# Best Practices for Adding Science Keywords to LTER Metadata

Draft: June 29, 2011

The goal of adding keywords to a metadata document is to assure that researchers who want to use your data will be able to locate it reliably and efficiently. Adding keywords from a controlled vocabulary means that your data can be linked to other similar datasets, greatly adding to its scientific value. Here are some best practices for keywording your metadata documents:

* **Provide keywords from as many of the different taxonomys (top-level groupings) as possible.** Ideally you should have at least one keyword from each of the different taxonomys in the controlled vocabulary. However, there may be some taxonomys that are simply not applicable to your dataset and these may be skipped. Using a broad selection of keywords is a good idea because, if a user is browsing down through a taxonomy to locate data, and you don’t have a keyword from that taxonomy associated with your dataset, they will never find your data.
* **Use the most specific possible keywords within the taxonomys.** When searching or browsing, higher-level the “parents” or higher-level terms for each keyword are implied, so choosing the most specific “child” term combines the highest level of discoverability with the maximum level of discrimination. For example, rather than choosing “transects” choose the more-specific child-term “vegetation transects. ”
* **Be willing to make reasonable compromises**. By its nature keywording requires compromise. Datasets vary widely, but if that uniqueness is fully expressed in the keywords, then searching becomes virtually impossible. Therefore you may need to make reasonable compromises in order to be able to use keywords from the controlled vocabulary. For example, you may have conducted a study on the population ecology of rodents, but when you go to the controlled vocabulary, “rodents” isn’t listed, but “small mammals” is. Rather than simply adding “rodents” as an uncontrolled keyword, use the next best term (“small mammals”) instead. If you want, you can also add “rodents” as an uncontrolled keyword, but also add the “closest” keyword from the list as well, because uncontrolled keywords don’t show up in browse-type searches.
* **If you really need to use a keyword not already part of the controlled vocabulary, put it in the proper form**. The international standard NISO Z39.19 (Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies) has recommendations the form of keywords. For example, nouns are preferred and they should be plural if they are something that is counted, but singular if they are something to which the question “how much” might be reasonably applied. See section 6 of NISO Z39.19 for details.
* **If you think a keyword really should be part of the LTER-wide controlled vocabulary, propose that it be added**. A proposal typically will consist of: the keyword, its definition, the rationale for adding the keyword, the packageID of at least one dataset that already uses the keyword, a suggestion on where it should be placed in the existing taxonomys, if there are any related terms to which it should be linked and and non-preferred synonyms.