Preservation Metadata: Another Chapter in the Metadata Story

Lynn Yarmey, CCE-LTER & PAL-LTER

The LTER Information Management community has devoted a great deal of time and attention to building relationships and communicating with local scientists in order to meet current site metadata needs. It is also valuable to periodically review our practices and documentation with respect to future user needs. At CCE and Palmer LTER, our iTeam's attention has thus far focused primarily on implementing an efficient and stable system to fully represent the environmental data of each site. From data acquisition practices to data processing procedures, our extended EML metadata schema uses a suite of qualifiers to describe the complex data collected and analyzed by local scientists. This description captured through metadata serves the immediate exchange and access needs of our local community.

In considering the long-term however, needs change as local 'use' shifts to 'reuse' by those outside the local community and data are moved into other long-term repositories.   As data get further away from their source both in time and location, new data representation and maintenance needs arise. From a broad data archival perspective, description of data handling at the local repository level is just as important as the initial description of the original data acquisition.   This type of information is called 'preservation metadata' (as opposed to 'descriptive', 'technical', 'administrative' or 'discovery' metadata) within the library science community; preservation metadata is defined as information needed to preserve digital objects. This metadata describes the data as an object in any form (e.g. a file or a database), and includes such information as the date of upload to a system, who completed the uploading, what format and content changes occurred either at the time of upload or subsequently, who is responsible for the changes and so forth. Preservation metadata maintains the authenticity and integrity of the data over the long-term through documenting changes occurring subsequent to the original data acquisition.

As with any type of metadata, standards provide helpful prompts that stimulate community thinking about data exchange.   In 2002, the library community formed a working group (sponsored by the Online Computer Library Center and the Research Library Group) to discuss a preservation metadata framework. This process resulted in a metadata model and exchange standard called PREMIS, the Preservation Metadata: Implementation Strategies (http://www.oclc.org/research/projects/pmwg/). I recently attended a tutorial on PREMIS offered through the University of California San Diego libraries. While the main focus of PREMIS is on digital documents, there are efforts underway to map scientific data into the PREMIS structure. We can view PREMIS as a container to bring the data and their EML metadata descriptions together with information about a site's data management. Metadata about such topics as the data handling processes (Has the data been changed? Was the original data file name changed?), file format (Was the original file a .dat, .txt, .xls, .mat?   What software created the data file? What software is needed to use or view the data?), related datasets or websites, versioning information and many other elements are included in PREMIS.

When considering a new procedure or practice, the implementation process may be seen as including the work of design, development, deployment as well as enactment (Baker and Millerand, 2007). Enactment, or the initiation of use within a community, is frequently problematic. As was the case with EML implementation, the enactment phase of PREMIS is not currently supported and buy-in is slow, hindered by few available tools, crosswalks and resources.   At CCE and Palmer, we are promoting readiness for data preservation by keeping up with developments in the library and archival realms.   As LTER Information Managers, we hold a key piece of the data stewardship role, namely for supporting local data use while simultaneously bridging to reuse communities as well.   Preservation metadata initiatives provide the LTER community an opportunity to better understand the full context of data stewardship.

References:

Baker, K.S. and Millerand, F., 2007. Articulation work supporting information infrastructure design: coordination, categorization, and assessment in practice. Proceedings of the 40th Annual Hawaii International Conference on System Sciences (HICSS'07), January 03-06, 2007, IEEE Computer Society, Washington, DC, 242a.