term value of ecological data, and its utility for advancing ecological understanding and solving important environmental problems depend on the availability of suitable and adequate metadata, or descriptive information describing data content, context, quality, structure and accessibility (Michener 2000). As a discipline, ecology is moving beyond its tradition of small-scale empirical observations and experiments conducted by one or a few investigators at relatively small scales (Kareiva & Anderson 1988; Brown & Roughgarden 1990). Increasingly, ecological and environmental research aims to understand complex ecological systems at broad spatial and temporal scales and requires access to data beyond what is typically contained within a single data collection effort. This changing focus has led to recognition within the community of the need for increased data sharing and long-term community access to data (e.g., Olson & McCord 2000; Andelman et al. 2004) and presents new challenges for integration of heterogeneous ecological information across a range of spatial, temporal and organizational scales.

Historically, investigations of many ecological phenomena, and the development of theory to explain them, have been limited by the availability of suitable long-term data. For example, the fragmentary nature of population dynamics research activities, which often has focused on analysis of individual data sets, has made it difficult to formulate general theory and to investigate large-scale spatial and taxonomic patterns. In response to this limitation, and motivated by the need to provide comprehensive access to biological population data, to facilitate discovery of general patterns and principles, to advance understanding of large-scale spatial and temporal patterns, and to enable researchers to acquire large numbers of datasets, without having to undertake repetitive, time-consuming and expensive searches, the Center for Population Biology at Imperial College, Silwood Park, the National Center for Ecological Analysis and Synthesis (NCEAS), and the University of Tennessee collaborated to develop the Global Population Dynamics Database (GPDD; <http://cpbnts1.bio.ic.ac.uk/gpdd/>). The GPDD is now the largest collection of animal and plant population data in the world, and brings together nearly five thousand time series in one database. It provides an important resource for ecologists, resource managers and environmental scientists interested in the dynamics of natural populations or in asking comparative questions about the nature of population variability (e.g., Kendall et al. 1998; Kendall et al. 2000; Fagan et al. 2001; Inchausti & Halley 2001; Inchausti & Halley 2002; Inchausti & Halley 2003).

Synthetic efforts such as development of the GPDD are limited by the heterogeneity of ecological data and metadata. Ecological data exhibit a range of formats, reflecting different underlying motivations for data collection, different suites of variables and different spatial and temporal sampling designs. In addition, ecological metadata typically also are highly variable in extent, depth, quality and associated types of uncertainty (Andelman et al. 2004; Regan et al. 2002), and may consist of mental notes, hand-written notes in a field notebook, a comments field in an Excel spreadsheet or other forms of documentation. Currently, there is no single standard to guide the decision about what quantity and quality of metadata are sufficient to enable data that initially were collected for a single, relatively narrow purpose, to be understood and used appropriately for a variety of purposes. Unfortunately, this often means that the value of ecological data diminish over time, because important details about the data may be forgotten or lost by the original investigator, because of career changes, or changes in data storage and management technology (Michener 2000).

This paper describes Ecological Metadata Language (EML; <http://knb.ecoinformatics.org/software/eml/>), a method for formalizing and standardizing the set of concepts that are essential for describing ecological data, and explains why and how creating metadata with EML will extend the long-term utility of ecological data and facilitate the processes of data discovery and integration.

Metadata

Metadata is the information that describes “who, what where, when, why and how” an ecological dataset was collected. Metadata literally means (meta ~ about) “about data”. Most of us have experienced the difficulty in using our own data after only a few months have passed since it was collected. Unless data are adequately documented, this difficulty only increases over time. Even the simplest analysis requires some level of metadata. For example, consider a simple data table, with no column headers (Table 1). Without metadata, a data table such as this one is useless. Unless we know the measurement units, the numbers in the columns are meaningless. Further, in this example, there is no metadata that would identify the location where data were collected, the focal organism or system, or the identity and location of the data owner. Table 2 shows the same data table, with some additional metadata. In this example we can see that data were collected in May, 2002, at a site, identified as VO. However, one can only guess at what VO might signify, or at the meaning of the columns with headers “S,” “R,” “Bm,” “P,” and “N.” signify. The lack of adequate metadata makes this dataset useless to anyone other than the original owner.

