



---

◆ **Feature Article**

## **Lessons Learned from EML about the Community Process of Standard Implementation**

*- Florence Millerand (PAL), Karen Baker (PAL/CCE), Barbara Benson (NTL), Matt Jones (NCEAS)*

Having worked with metadata for over a decade, the LTER Information Management community endorsed the Ecological Metadata Language (EML) standard in 2001 as a strategy to support data discovery and integration. EML was developed for the ecology discipline under the Knowledge Network for Biocomplexity project. After a number of years of design, development, and deployment with EML, LTER sites are in the midst of enacting this standard locally. The LTER information managers agreed upon a metadata standard with a machine readable format to create an infrastructure base upon which to build more sophisticated information systems.

To develop a metadata standard for a community is a big endeavor. The design, development and deployment of EML within the ecological community is a far reaching project (Jones et al., 2001 ). The need for community involvement in the development cycle was recognized, and mechanisms such as training workshops were used to involve the community and broaden participation in the development of the standard. As we enter a century of increasing digital infrastructures, the multi-faceted and long-term work with EML provides a unique opportunity to consider the process of developing a community standard.

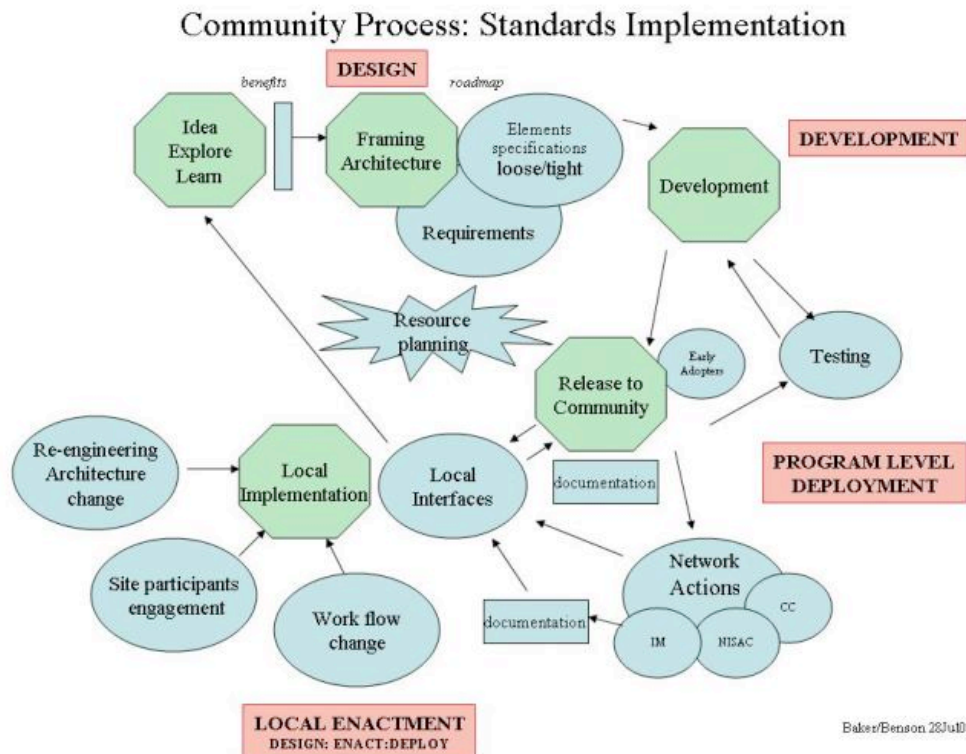
EML, with its wide scope and ecological specificity, has provided a valuable prompt and unique coordination mechanism to the LTER community for preparing datasets to be integration-ready. There are a variety of strategies used today in organizing data - from controlled vocabularies and dictionaries to metadata and ontologies - each addressing different but important aspects of data interoperability. We are not addressing the viability of these efforts here - although we do recognize them as interdependent, not linear or exclusive. The LTER community has made a commitment to metadata preparation a priority. As a result, EML helps the community to focus its

immediate efforts as well as to establish robust elements that can contribute to future or alternative efforts.

