

Controlled Vocabulary Working Group – Progress Report

John Porter (VCR), Margaret O'Brien (SBC) February, 2013

The Controlled-Vocabulary Working Group is in the process of enhancing the LTER Controlled Vocabulary to support data synthesis by incorporating new terms and definitions, analyzing how existing keywords are being used and completing work on a Best Practices document. The group scheduled a workshop for mid-May and began work on pre-workshop tasks (below). We have also engaged in communications with the EnvThes European thesaurus regarding automating acquisition of shared definitions..

Activities October 2012 to February 2013:

❖ These activities use the following definitions:

- term = a word or phrase in the LTER controlled vocabulary
- EML-keyword = content in an EML metadata document's keyword element.

Using these definitions, a 'term' is not a 'keyword' until someone uses it to describe a data package by putting it into an EML <keyword> element.

- A. Pre-workshop goals with sub-WGs formed to address each, with results to be vetted during workshop:
 - a. all current terms have definitions and relationships (mainly synonyms) - **IN PROGRESS**
 - i. work to be vetted during workshop
 - ii. identify domain experts to be available via email for questions.
 - b. additional terms have been suggested by 2 mechanisms,
 - i. by sites, with definitions, relations and docids - **REQUESTED BY FEB 28**
 - ii. with some new queries to Metacat (orig list dates to 2006). Includes accumulated statistics on usage in current EML documents- **IN PROGRESS**
 - c. outline external resources for geographic and taxonomic terms - - **IN PROGRESS**
- B. Possible activities at IMC annual meeting could include either (or both) soliciting feedback from the IMC on several aspects of vocabulary use, or the working group disseminating more information about using the vocabulary in site IM systems.
 - a. Feedback the working group needs from the IMC:
 - i. Input for guidelines on how vocabulary terms should be used for returning data There are implications for catalog interface design. Examples:
 1. What parts of EML documents should be searched?
 2. Should the network plan for keyword-only searches? More precise matches would be possible, but higher demands would be placed on all packages. Adoption of such a plan depends on catalog features and implementation, and requires a timeline.
 3. What other uses of EML-keywords could the Network take advantage of in the future? For example, an EML-keyword could include the name of a synthesis project.
 - ii. Testing of searches on current terms via the tematres interface. E.g, searching RTs increases the size of the result set (# hits); is this a desirable outcome or not?
 - b. Additional information the IMC might need from the working group
 - i. An introduction or tutorial on using web services to retrieve terms from tematres (e.g., for insertion as EML keywords)
 - ii. Outline or review of best practices drafts (2 aspects of BP)
 1. How to suggest a new term in the vocabulary
 2. What to include in your EML-keywords