Table 3 illustrates a dataset with more comprehensive documentation. The data owner is identified, column headers are defined, and some general information is provided regarding how and where the data were collected. Depending on the intended use, access to this information may make this a useable dataset. However, the original owner of this data probably has additional information that would facilitate a more comprehensive description of the data and potentially make it more suitable for a broader range of research activities in the future. For example, inclusion of geographic coordinates or other information that describes the spatial location where the data were collected, or information about codes or numbers used to indicate missing data would increase the potential utility of these data.

This example illustrates another common feature of metadata; the information provided is what the data owner decided to document. Without standards or guidelines for metadata content, if the same individual were to document another dataset, the same or different information might be included, and the format might or might not be the same. In this way, if analyses require multiple datasets from different owners, locations or times, it is unlikely that all relevant datasets would have metadata with equivalent levels of detail, use consistent terminology, or use consistent formats for metadata. The potentially infinite formats and types of information in which to document data suggest the need for metadata to be standardized. Table 4 provides an example of a much more detailed metadata document that was made using EML, in which each metadata concept (from dataset title to geographic description) has been formalized and standardized.

Creating Metadata with EML

NOTE: This for non-RDBMS datasets. Contact jones@ecoinformatics.ucsb.edu for more information.

Ecological Metadata Language (EML) is a method for formalizing and standardizing the set of concepts that are essential for describing ecological data, as well as the format for recording this information. EML evolved from an

open source, community-based effort involving ecological researchers, information managers, and software developers, led by the National Center for Ecological Analysis and Synthesis (NCEAS) and the Long Term Ecological Research Network (LTER). The need for EML or a similar method to promote preservation and long-term utility of the growing archives of ecological data has been recognized for some time (FLED Report 1995, Michener et al, 1997; Olson & McCord 2000).

EML is intended for use by any ecologist or manager of ecological information. It describes a range of essential aspects of ecological data, such as data attribute (usually thought of as variables by ecologists) names and definitions; units of measurement; date, time and location of data collection; who collected the data sampling design; etc. EML attempts to reduce ambiguity and uncertainty by formalizing these metadata concepts into a comprehensive yet standardized set of terms and definitions intended specifically for ecological data. The metadata in Table 4 provide an example of dataset that has been reasonably well-documented using EML.

The question of “How much metadata is enough?” does not have a clear-cut answer. As a general guideline, an ecologist should be able to get a thorough understanding of a moderately complicated dataset after reviewing the metadata for 20 minutes. If in doubt, assume that “more is better,” because omitting detail from metadata at the outset may lead to problems later on (e.g., hours of discussion or exploratory analyses), and in the worst case may render the data unusable. In general the more metadata you create the longer the lifespan of your dataset.

A walk through each of the formalized EML metadata concepts shown in Table 4 will clarify some of the more important metadata concepts provided in EML. The information in Table 4 is arranged in six broad metadata categories, each of which contains more detailed metadata. These categories are somewhat arbitrary, but are intended to categorize EML metadata fields in a way that is intuitive to ecologists. The categories include the General Dataset, Geographic, Temporal, Taxonomic, Methods and Data Table Metadata sections.

The General Dataset Metadata category contains EML metadata concepts that describe the purpose of the data collection and the questions the data were originally intended to address being collected to address. Some types of metadata, such as the title and abstract are self-explanatory, but others may not be. The *usage rights* field provides a place for information about who can use the dataset, and what, if any, restrictions there are on usage. Other general dataset metadata information includes contact information for people who had a significant role in collecting or managing the data. At least one primary contact is recommended. This should be the person to whom further questions regarding the data and metadata should be addressed.

Also, in this category, information about the data or data collection methods may be provided in published papers. EML provides fields for entering bibliographic information such as citations to journal articles or books. EML supports a range of standard reference styles, and permits importation directly from EndNote and other bibliographic software.

As the name suggests, the Geographic Metadata category is used for geographic and spatial metadata. The geographic description field contains information about where the research project took place, where samples were collected and any spatial or geographic references that may provide a context for the data. Latitude and longitude may also be entered here to increase geographic accuracy.

