

## Virtual Update Notes April 6 & 7, 2009 - Controlled Vocabulary WG Update



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

Home > Virtual Update Notes April 6 & 7, 2009 - Controlled Vocabulary WG Update

---

### Virtual Update Notes April 6 & 7, 2009 - Controlled Vocabulary WG Update

Wed, 03/25/2009 - 1:09pm — sremillard

Controlled Vocabulary Working Group Update

See Powerpoint presentation: <http://intranet.lternet.edu/im/node/432> [1]

Participants:

Monday (4/6): John Porter (VCR; moderator), Don Henshaw (AND), Duane Costa (LNO), Hap Garritt (PIE), Barbara Benson (NTL), Kristin Vanderbilt (SEV), Gastil Buhl (MCR), Margaret O'Brien (SBC), Mark Servilla (LNO), Suzanne Remillard (AND)

Tuesday(5/7): John Porter (VCR; moderator), Emery Boose (HFR), Corinna Gries (CAP), Duane Costa (LNO), Nicole Kaplan (SGS), Hap Garritt (PIE), Wade Sheldon (GCE), Theresa Valentine (AND), Mark Servilla (LNO), Jim Laundre (ARC), James Brunt (LNO)

Monday Notes:

Notes by Kristin Vanderbilt

Would be nice to have people review the list.  
See issues and discussion topics.

Tuesday Notes:  
by Emery Boose

See power point on IM website for more details.

The process for controlling the list must be determined. Who will manage it?  
How will additions be decided? What related activities should take place at ASM?

The purpose of a controlled vocabulary is not to provide the best description of a dataset, but to facilitate searching by providing the best bins.

Problem: 90% of keywords are used at 4 or fewer LTER sites.

Goal is to aid in the discovery of data by researchers. Develop browseable structures (taxonomies, thesauri, ontologies).

Project has been underway for a couple of years. It began with studies of existing systems. LNO Metacat now provides a list of keywords currently in use. Develop tools to aid in keywording and searching.

Draft list of about 650 words. Species names and locations were removed. These can be handled in separate lists (e.g. species list, gazetteer). Criteria for word selection: (1) used at 2 or more sites, or (2) used at 1 site plus NBII, GCMD, KNB Metacat browse list or recent searches.

ISO standard for creating controlled vocabularies identifies preferred forms. E.g. amounts are singular, countable quantities are plural. Synonym ring to link existing forms to preferred form.

Formalize draft list into a database. Approaching a thesaurus.

How to handle site-specific forms (e.g. kelp forest)?

Polytaxonomy means different taxonomies for different purposes.

What should be the chain of authority for managing the list?

Other issues: site-specific words, human dimensions (largely absent), locations, homographs (words that mean different things in different contexts).

Next steps: give sites opportunity to propose additions and substitutions, allow sites to vote on changes.

At ASM, present different approaches, new tools, how to find and use data.

Need to discuss words not in the draft list. E.g. general taxonomic terms for browsing.

Taxonomies: Process to date has been bottom up. Also need a top down approach. See KNB website for browse categories. Propose a working group to consider polytaxonomies? Is expert help available for taxonomies?

Work flow considerations. When should the list be implemented at sites? How much time will be required to manage the list when finished?

Site keywords should be flagged so they don't show up at every site. Create filters to show / not show local keywords.

Scientific names and classifications change over time. USDA plant database could serve as a standard. Are there other comparable standards?

Non-inclusion of jargon is best practice. When to commit and begin using the list? Develop target dates at ASM. Avoid "better is the enemy of the good."

Wait until polytaxonomy is complete? Or just ensure that words in the list will appear somewhere in the polytaxonomy (placeholder approach). We probably need to move forward before the polytaxonomy is complete. Later changes to navigation aids but not to datasets. Avoid too many mutual dependencies.

Polytaxonomy information might be maintained separate from datasets.

Andrews LTER local preferred keyword list is also about 650 words.

A good set of synonyms for preferred words would enable search and replace in documents.

GIS working group might help with place names and gazetteers.

The value of having a list far exceeds the value of working on a list. Much progress to date.

Send list to sites for comment and vote.

Chain of command? Who needs to approve decisions (IMC, EB, etc)? This is an issue for the governance working group, which could make recommendations.

The EB will want to make a decision on the final proposal: final keyword list, how to maintain, how much work for each site to implement, programming efforts for database. EB is waiting for IMC action because this project has been mentioned many times in the past. Include the future: how to handle new words, etc. Need to expedite the site vote (e.g. by a deadline).

A manuscript for Ecological Informatics could describe the process. Might help with the EB.

Get consensus from IMC and NISAC regarding best practice. Might be informational only (no need to review). But inclusion in site review criteria would need EB approval.

NISAC can help by developing strategies for infrastructure to implement and maintain at network level.

Document time and effort already invested. We could have used more support in past. Possible additional resources in future.

John P. will circulate a list / description of related tools. Use case scenarios to guide software development. Get PIs to look at tools.

How does draft list compare to journal keywords? Not consulted to date.

See ISO document. Good on general principles but lacks detail on implementation.

Long-term maintenance. Need to add keywords (not delete). Committee oversight, perhaps with rotating membership. Need some rationale for decisions.