A Community Process Working Group (CPWG) held at the LTER Information Manager's annual meeting in Montreal in August 2005 looked back on the standard's implementation as a model of community processes. We expect this reflection and evaluation to help in looking forward to future efforts including the work of dictionary and ontology building. Considering the full history of successes and frustrations with the EML process, what lessons have we learned? The goal of the CPWG was to share information managers' and developers' experiences with EML and distill insights and recommendations on how to improve the process.

The CPWG meeting began with a survey for participants and brief statements by the organizers (Karen Baker, Barbara Benson, Matt Jones, and Florence Millerand) followed by workshop participants sharing their experiences and a final survey. Preparations prior to the workshop included design of the two surveys to collect participant input as well as to prompt participant learning. In addition, a diagram (see [figure 1](#)) to serve as a shared meeting visual was created to capture the full life cycle of standards implementation and to draw attention to the need for language and terms to describe the complex arena of community standards.

**Figure 1. Community Process in Implementing Standards\_**



## Survey results

The first survey - distributed at the beginning of the working group - asked the participants to describe their experiences with EML in terms of degree of success as well as in terms of frustrations and barriers. The second survey - distributed at the end of the working group - asked for the critical factors that the process of developing EML has identified as well as ways of improving the learning process for other projects that may be similar to the EML process.

These surveys ask respondents their opinion about a process, and therefore elicit individual responses. Responses within the surveys represent individual perceptions of respondents who volunteered to participate in the surveys and can't be interpreted as being representative of the entire LTER community. Using qualitative methods, the surveys intend to provide interpretation and understanding of standardization processes through a large variety of responses and respondents (Denzin and Lincoln, 1994 ).

Twenty-four persons responded to the first survey and 14 to the second. Of the 24 participants of the first survey, 18 are site information managers, 1 is a site programmer analyst, 2 are community 'deployers', and 3 are community developers/users of EML. Each of the four questions and responses are summarized below.

### ***Experiences with EML standard: degree of success, frustrations & barriers***

Question 1: Characterize your site experience with EML implementation in terms of degree of success.

A large majority of the participants in the survey reported successful experiences with EML (18 of 24). It is interesting to note that the criteria used to measure this success are different between the community 'deployers' or developers/users of EML and the site information managers.

Overall, the community 'deployers' and developers/users of EML reported a strong recognition of success in terms of "efficient use" of the standard at their local organizations and useful implementations that provided "valuable knowledge about metadata needs that ultimately led to new versions of EML".

With the information managers, two-thirds (13 of 18) reported successful experiences with EML implementation at their site. Note that only 2 information managers reported "making difficult progress", and 3 responded that they were at a "too early stage" to characterize their site experience with EML implementation.

Half the information managers reported successful experiences in terms of a

"full implementation" of EML so that EML metadata can be generated at the site, whereas the other half reported a successful experience that was qualified by some limitation including: a successful but "partial" implementation; some great success even if "some problems still have to be solved"; a success despite the fact that it has required "too much work and time than anticipated".

Also, it is interesting to note that the success of an EML standard is not only measured in terms of the site capacity to generate EML metadata, but also as some broader positive outcomes, such as: "successful tools have been developed locally", "the quality of our metadata is enhanced", "it is more complete, more descriptive", and "the IM community has been brought together more closely".

With respect to success factors, a good socio-technical infrastructure already in place appears to be one of the key advantages for a successful experience. According to the information managers who have succeeded in enacting EML: "we had the advantage of a richly structured database of metadata", "we were able to hire a student to help us", "we had the capacity to develop good local tools".

Finally, the site information managers who participated in the EML development process made exceptionally strong statements regarding success compared to the information managers in general (e.g. "we have about 8-9 degree of success", "EML has had a positive influence", "all our data are on EML"). These views of success are perhaps due to particular infrastructure synergies or to their increased understanding for deployment/enactment resulting from their involvement in the process at an early stage.

Question 2: Describe any frustrations or barriers in implementation of EML at your site.

The two main barriers that information managers have encountered in the EML project are related to timing issues and to the lack of suitable tools. EML limitations and lack of resources in terms of both expertise and funding at the site level are also mentioned as sources of frustrations for some information managers.

These issues point to distinct stages in EML cycle from its design and development, to its deployment and enactment (see [figure 1](#)). We describe the frustrations and barriers that IMs have experienced that relate to each of these stages - as far as it is possible to distinguish these steps given that they are not as isolated as they seem to be.