The Temporal Metadata category contains information about when the research project started and ended. Information can be stored either as a range of dates (e.g. data was collected every month between June 2002 & 2003) or specific time periods (e.g. May 2002 and June 2003). In addition, information about potential gaps in data collection or in the collection of some variables

If the dataset has species information there are Taxonomic Metadata fields that describe them. Information such as the taxonomic authority (i.e. the book or system that is used to identify a species) and the taxonomic class (i.e. Family, Genus, Species) can be described here. To facilitate entry of taxonomic information, tables or text files containing lists of species names can be uploaded as a single file.

The Methods Metadata category contains information on the methods used for data collection. General methodological information such as the experimental design and the machines or devices used to collect data can be described here. The methods documentation should be sufficiently detailed to allow someone to recreate the research project.

The Data Table contains information regarding the data table itself. There is physical data information such as the file name, whether or not the letters in the data table are case sensitive, the number of records and the structure of the data table (i.e., attribute names in columns or rows). This category also contains metadata regarding the columns of data themselves. The *label* is a word or phrase that describes the column, because acronyms or ambiguous abbreviations often are used as column headers. The “definition” field contains more descriptive information, indicating what the numbers in individual columns or rows represent. The *unit* and *type* fields contain the units and data type (e.g., integer, floating point, etc.) for each column. *Missing* represents a number or symbol used to indicate that no data was collected (e.g., 9999). *Precision* pertains to the accuracy of the measurement. For example, if the numbers in a column represent output from a machine, there is often some level of precision or measurement error associated with the data. Similarly, an investigator may have estimates of the precision associated with a particular data collection method. The *attribute description* column provides definitions of any codes used in that column (e.g. VO = Vally Oak) and the range of values in a column (e.g. biomass values range from 10.04g/ m² to 88.82g/m²).

The metadata contained in Table 4 represent a minimum level of detail that would be required by someone with little or no prior knowledge about the dataset (or yourself after not working with the data for a few years) to determine whether or not the data are appropriate for some intended use. Additionally, once someone has decided to use the data for a particular purpose, the metadata should be sufficient to enable the next research steps (e.g., contact the data owner for the dataset, or if the data are public and accessible, begin preliminary analyses).

Documenting Data With Ecological Metadata Language (EML)

At this point you have a basic understanding of the advantages of using EML, and you have seen an example of a reasonably well-documented dataset. Currently, there are two mechanisms for creating EML to document your data: Morpho, and web registries.

Morpho

Morpho is a cross-computer platform data and metadata management software program. It enables an ecologist to create, edit and manage metadata and data tables. Morpho also provides capabilities to search and query ecological data, both locally and remotely, on publicly accessible ecological data archives (i.e. ecological archives accessible via the internet). Morpho includes wizards that facilitate using a subset of EML (e.g. Table 4) to document the most common attributes of your data. In addition there are tools that provide access to the entire contents of EML, which currently include over 2000 metadata concepts or terms for describing ecological data. For more information see <http://knb.ecoinformatics.org/software/morpho>.

Another option is to document your data using a subset of EML through a web registry at <http://knb.ecoinformatics.org/index.jsp>. To use this tool you first must register (i.e., provide basic information about yourself and how you may be contacted). Then you can create EML compliant metadata without installing Morpho. As when using Morpho, the user can create EML metadata and make it broadly available to the ecological community through the internet. Making EML metadata available facilitates the discovery of your dataset by ecologists around the world. Additionally, large organizations and large research projects may want to create customized EML web interfaces (contact jones@nceas.ucsb.edu for information about how to do this). Currently, the web registries provide a mechanism for creating and querying metadata, however they do not provide direct access to the data, which requires Morpho.

The Future

By systematically documenting your data in a standardized and structured format, you will contribute to advancing ecological knowledge. As ecological data and metadata archives grow, the value of these resources to the ecological community will increase. EML provides a *common structure* that ecologists can use to document, share and interpret ecological data. The formal structure of EML also will facilitate the development of software applications that process the metadata. EML is implemented in XML (Extensible Markup Language), a growing standard for marking up documents on the Web. This means that EML metadata eventually will enable the use of a wide range of software, from basic search and query tools that can be used remotely through the web, to remote integration of heterogeneous datasets, to analysis and visualization. For more information about EML and tools for creating metadata and sharing data, go to <http://knb.ecoinformatics.org/index.jsp>.

