**Controlled Vocabulary WG**

2010 IMC meeting

John Porter

* Focus on next steps for this project.
* Basic problem: little commonality in current LTER keywords.
* Past accomplishments: list of 650 terms based on widely-used LTER EML keywords, auto-completion feature added to Metacat.
* Goal is to include at least one LTER keyword in each LTER EML file.
* HIVE = helping interdisciplinary vocabulary engineering. Duane transformed list and synonym ring into SKOS format, created HIVE-RS (restful service), created service to identify candidate LTER keywords from an EML file (uses sophisticated algorithms developed by HIVE project).
* Reliability & efficiency are key concepts for librarians: Is document well represented? Are the results useful? As a rule of thumb, it can be 60:40 either way.
* See prototype at: <http://scoria.lternet.edu:8080/lter-hive-prototypes/emlTagger.jsp>
* Needed activities:
* Draft “vocabulary management plan”. Who will oversee updates?
* Develop or adapt database structure and tools for managing the list, synonym ring and hierarchical relations.
* Develop a polytaxonomy for placing terms into a hierarchical framework. This will result in new terms added to the list. Four or five levels would be desirable. Ideally lowest level search would return ~ 10 datasets.
* How can the HIVE keywording tool be enhanced to enable its use by LTER sites?
* Different scopes for site and network vocabularies.
* Current focus on general science vocabulary. Also a need for taxonomic and place vocabularies.
* Keep core list to a reasonable size. List of synonyms may grow as needed.
* What is the best way to incorporate suggested terms back into site systems? Ideally one would select from suggested terms (after other metadata is entered) and these would be added to the EML file.
* How can the vocabulary be incorporated into searching and browsing?
* Possible relationships: preferred term, alternate term, related term, broader / narrower term.
* Also a web service to return HIVE info for a single term.
* Discussion:
* Creating a polytaxonomy is a priority. A series of taxonomies (habitat, etc).
* Kristin created a taxonomy browser for SEV using Drupal.
* Margaret is creating ontologies for SBC. Working from bottom up.
* Adapt other taxonomies?
* Need relational database to hold content.
* How to map polytaxonomies to EML? One could include only lower level terms, since these imply higher levels.
* SKOS is good at broader, narrower, and related. OWL handles more relationship types: is a member of, is an owner of, etc.
* Tools include Protégé (full featured), MindMap (simple, easy).
* Polytaxonomy or hierarchy? If lower terms occur in more than one tree, then it’s a polytaxonomy.
* Next steps (drafts by Jan 1):
* Vocab organization and governance, procedures for modifying.
* Polytaxonomy creation.
* Tools for ingestion & searching.