(1) *Design-development stage: EML limitations* Although the majority of respondents are working with the EML structure and by-and-large did not comment on it, a few information managers reported some intrinsic EML limitations in terms of metadata structure and formats that have contributed to making the standard implementation more complex at their site, (e.g.

"validation issues", "difficulty to encode QA/QC rules for prescriptive purposes"). In addition, EML was claimed to be "poorly suited to working with legacy data", although the survey did not elicit details so specifics are unclear.

(2) *Deployment stage: Timing issue*—The information managers reported the timing issue as the main barrier in their implementation of the standard at their site. This timing problem relates to lags in tool development and EML version releases as well as to considerable gaps between expectations at the deployment level and the reality of EML implementation at the sites at the enactment level.

The "moving metadata standard target" (as one information manager put it) from FLED to FGDC to EML caused extra work, notably "redo work", as well as did the changes in EML itself, requiring the sites to adapt to new EML versions. As the schema was evolving, it was difficult "to come up with one consistent approach" for many information managers. Overall, in working to meet the expectations for EML implementation at the sites that were shared by the entire LTER community, the gap between the amount of time and resources that was needed to achieve this goal versus the effective resources that information managers had at their disposal was a huge source of frustration. "Just the lack of [related xml] tools couldn't match the level of expectations" as one information manager put it. Also, the "changes in support personnel at the network office had a very negative effect" at some sites, because the previous collaborative work did not get incorporated ultimately in sites solutions.

(3) *Enactment stage: Lack of suitable resources*—Although people frequently think of resources as funds and/or personnel, community and open source tools may also be regarded as a resource. The lack of suitable tools was reported as another obstacle in EML implementation. Basically, such tools were either "under-developed", "too site specific", "too buggy" or "over-complex", according to most information managers. And it was only when the site information managers started to implement EML that they discovered the difficult work of adapting or modifying their local metadata to match EML structures.

The "lack of expertise" in metadata standards, the "lack of appropriate documentation", and the "lack of good training materials and examples" to learn from (before the Best Practices document came out) were critical barriers for most of the information managers, and who didn't have enough time (planned) to devote to the learning and testing processes needed.

The "lack of site PI interest" in the process of metadata standard implementation was also reported as a source of frustration for information managers, and several of them regretted the "absence of incentives and easy tools" for the researchers to use and worried about how to get them involved.

Finally, the "funding issue" was reported as an additional barrier to EML

implementation at the sites, not only in terms of inability to hire additional staff to develop tools or work with metadata content but also in terms of information managers difficulty to "justify the amount of effort that was required to the PIs". There was clearly a lack of appropriate resources as well as recognition of scope that made all of the enactment stage problematic.

In contrast, the community developers'/deployers'/users of EML reported domain scientists' "unwillingness" or "resistance" to share their metadata, largely due to the time needed to provide useful metadata, as the main barrier in EML implementation at their own organizations.

### ***Lessons learned: Critical factors and learning process***

Note: This second part of the survey didn't ask for participant identification so we don't distinguish site information managers' and community developers'/deployers'/users' answers.

Question 3: What distinctions, terms or principles best capture critical factors to keep in mind for our next EML-like project?

Drawing upon both the successes and frustrations of EML implementation in the LTER Network, two main critical factors have been identified by the survey participants as important for future projects: community involvement and communication. Training, resources and funding have also been reported as important factors as well as a more structured and staged implementation process.

*Community involvement:* \_ "Being involved early in the process" of such EML-like projects would help in the general understanding of it and facilitate better planning in terms of resource mobilization and allocation. In addition, the need to have "practical experience with the task at hand in order to provide good input" has also been pointed out as a major critical factor.

*Communication:* \_The need to "keep open to communication and participation with vested stakeholders" in such projects would facilitate its design and development as well as its deployment and enactment. One of the challenges becomes establishing effective communication mechanisms between working members of the project.

*Staged implementation process:* \_Exploring and defining a "structured" and "staged" implementation would allow early adopters or testers of the new standard to "test it in real situations and provide feedback" to the rest of the community. This would be facilitated by some compensation or other form of recognition for the testers to ensure timely reporting to the developers.