Literature Cited

Andelman, S. J., Bowles, C. M., Willig, M. R. & Waide, R. B. (2004). Understanding environmental complexity through a distributed knowledge network. *Bioscience* 54, 240-246.

Brown JH, Roughgarden J. 1990. Ecology for a changing earth. *Bulletin of the Ecological Society of America* 71: 173-188.

Fagan, W. F., Meir, E., Prendergast, J., Folarin, A. & Karieva, P. (2001). Characterizing population vulnerability for 758 species. *Ecology Letters* 4, 132-138.

Gross, K et al. Report of the Committee on the Future of Longterm Ecological Data (FLED), Ecological Society of America. Ecological Society of America, Washington, D.C., 1995.

Inchausti, P. & Halley, J. (2001). Investigating long-term ecological variability using the global population dynamics database. *Science* 293, 655-657.

Inchausti, P. & Halley, J. (2002). The long-term temporal variability and spectral colour of animal populations. *Evolutionary Ecology Research* 4, 1033-1048.

Inchausti, P. & Halley, J. (2003). On the relation between temporal variability and persistence time in animal populations. *Journal of Animal Ecology* 72, 899-908.

Karieva, P, Anderson, M. (1988) Spatial aspects of species interactions: The wedding of models and experiments. In A Hastings (Ed.) *Community Ecology* (pp. 35-50). New York: Springer-Verlag.

Kendall, Bruce E.; Briggs, Cherie; Murdoch, William W.; Turchin, Peter; Ellner, Stephen P.; McCauley, Edward; Nisbet, Roger M.; Wood, Simon. 1998. Why do populations cycle?: A synthesis of statistical and mechanistic modeling approaches. *Ecology*. Vol: 80. Pages 1789-1805.

Kendall, Bruce E.; Bjornstad, Ottar N.; Bascompte, Jordi; Keitt, Timothy H.; Fagan, William F.. 2000. Dispersal, environmental correlation, and spatial synchrony in population dynamics. *American Naturalist*. Vol: 155. Pages 628-636.

Michener, William and J. Brunt, eds. (2000) *Ecological Data—Design, Management and Processing*, Blackwell Science, Malden, MA:

Michener, William K., James W. Brunt, John Helly, Thomas B. Kirchner, and Susan G. Stafford. 1997. Non-GeoSpatial Metadata for the Ecological Sciences. *Ecological Applications*. 7(1):330-342.

Olson, RJ & McCord, RA. (2000) Archiving Ecological Data and Information. In: W. Michener & J. Brunt (Eds.), *Ecological Data – Design, Management and Processing* (pp. 117-141). Malden, MA: Blackwell Science.

Regan, H. M., Colyvan, M. & Burgman, M. A. (2002). A taxonomy and treatment of uncertainty for ecology and conservation biology. *Ecological Applications* 12, 618-628.

Table 1. Ecological data with no metadata

VO	5/30/2002	1	S1	1	5	16.25	3.19	0.01
VO	5/30/2002	1	S2	1	5	16.25	3.19	0.01
VO	5/30/2002	1	S3	1	5	16.25	3.19	0.01

Table 2. Ecological data with a limited amount of metadata.

Site	Date	Plot	Sp	Pa	R	Bm	P	N
VO	5/30/2002	1	S1	1	5	16.25	3.19	0.01
VO	5/30/2002	1	S2	1	5	16.25	3.19	0.01
VO	5/30/2002	1	S3	1	5	16.25	3.19	0.01
VO	5/30/2002	1	S4	1	5	16.25	3.19	0.01
VO	5/30/2002	1	S5	1	5	16.25	3.19	0.01
VO	5/30/2002	2	S1	0	3	88.82	11.91	0
VO	5/30/2002	2	S2	1	3	88.82	11.91	0
VO	5/30/2002	2	S3	1	3	88.82	11.91	0
VO	5/30/2002	2	S4	0	3	88.82	11.91	0
VO	5/30/2002	2	S5	1	3	88.82	11.91	0
VO	5/30/2002	3	S1	1	4	22	9999	9999
VO	5/30/2002	3	S2	1	4	22	9999	9999
VO	5/30/2002	3	S3	1	4	22	9999	9999
VO	5/30/2002	3	S4	0	4	22	9999	9999
VO	5/30/2002	3	S5	1	4	22	9999	9999
VO	5/30/2002	4	S1	0	2	12.36	5.14	0.22
VO	5/30/2002	4	S2	1	2	12.36	5.14	0.22
VO	5/30/2002	4	S3	0	2	12.36	5.14	0.22

