Coweeta LTER Guide to Assigning Key Words

Keywords are an integral part of metadata, the set of instructions or documentation that describe the content, context, quality, structure, and accessibility of a data set (Michener et al. 1997). According to Michener et al., keywords fall under the broader category of Class I data set descriptors, whose aim is to alert secondary users to the existence of data sets that fall within specific temporal, spatial, and thematic domains (1997).

Keyword assignment can be considered a fairly subjective activity. However, with the goal of standardizing keywords used in relation to Coweeta LTER-affiliated data, the Coweeta LTER Information Management Office has elected to recommend some guidelines for keyword selections. Appropriate, standardized, and comprehensive usage of key words will assist users in performing targeted searches for data sets and relating different data sets to each other based on their key word commonalities and differences (Blankman and McGann 2003).

One of the main goals of Ecological Metadata Language, or EML, is to define a common structure that all ecologists can use to document ecological data in order to assist other ecologists in interpreting the data (Blankman and McGann 2003). According to the “Ecological Metadata Language: Practical Application for Scientists” guide, keywords fall under the scientific description “kingdom”, which also includes the abstract, creator name, and other descriptors. The scientific description “kingdom” encompasses metadata needed to allow a researcher to make a targeted query and to describe the scientific and organizational research queries. According to EML, the goal of keywords is to answer the question “What are some of the key concepts that refer to the data?” Keywords are useful in describing the nature of the research being documented, particularly when the data is being designed for use with a search engine. (Blankman and McGann 2003)

Generally, keywords fall into two broader groups based on what they describe. One category of keywords describes what the dataset is about. Examples of such keywords are words or phrases describing the geographic scope (such as “Otto, North Carolina”) or institutional affiliation (such as “Duke University”). The second category of keywords, which the Coweeta LTER is placing a stronger emphasis on, consists of keywords describing what the dataset contains. Examples of keywords describing what the dataset contains include words or phrases related to the variables measured (examples include “phosphorus”, “soil moisture”, and “abundance”).

When selecting keywords, it is recommended to use the LTER Controlled Vocabulary whenever possible (http://vocab.lternet.edu/vocab/vocab/). If the precise phrase, concept, discipline, or other type of keyword is not found in the LTER Controlled Vocabulary, try to adhere closely to the listed vocabulary and phrases. In some cases, a keyword such as “fish population dynamics” can easily be broken down into “fishes” and “population dynamics”, both of which are listed in the LTER Controlled Vocabulary website.

Keywords should define the spatial scale, temporal scale, and thematic scale of the study (Michener et al. 1997). The Coweeta LTER has determined that a set of key words must, at the very minimum, include keywords falling into several categories. Researchers should aim to generate at least one keyword for each category. For many of these categories, it is perfectly acceptable or even recommended to list more than one keyword or phrase in order to adequately describe the data set or research project. First of all, the keyword section must include at least one word or phrase identifying the LTER site by its full name, such as “Coweeta LTER” or “Harvard Forest LTER.” If the research project is affiliated with more than one LTER site, it is best to list both site names. In addition to having a keyword associated with the full name of the LTER site, researchers should separately list the three-letter acronym for the site, such as “CWT” for Coweeta or “HFR” for Harvard Forest LTER.

