

Spring, 2006

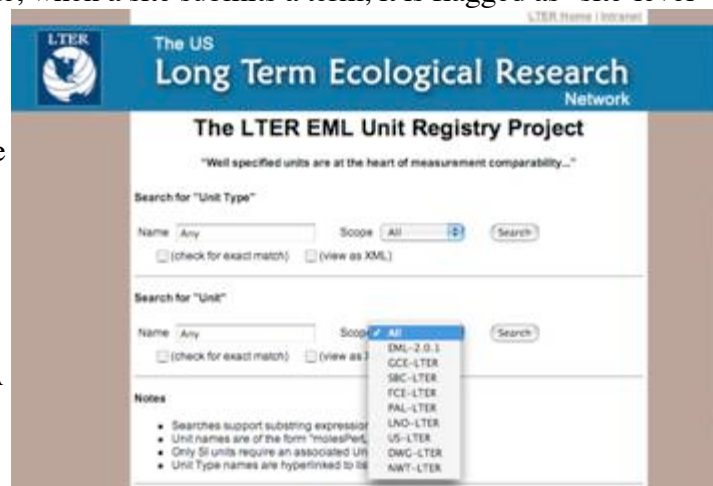


Creating Information Infrastructure through Community Dictionary Processes

-Karen Baker, Lynn Yarmey*, Shaun Haber*, Florence Millerand*, Mark Servilla** *PAL/CCE; **LNO*

A Dictionary Working Group meeting and prototype demonstration were held during the LTER Information Manager 2005 Meeting. The group formed in order to consider and broaden the work of a Unit Registry Design Team that collaborated periodically in the months preceding the LTER IM Meeting. This effort included participants spanning five sites and the network office (KBaker, MServilla, LYarmey, LPowell, MO'Brien, WSheldon, SHaber, FMillerand, ISan Gil). From the demonstration and working group discussion, several information managers explicitly expressed interest in joining the working group (TAckerman, HGarrett, JWalsh). The group's aims may be summarized as: preparing a prototype, developing a dictionary process, and creating a community process. Having already prepared a prototype unit registry, the group's next steps include assessing the prototype and developing federating mechanisms between local and community dictionaries for adding new units and unit types. Additionally, the development of an attribute (or parameter) dictionary is of parallel interest to address the organizational needs at the local level for integrating dictionaries. Capturing and embedding the interdependence

A dictionary is a list of words with information about their definitions and characteristics. Having a unit registry associated with an LTER unit dictionary is a design feature that transforms the dictionary from a passive static list to an interactive community tool - a living dictionary. An important aspect of the unit dictionary is the concept of 'scope'. The scope represents a strategy to define the working acceptance of a specific unit within the community. For example, when a site submits a term, it is flagged as "site-level" (e.g. PAL-LTER). Site-level implies that the unit is accepted only at the site for use in their dictionary process. Units that are under Network review by the dictionary working group are designated as 'DWG', approved at the Network level as 'US-LTER', and at the community level as 'EML'. The initial unit dictionary included all units defined within the EML 2.0.1 specification, and thus have an EML designation. Such a plan is technically straightforward though it would take extensive community and organizational work to enact. A view of the prototype is included in Figure 1.



The prototype is described in the meeting report

[For additional information \(Motivation & Process, Products, Needs, and Design Team\)](#)

(http://gce-fter.marsci.uga.edu/fter_im/2005/index.htm) and available online (<http://fire.lternet.edu/customUnit>) as is the unit registry code (<http://cvs.lternet.edu/cgi-bin/viewcvs.cgi/NIS/projects/customUnit/>).

The initial focus of the LTER Dictionary Working Group was to use a participatory design methodology for considering the dictionary process (see Living Dictionary, Databits Spring 2005). Our first case study yielded the unit registry prototype. One lesson learned from this work is the importance of recognizing the addition of new terms to a dictionary as a community negotiation and learning process. This means that issues like duplicate entries are not bugs, but they are markers of agreements-to-come. At minimum they serve as dialogue prompts, and at best they produce knowledge making moments where local everyday assumptions are considered and differences revealed as options to discuss.

Subsequent to the summer LTER IM Meeting, the unit registry code was checked out from the CVS Repository to a local site (PAL) in order to test its portability and to investigate integration with a multi-project site attribute dictionary. This migration, if viewed as a robustness indicator or prototype benchmark, was successful in practice. Local efforts were able to leverage without any major difficulty from the previous unit registry work. Additionally, a web interface has been developed to enable input as well as administration of unit and attribute dictionaries (see Figure 2). Further working group discussion and resource arrangements will decide the trajectory of this work.



Data dictionaries provide a mechanism to gather and preserve information about field observations as well as to inform both data collectors and data users. Unit and attribute dictionaries represent an organizational strategy and are one element of an information infrastructure. As we gain experience with the scope of our local and community data as well as with information classification, we begin to build an understanding of data typologies, units, and attributes. The process of creating dictionaries establishes a unique setting for dialogue between information system requirements, information managers, and earth scientists.

A dictionary can create a bidirectional forum - one of both elicitation as well as prescription. It serves as a mechanism prompting self-organization; with an explicit organization, it exerts control. What's in the dictionary informs, yet is subject to discussion and update when appropriate processes are in place. So after all, is a dictionary just a controlled vocabulary list? Or is it a moderated forum informed by community needs? Dictionaries are an infrastructure element that may be enhanced by technical structure, organizational flexibility, and community use.

Communication among the three LTER Working Groups - controlled vocabulary, data dictionary and knowledge representation - was facilitated by a visit of Deana Pennington to UCSD/SIO on 16 March this year as part of an ongoing series of Ocean Informatics Exchange events. Ocean Informatics is a conceptual framework for marine science information management efforts including the Palmer LTER, the California Current Ecosystem LTER, the California Cooperative Fisheries Investigations as well as some elements of the Southern California Ocean Observing System (SCCOOS). A hands-on workshop with Ocean Informatics participants created an active forum for data-to-knowledge mental maps; a cyberinfrastructure presentation to the LTER CCE community provided a broad context for local efforts. The visit included a conference call with John Porter as lead of the Controlled Vocabulary Working Group and discussions about LTER related social informatics and articulation work supported by the ongoing Human Social Dynamics Comparative Interoperability Project. The meeting was a continuing recognition of the need for new forums and formats that cross projects, institutions, and traditional task structures. In overload with existing requirements for collecting, managing, and publishing data, there is

typically a lack of organizational support and of resources to dedicate to creating new structures that function as infrastructure building information exchanges.

With data collection and research ongoing, there is a question of where to start in organizing efforts for wider data sharing activities. A dictionary is one place to start. It creates a language held in common at a site. Merging a site dictionary with a community dictionary may be viewed as problematic or, alternatively, as creating an opportunity, a venue for discussion and negotiation. The dictionary becomes a mechanism to make visible the process of information sharing and of community arrangement making. The dictionary initiative emerged in the midst of EML implementation efforts, when information managers were looking for tailored pragmatic solutions to facilitate the transition of local arrangements to support EML. From the Dictionary Working Group development process emerged something more than an additional unit registry to augment the EML unit dictionary. The Working Group served as a coordination tool between the sites themselves and between the sites and the LNO, initiating a community process upon which further information infrastructure efforts may build.

Dictionaries are one of a suite of semantic tools for developing local and federated information infrastructure. In the semantically rich and chaotic realm of observational research, data dictionaries serve as a point of engagement for participants in preparing for data sharing. They provide a place to start for data collectors to engage with community expectations that are semantically demanding and to align with complex information system requirements.