VO	5/30/2002	4	S4	0	2	12.36	5.14	0.22
VO	5/30/2002	4	S5	1	2	12.36	5.14	0.22
VO	5/30/2002	5	S1	0	3	22.18	7.15	0.35
VO	5/30/2002	5	S2	0	3	22.18	7.15	0.35
VO	5/30/2002	5	S3	1	3	22.18	7.15	0.35
VO	5/30/2002	5	S4	1	3	22.18	7.15	0.35
VO	5/30/2002	5	S5	1	3	22.18	7.15	0.35
VO	5/30/2002	6	S1	0	4	15.5	14.35	0.15
VO	5/30/2002	6	S2	1	4	15.5	14.35	0.15
VO	5/30/2002	6	S3	1	4	15.5	14.35	0.15
VO	5/30/2002	6	S4	1	4	15.5	14.35	0.15
VO	5/30/2002	6	S5	1	4	15.5	14.35	0.15
VO	5/30/2002	7	S1	1	4	61.49	12.65	0.45
VO	5/30/2002	7	S2	0	4	61.49	12.65	0.45
VO	5/30/2002	7	S3	1	4	61.49	12.65	0.45
VO	5/30/2002	7	S4	1	4	61.49	12.65	0.45
VO	5/30/2002	7	S5	1	4	61.49	12.65	0.45
VO	5/30/2002	8	S1	0	3	10.04	4.42	0.76
VO	5/30/2002	8	S2	1	3	10.04	4.42	0.76
VO	5/30/2002	8	S3	1	3	10.04	4.42	0.76
VO	5/30/2002	8	S4	0	3	10.04	4.42	0.76
VO	5/30/2002	8	S5	1	3	10.04	4.42	0.76
VO	5/30/2002	9	S1	1	5	75.58	3.55	0.76
VO	5/30/2002	9	S2	1	5	75.58	3.55	0.76
VO	5/30/2002	9	S3	1	5	75.58	3.55	0.76
VO	5/30/2002	9	S4	1	5	75.58	3.55	0.76
VO	5/30/2002	9	S5	1	5	75.58	3.55	0.76
VO	5/30/2002	10	S1	0	1	52.48	18.53	0.91
VO	5/30/2002	10	S2	1	1	52.48	18.53	0.91
VO	5/30/2002	10	S3	0	1	52.48	18.53	0.91
VO	5/30/2002	10	S4	0	1	52.48	18.53	0.91
VO	5/30/2002	10	S5	0	1	52.48	18.53	0.91

Table 3. Relatively Comprehensive, but Unstructured Metadata.

Written Metadata

This experiment was designed to collect productivity, diversity and soil data for Northern California grasslands. The results were published in a paper titled “Soil Nutrients and the Relationship between Diversity and Productivity” (Doe and Smith 2003). Data were collected at two sites, the Coastal Hills Reserve and the Valley Oak Reserve, within the coastal mountains of Northern California. The area is primarily oak (*Quercus spp.*) savannah and grasslands on limestone soil. In spring of 2002, 10-1 m² plots were randomly distributed throughout a 100 km² area of each location. All plots were placed on grasslands. In each plot, presence and absence of all plant species was recorded. All plants were then clipped at root level, dried and weighed to obtain above-ground peak standing biomass. As most of the production is from annual plants, peak standing biomass can be used as an approximate measure of annual productivity. Approximately 0.5 g of soil was collected from the mid point of each plot. This soil was taken back to the laboratory and analyzed for total nitrogen and phosphorous content.

Five species were observed in all plots. Nonnative plants observed included: *Avena fatua* and *Bromus hordeaceus*. Native plants included *Eschscholzia californica*, *Nassella pulchra* and *Calochortus luteus*.