Most keyword lists are maintained by a small group of representatives.

If adding a keyword means that all datasets must be evaluated for that keyword, then the process is expensive. Alternatively, a new keyword could be used only for new datasets.

How to teach searching to others? Some people will always find data, others never will.

Try not to be too insular. Be aware of researcher issues early in the game. Get researcher feedback as much as possible. Include a grad student who tried to find data. Invite scientists who have expressed concerns. Assemble a panel. Include NSF concerns about finding data.

Scientists should be more concerned that their data won't be stolen on the web!

Good story to tell: problems, community efforts, solutions.

Here are some follow-up tasks (as provided by John Porter via e-mail):

- Vetting the list of LTER Data Keywords. I've created a Google Docs version of the list and sent it to all the IMs (about 1/4 of all of you) for which I had preferred Google Docs email addresses. Others should send me the email address you want to use for accessing the spreadsheet in Google Docs. The deadline for new additions is April 24, so I can put out the suggested changes to an IM-wide vote by May 1. Emphasis should be on terms that could be broadly applied. We will deal with adding site-specific terms later.
- Prepare a proposal for an ASM session on "Finding and Using Data: Challenges and Opportunities". The basic plan is to have a panel discussion wherein data users, IM's and funding agencies will be able to present their frustrations with finding and using data, and how those problems might be addressed. However, I need volunteers to help me pull together the proposal, ID panelists etc. I'd like to have the proposal in by June 1, 2009. Email me if you would like to work on organizing this working group session.
- Other: Wade and Barbara will be working on a plan for developing a polytaxonomy or other, higher level, structure to support browse-type searches. I will be working on pulling together all the ideas for systems to aid in keywording and searching from our past meetings. We will then need a group to develop these into use-case scenarios so we can have some prototypes to demo at the ASM.

Let me know if you want to volunteer. Theresa will be taking our need for a gazetter of places to the LTER GIS Committee to see how they can help us out.

— Updating the “Vocab” email group. The following list of people are currently on the self-selected VOCAB group. Please let me know if you want to continue to be on the list, or if you wish to be added to the group.

Benson, Barbara (bjbenson@wisc.edu [2]) - NTL

Bohm, Sven (bohms@msu.edu [3]) - KBS ; Brunt, James (jbrunt@LTERnet.edu [4]) - LNO ; Costa, Duane (dcosta@lternet.edu [5]) - LNO ; Gries, Corinna (corinna@asu.edu [6]) - CAP ; Henshaw, Donald (don.henshaw@oregonstate.edu [7]) - AND ; Kaplan, Nicole (Nicole.Kaplan@colostate.edu [8]) - SGS ; Koskela, Rebecca (fnrjk@uaf.edu [9]) - BNZ ; O’Brien, Margaret (mob@icess.ucsb.edu [10]) - SBC ; Porter, John (jhp7e@virginia.edu [11]) - VCR ; Remillard, Suzanne (suzanne.remillard@oregonstate.edu [12]) - AND ; San Gil, Inigo (isangil@lternet.edu [13]) - LNO ; Schildhauer, Mark (schild@nceas.ucsb.edu [14]) - NWK ; Servilla, Mark (servilla@lternet.edu [15]) - LNO ; Vanderbilt, Kristin (vanderbi@sevilleta.unm.edu [16]) - SEV ; Walsh, Jonathan (WalshJ@EcoStudies.org [17]) - BES ; Zolly, Lisa (lisa\_zolly@usgs.gov [18]) - NWK

Thanks for all your help!

-John Porter

Attachment	Size
Notes_VTC_4.6.2009.doc [19]	36 KB

- Virtual Updates [20]

- 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/431>

**Links:**

- [1] <http://intranet.lternet.edu/im/node/432>
- [2] <mailto:bjbenson@wisc.edu>
- [3] <mailto:bohms@msu.edu>
- [4] <mailto:jbrunt@LTERnet.edu>
- [5] <mailto:dcosta@lternet.edu>
- [6] <mailto:corinna@asu.edu>
- [7] <mailto:don.henshaw@oregonstate.edu>

- [8] <mailto:Nicole.Kaplan@colostate.edu>
- [9] <mailto:fnrjk@uaf.edu>
- [10] <mailto:mob@icess.ucsb.edu>
- [11] <mailto:jhp7e@virginia.edu>
- [12] <mailto:suzanne.remillard@oregonstate.edu>
- [13] <mailto:isangil@lternet.edu>
- [14] <mailto:schild@nceas.ucsb.edu>
- [15] <mailto:servilla@lternet.edu>
- [16] <mailto:vanderbi@sevilleta.unm.edu>
- [17] <mailto:WalshJ@EcoStudies.org>
- [18] [mailto:lisa\\_zolly@usgs.gov](mailto:lisa_zolly@usgs.gov)
- [19] [http://im.lternet.edu/sites/im.lternet.edu/files/Notes\\_VTC\\_4.6.2009.doc](http://im.lternet.edu/sites/im.lternet.edu/files/Notes_VTC_4.6.2009.doc)
- [20] <http://im.lternet.edu/taxonomy/term/169>