*Training/Support/Resources:* \_"Good training" early in the project, and "good documentation" - even if it is hard to develop before having experience using the technology - would facilitate community involvement in the project. Also, support and resources - including funding - all the way through as the project

develops would keep the community engaged.

Question 4: How might the learning process for you be improved in future standards implementation projects?

The need for more and better discussions within the community was reported as one of the main ways of improving the learning process in future standards implementation projects. In that respect, better mechanisms to "share information", to facilitate "communication" and "mutual mentoring" have also been mentioned, as well as better "cross-representation of stakeholders" throughout the process. Finally, "advanced planning" would ensure training and time to participate in all aspects of the project and would contribute to significant improvement of the learning process.

### *Discussion*

EML implementation is reported as a "successful experience" for a majority of information managers, which may be a bit surprising given the nature and number of the frustrations and barriers mentioned (e.g. timing issue or lack of suitable resources).

Interestingly, the desire to be involved early in the EML development process (see question 2(2)), as claimed by some respondents, would exacerbate the difficulties mentioned of a moving standard. There appears to be a tradeoff between early involvement and implementation stability. This may be a reminder of important timing issues as well: the EML project is a research and development process but must be a product as well in order to support the community catalog, a product promised in the short-term.

Another interesting result is that individual respondents perceived a need for improvements in community involvement and communication (see question 3) yet many did not take advantage of all the processes that did exist. For example, anyone was welcome to be involved early in the process, as was clearly stated from 1999 onward on the EML web page. In terms of communication, emails went out to the whole IM group with each of the 13 EML beta releases summarizing the changes from the previous, EML's status was presented annually at the IM meeting, and 3 IM training workshops were held at CAP in the final year of the supported metadata research projects.

A similar argument applies to the "staged implementation process", in that feedback was called for at every stage of the 13 structured beta releases, and sites were asked to implement in a staged way (catalog level first, then more complete metadata later). EML was even designed around the idea that different 'implementors' would provide varying levels of detail in a staged manner.

Exploring when these community involvement processes have not been used or fully effective would be illuminating. Is it because of unawareness? Is it because of inappropriateness of the communication mechanisms? Is it

because of other community, organizational, or technical reasons? It is interesting to note that the main barriers encountered in the EML project for the site information managers are not related to the EML standard itself but to the general context of its design/development/deployment/enactment including its related components (e.g. tools or resources). In other words, what has constituted the top source of frustration for information managers in the EML project is not the product but the *support environment*.

Moreover, some of the critical factors mentioned concern mainly organizational and social rather than strictly technical aspects of the EML cycle e.g. "community involvement", "communication", "share information", "staged implementation", "mutual mentoring", and so forth.

Our interest in surveying about experiences with EML is to consider the new and underappreciated elements of design that arise in conducting collaborative science - science on a larger scale than within a laboratory or an organization. Design does create products that require a support environment as part of a community-wide project where this environment is a process of cooperating, coordinating, consensus building, compromising and learning. Standard efforts have traditionally been slow long-term processes to emerge or alternatively have been imposed (with varying degrees of success). In contrast, the work of the Community Process Working Group at the Information Manager Meeting may be seen as a communication mechanism together with those mentioned above. With such a forum we seek to take a progressive approach to the continuing work of creating community standards both rapidly and collaboratively that represents a new demand for research and development efforts.

To conclude, EML is a resource being used by several communities of which the LTER IM community is only one. The results of the surveys presented here do not reflect EML experiences of all communities but rather provide a partial view of the whole picture. We would like to thank the participants for their thoughtful contributions. Though the surveys ask about the enactment phase of EML, an often neglected design activity, the lessons learned are a part of the larger implementation cycle. As the LTER Information Manager community looks forward to future efforts (dictionaries and ontology building, for example), it is important to acknowledge the critical nature of lessons that can be drawn from the EML project in terms of community processes.

## **References**

Jones, M.B., Berkley, C., Bojilova, J., Schildhauer, M. (2001). *Managing Scientific Metadata*. IEEE Internet Computing, 5(5), 59-68.

Denzin, N.K, Lincoln, Y.S. (1994). *Handbook of qualitative research*. Sage Publication