Codes used in data tables are given below:

Site: Site at which data were collected. CH=Coastal Hills Reserve, VO=Valley Oaks Reserve, Date: Date data were collected mm/dd/yy format, Plot: Randomly assigned number of plot, Sp1-Sp5: Presence absence of each of five species. For each species, a value of 1 indicates presence and a value of 0 indicates absence.

Species Name	Code
<i>Avena fatua</i>	S1
<i>Bromus hordeaceus</i>	S2
<i>Calochortus luteus</i>	S3
<i>Eschscholzia californica</i>	S4
<i>Nassella pulchra</i>	S5

R: Species richness, Bm: Biomass, measured in grams, P: Phosphorous in soil, recorded in ppm (parts per million), N: Nitrogen in soil, recorded as a percentage

Data were collected by PI Jane Doe with assistance from graduate student John Smith in conjunction with the staff of the Coastal Hills Reserve and the Valley Oak Reserve. Collection of the data was funded by NSF grant #12345. Data may be used freely. Please acknowledge persons, grants and reserves in any resulting publications.

Contact information:

Jane Doe
 Department of Biology
 Northern California University
 University Town, CA 95666
 (321) 654-0987 doe@uncal.edu

Table 4. Standardized and structured metadata created with EML.

Metadata for: Diversity, Productivity and Soil Data for North American Grasslands

General Dataset Metadata:	
Abstract:	This research program was designed to collect productivity, diversity and soil data in Northern California grasslands. Data collected include species richness, presence absence of plant species, peak standing biomass and nitrogen and phosphorus soil content. The relationship between diversity and productivity can take many different shapes. Soil nutrients can affect species composition, diversity and productivity. This research program will attempt to investigate soil nutrients to as a possible factor in determining the shape of the diversity productivity curve. Funding was used to collect richness, plant presence / absence data, biomass and soil nutrient data.
Keywords:	Richness, productivity, grasslands, biomass, northern California, soil nutrients
Usage Rights:	This dataset is publicly available through the Knowledge Network for Biocompletxity. www.knb.ecoinformatics.org Please acknowledge the Knowledge Network for Biocomplexity, NSF Grant 12345 and Dr. Jane Doe in any publications that use this data.
Funding:	NSF Grant 12345

Individual: <i>Owner</i>	Dr. Jane Doe
Position	Associate Professor
Address:	Department of Biology, Northern California University, University Town, CA 95666 USA
Phone:	321-654-0987 (voice)
Email Address:	doe@uncal.edu
Web Address:	http://www.uncal.edu/doe
Individual: <i>Primary Contact</i>	John Smith
Position	Data Manager
Address:	Department of Biology, Northern California University, University Town, CA 95666 USA
Phone:	321-654-0987 (voice)
Email Address:	smith@uncal.edu
Web Address:	http://www.uncal.edu/smith
Organization:	Department of Biology, Northern California University

Article Citation:	
Author:	J. Doe
Author:	J. Smith
Date:	2002
Title:	Diversity, productivity and soil nutrients at a Northern California grassland
Journal:	Plantae
Volume:	23
Issue:	2
Page Range:	1-10

Geographic Metadata:		
Geographic Description:	Data were collected in the coastal mountains of Northern California, in the Valley Oak Reserve. The Valley Oak Reserve is adjacent to and managed by Northern California University (NCU). NCU is located in University Town in Sonoma County, approximately 150 km northeast of San Francisco. The Valley Oak Reserve is located on the east facing slope of the California coastal mountains. The area is primarily oak (Quercus spp) savannah and grasslands. The reserve is 100 km² in area.	
Bounding Coordinates:	Longitude:	-120°15'00" degrees
		-120°30'00" degrees
	Latitude:	39°00'00" degrees
		38°45'00" degrees

Temporal Metadata:	
Temporal Description	The observations in the data were made in the late spring, at approximately peak biomass. All observations were made during daylight hours. Observations were made during the following range of dates.
Begin:	5/30/2002
End:	06/05/2003

Taxonomic Metadata:

