

## Virtual Water Cooler Session - Controlled Vocabulary Presentation - April 6&7 2009



Published on *LTER Information Management* (<http://im.lternet.edu>)

Home > IM Working Groups > Virtual Water Cooler Session - Controlled  
Vocabulary Presentation - April 6&7 2009

---

### Virtual Water Cooler Session - Controlled Vocabulary Presentation - April 6&7 2009

Mon, 04/06/2009 - 7:56am — jporter

The attached Powerpoint is for use in the session. Also attached is the draft word list in Excel.

In Day 1 of the meeting, Margaret O'Brien mentioned a site working on the semantic/ontology approach. The site is: <https://semtools.ecoinformatics.org/>  
[1]

Below is the text of the email circulated prior to the meeting:

IN BRIEF: In preparation for next week's VTC, attached is a prototype list of "LTER Data Keywords". Please try keywording three datasets from your site using the list.

Discussion topics for the VTC:

- 1) Goals for the keywording effort
- 2) Evaluation of the list
- 3) Developing a process for adding or subtracting words

- 4) Developing a browse polytaxonomy or thesaurus or ontology
- 5) Tools to aid in keywording datasets and searching

IN FULL: In preparation for our April 6-7 IM VTC, I've produced a draft "straw man" controlled vocabulary list for LTER Data Keywords (attached). It consists of ~640 keywords. Additional "sheets" in the spreadsheet contain synonyms, taxonomic and place keywords. There are also some terms I have questions about (Productivity vs. Production). I've sorted the list to put the questionable terms at the TOP.

Please test the list by using it to keyword three of your datasets.

A term was included in the candidate list if either of the following was true:

- 1) Two or more sites used the term as an EML Keyword; OR
- 2) the term was used at least once as an EML Keyword and was also included in either the NBII Thesaurus, the GCMD keyword list, a list of words used recently in searching the LTER Metacat OR
- 3) it was in the list of browse terms used in the KNB/LTER Metacat browse lists, OR
- 4) it is a preferred form of a word that otherwise qualified.

The first criteria simplifies the task by eliminating 5,756 words that were only used at a single site (see below on more about single site words). For adding "Preferred Terms", I've tried to follow the NISO Z39.19-2005 standard:

Chapter 6 focuses specifically on the forms words should take. They recommend nouns, with plurals where you use them for counting, but singular where they refer to amounts rather than counts. However, I'd recommend taking a look at that (large, but very useful) document, to make sure I've gotten it right. One note, if NBII or GCMD already was using the term, I did NOT create a (different) preferred form, as we want to easily match up with those systems. (Oops - I noted one exception - FISHES vs FISH)

Note that we can still use the non-preferred terms by creating a "synonym ring" that can be used to aid searching, even if we only recommend the preferred terms for adding to EML documents.

An issue that we have discussed, but not resolved, is the issue of EML Keyword terms that are used only at a specific LTER site, but may be important to that site (e.g, Kelp Forest). The process above largely eliminated such terms from the list. Issues for you to consider are whether we should add these to the general list (where they may impede keywording at other sites by adding more "chaff" to wade through) or whether the tools we design for using the list that allow both LTER-wide and site-specific terms (which makes the tools harder to design and implement).

You might also want to look at the parts of NISO Z39.19-2005 that deal with taxonomies, and thesauri. My thinking is that our next step would be to create a polytaxonomy, perhaps starting with the high-level concepts used in the current Metacat browse interface, but including a few more levels. . . . But Barbara and Wade have been working on this, so it will be good to hear from them.

Many thanks to Duane, Giri and all the rest of you who have contributed to providing lists, cross checking them etc.! Thanks also to those of you who will contribute by helping vet the list!

I look forward to seeing you at the VTC!

-John Porter

Attachment	Size
LTER_Keywords_V0.1.xls [2]	217 KB
ControlledVocabulary-IMVTC200904.ppt [3]	949 KB

- ControlledVocabulary [4]

- Copyright © 2012 Long Term Ecological Research Network, Albuquerque, NM - This material is based upon work supported by the National Science Foundation under Cooperative Agreement #DEB-0236154. Any opinions, findings, conclusions, or recommendations expressed in the material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Please contact us with questions, comments, or for technical assistance regarding this web site.

---

**Source URL:** <http://im.lternet.edu/node/432>

**Links:**

- [1] <https://semtools.ecoinformatics.org/>  
[2] [http://im.lternet.edu/sites/im.lternet.edu/files/LTER\\_Keywords\\_V0.1.xls](http://im.lternet.edu/sites/im.lternet.edu/files/LTER_Keywords_V0.1.xls)  
[3] [http://im.lternet.edu/sites/im.lternet.edu/files/ControlledVocabulary-IMVTC200904\\_0.ppt](http://im.lternet.edu/sites/im.lternet.edu/files/ControlledVocabulary-IMVTC200904_0.ppt)  
[4] <http://im.lternet.edu/taxonomy/term/45>