There are several categories of keywords which pertain to the broader themes found within the research project or data set. The keywords section must include at least one word or phrase identifying the LTER core research area. There are currently five LTER core areas, which are “primary production”, “population studies”, “movement of organic matter”, “movement of inorganic matter”, and “disturbance patterns.” In addition, the keywords section should include a word or phrase describing the overall research theme. Research themes for monitoring studies include “bacterial productivity”, “fungal productivity”, “hydrology”, “terrestrial insect ecology”, “aquatic invertebrate ecology”, “meteorology”, “nutrient chemistry”, “organic matter chemistry”, “phytoplankton productivity”, and “plant ecology”. Research themes for directed studies include “anthropology”, “botany”, “chemistry”, “geology”, “geographic information system analysis”, “geophysics”, “microbiology”, “physics”, and “population ecology.” Once again, whenever applicable, it is perfectly acceptable to list more than one research theme per data set. Another category of keywords focuses on decadal themes, which are recurring themes specific to the Coweeta LTER program. The decadal themes are “forest structure and function”, “water quality and water quantity”, “biodiversity”, “population and community ecology”, “synthesis and scaled integration” (indicative of modeling studies), “land use and regional decision-making”, and “cross-site collaboration.”
One of the keywords should be a word or phrase identifying some meaningful geographic place names (e.g. state, city, county) which relate the location and spatial scale. An example of a keyword is “Otto, North Carolina”, commonly used in reference to research projects based in the Coweeta Hydrologic Laboratory. The keyword section should also include one network acronym, such as “LTER” (for the Long Term Ecological Research Network), “ILTER” (for the International Long-Term Ecological Research Network), or “NEON” (for the National Ecological Observatory Network). One word or phrase should describe the organizational affiliation of the data set. Examples of organizational affiliations are “University of Georgia”, “USDA Forest Service”, and “Duke University.” Once again, data sets affiliated with multiple researchers, universities, and/or funding sources will often have more than one organizational affiliation, and all organizational affiliations should be listed.

Several categories of keywords describe the content of the data set in varying broad to specific levels. Briefly, these categories of keywords serve to answer the question, “Do I want to download these data?” For the following categories of keywords, examples will be drawn from the Coweeta LTER “Habitat suitability and the distribution of species” (Pulliam 2000). One category of keywords should be a word or phrase describing the study at the discipline level, such as “landscape ecology.” Another category of keywords should describe the study at the theoretical level, such as “metapopulation theory.” One word or phrase should describe the data set at the study level, such as “habitat suitability.” One word or phrase should describe the study at the measurement level, such as “soil moisture.” Finally, a word or phrase should describe the sampling interval in the study. For the sake of consistency, researchers should classify the sampling interval in several standard categories (“continuous”, “hourly”, “weekly”, “biweekly”, “monthly”, “bimonthly”, “quarterly”, “annually”, “biannually”, or “uneven sampling interval”).

According to the “EML Best Practices for LTER Sites” guide (2011), one category of keywords should describe the funding source (i.e. “LTER funding”, “co-funded with other sources”, “non-LTER funding”, etc.). Thus, an example of an adequate keyword list for a data set is “Coweeta LTER, primary production, population and community ecology, synthesis and scaled integration, CWT, Otto, North Carolina, LTER, University of Georgia, landscape ecology, metapopulation theory, habitat suitability, soil moisture, biannually, LTER funding.”

**Summary of Key Word Categories:**

1. One word/ phrase identifying the LTER site (ex. “Coweeta LTER”)
2. One word/phrase identifying the LTER core area (ex. “primary production”)
3. One word/ phrase identifying the research theme (ex. “population and community ecology”)
4. One word/ phrase identifying the decadal theme (ex. “synthesis and scaled integration”)
5. The three-letter acronym for the site (ex. “CWT”)
6. One word/phrase identifying some meaningful geographic place names (ex. “Otto, North Carolina”)
7. One network acronym (ex. “LTER”)
8. One word/ phrase about organizational affiliation (ex. “University of Georgia”)
9. One word/phrase at the discipline level (ex. “landscape ecology”)
10. One word/phrase at the theoretical level (ex. “metapopulation theory”)
11. One word/phrase at the study level (ex. “habitat suitability”)
12. One word/phrase at the measurement level (ex. “soil moisture”)
13. One word/phrase concerning the sampling interval (“continuous”, “hourly”, “daily”, “weekly”, “biweekly”, “monthly”, “bimonthly”, “quarterly”, “annually”, “biannually”, or “uneven sampling interval”)
14. One word/ phrase describing the funding source (ex. “LTER funding”)

**Citations:**


Article ID: 137

Last updated: 05 Feb, 2016

Revision: 1

Information Management -> IM Processes -> Coweeta Data Processing via the Matlab Data Toolbox -> Coweeta LTER Guide to Assigning Key Words

http://coweeta.uga.edu/cwt_kb/entry/137/