Taxonomic Authority Citation:	Type:	Book
	Author:	W. L. Jepson
	Editor:	J. C. Hickman
	Date:	1993
	Title:	The Jepson Manual: Higher Plants of California
	Publisher:	University of California Press
	Publication Place:	Berkeley
General Info	All herbaceous plant species were recorded.	
Classification:	Taxonomic Level:	Family
	Taxonomic Name:	Poaceae
	Taxonomic Level:	Genus Species
	Taxonomic Name:	Avena fatua
Classification:	Taxonomic Level:	Family
	Taxonomic Name:	Poaceae
	Taxonomic Level:	Genus Species
	Taxonomic Name:	Bromus hordeaceus
Classification:	Taxonomic Level:	Family
	Taxonomic Name:	Liliaceae
	Taxonomic Level:	Genus Species
	Taxonomic Name:	Calochortus luteus
Classification:	Taxonomic Level:	Family
	Taxonomic Name:	Papaveraceae
	Taxonomic Level:	Genus Species
	Taxonomic Name:	Eschscholzia californica
	Common Name:	California poppy
Classification:	Taxonomic Level:	Family
	Taxonomic Name:	Poaceae
	Taxonomic Level:	Genus Species
	Taxonomic Name:	Nassella pulchra
	Common Name:	Bunchgrass

Methods Metadata:

General Sampling Design
Twenty five square 1m ² plots were randomly placed throughout the Valley Oak Reserve. Due to destructive biomass harvest, plots are relocated each year.
Two plastic sample bag (Ziplock) are labeled with the randomly assigned plot number and contents (plant or soil). If more than one bag is needed, all bags are labeled with the contents, plot number and bag number (i.e. the second of four bags of plant clippings from plot 6 will be labeled "Plant, Plot 6, Bag 2/4")
Species Presence- Absence
At each plot, one person, starting at the south-east corner of the plot, identifies each plant according to the Jepson manual. Species names are recorded in the field notebook.
Species Richness
Species names are used to calculate species richness per plot.
Biomass Measurement
All plant material within each 1 m ² plot is clipped at soil level and placed in the sample bag.
Sample bags containing plant matter are brought to the laboratory. If wet, plant matter is dried using paper towels.
Plant material is dried in a drying oven at 80 degrees C for 24 (+/-2) hours.
Plant matter is weighed within 2 hours of drying.
Soil Data Measurements
Approximately 0.5 g of soil, free from plant debris, is collected from the middle of the plot.
Soil is placed in appropriate sample bag.
Soil samples are placed into aluminum sample trays and placed into a drying oven and dried at 80 degrees C for 24 (+/- 2) hours.
Soil is ground using a ball mill until powdery and weighed.
Soil sample is analyzed for Phosphorus and Nitrogen using a SoilPro v. 10 machine. All procedures for this machine are followed.
Quality Control
All sampling was done by Jane Doe and John Smith.

Data Table Metadata:

File Name:	EML_simpleex_table.txt
Case Sensitive?:	no

Number Of Records:	50
Orientation:	Th

The data are arranged with major variables in columns.

Data Table Structure & Attribute Description

Attribute Name	Label	Definition	Unit	Type	Missing	Precision	Att.	Description
Site	Site	Site at which data were collected		integer				enumerated: code: defin: VO Vally Oak Reserve
Date	Date	Date data were collected	mm/dd/yy	date				
Plot	Plot	Randomly assigned 1m ² plot numbers		integer				numeric: min: max: 1.0 10.0
Sp	Species	Species codes		text				enumerated: code:defin: S1 Avena fatua S2 Bromus Hordeasceus S3 Calochortus luteus S4 Eschscholzia californica S5 Nassella pulchra
Pa	Presence/Absence	Presence/Absence of species in plot	species	integer				enumerated: code: defin: 1 Present 0 Absent
R	Richness	Total number of species in plot	species	integer				numeric: min: max: 1.0 5.0
Bm	Biomass	Peak standing biomass in plot	g/ m ²	float				numeric: min: max: 10.04 88.82
P	Phosphorous	Phosphorous in soil. Measured per plot	ppm (parts per million)	float	9999	±0.05		numeric: min: max: 3.19 18.53
N	Nitrogen	Nitrogen in soil. Measured per plot.	proportion (of total soil mass)	float	9999	±0.01		numeric: min: max: 0.0 0.91