LTER DataBits

01001100 01010100 01000101 01010010

LTER DataBits
Information Management Newsletter of
The Long Term Ecological Research Network

01001100 01010100 01000101 01010010

◆ **Feature Articles**

◆ **Editorials**

◆ **Good Reads**

◆ **Frequently Asked Questions**

DataBits: An electronic newsletter for Information Managers ----- Fall 2005 Issue (http://intranet.lternet.edu/archives/documents/Newsletters/DataBits/05fall/)

Featured in this issue:

After most of the LTER sites have produced their Ecological Metadata Language (EML) packages up to Level 3, the subject still occupies a main line in our agendas. Not only do we want to share our methods to achieve this task with other communities, but we want to study the lessons that have we learned in the process. The article Transforming HTML into EML Text presents methods to "transform HTML input by users into the XML DocBook subset required for incorporation in EML" and the article Lessons Learned from EML about the Community Process of Standard Implementation analyses the process of designing, developing and implementing a standard in our community. As we engage ourselves in this task, we evaluate the adequacy of our metadata to have an impact in future generations of scientists in the editorial Tacit Knowledge Acquisition: Approaching Replicability which give us a starting point for this kind of evaluation. On the other hand the article Baltimore Ecosystem Study Data Document, Management, and Sharing: Preliminary Findings of a Site-Specific Study gives us an insight of the envolement of some of our scientists in complying with the sites' metadata standards. Our commitment to share our sites' information on our web sites targets the search of new methods to facilitate this task as the article Life at the Command Line exposes, and, at the same time, to examine again what lessons have we learned in the process in the article Web Communication Strategies in a Collaborative Environment: Lessons Learned. We report our outreach activities were we have the chance to give and receive knowledge with the rest of the world in the articles Site-based Interactions with TaiwanSite-based Interactions with Taiwan and International Workshop on Information Management, Beijing, China July 20-22, 2005.

DataBits continues as a semi-annual electronic publication of the Long Term Ecological Research Network. It is designed to provide a timely, online resource for research information managers and to incorporate rotating co-editorship. Availability

is through web browsing as well as hardcopy output. LTER mail list IMplus will receive DataBits publication notification. Others may subscribe by sending email to majordomo@lternet.edu with two lines "subscribe databits" and "end" as the message body. To communicate suggestions, articles, and/or interest in co-editing, send email to databits-ed@lternet.edu.

----- Co-editors: Eda C. Meléndez-Colom (LUQ), Brian Riordan (BNZ)

◆ Feature Articles

Life at the Command Line

- John Porter (VCR)

Today, most software applications we interact with use sophisticated graphical interfaces. These interfaces make using software easy, but at the same time, make it difficult to create automated processes that can act when no person is present. In contrast, command line tools, where all the commands to run a program go on a single line, are easy to incorporate into scripts for automating processing of data or creating web products.

Here I'm going to talk about two of my favorite command-line utilities WGET, from GNU (the Free Software Foundation), and IMAGEMAGICK, from Imagemagick.org. Both these programs have two things in common: they run on multiple platforms (Win, Linux, Unix) and have lots and lots of options. This latter commonality is both a blessing and a curse when it comes to using the programs, as finding the right options in a long list is always a challenge.

As I use it, WGET is a command-line web browser that I use to assemble web pages on-the-fly from multiple sources. For example, in my display of dataset metadata, I want to include information from our personnel database, along with information about the data itself. I already have an attractive display format for my personnel records available on a separate web page. Rather than duplicating that database code in my dataset display, instead I embed a WGET command in a script that fetches and includes that page. When run, WGET goes out and fetches the personnel page - inserting that HTML code into the current HTML document. In principle, a CGI-script or PHP program could assemble into one web page source materials from many different web pages, even if they are on different servers. Similarly, when I want to harvest images from our webcams, I use a script containing multiple WGET commands to aim, focus and capture images, issuing the same requests as would come from a "normal" web browser - but I don't have to be there every 10 minutes when the script runs!

Here is a sample WGET command with the options I use for these simple harvests, such as the one described above:
wget -O - -q "http://www.vcrlter.virginia.edu/pda/index.html"
The options used are -O which specifies the name of the file to which output will be sent, in this case "-", meaning standard output and -q (quiet) which suppresses messages from WGET itself. What is returned when I run the program is the raw HTML source code from the specified web page. This code could then be routed to a file, or to another program for further processing using standard command "pipes."

There are many other capabilities of WGET that I do not currently use, but are very useful for other tasks. WGET includes options that allow the capture not only of a single web page, but of entire sets of interlinked web pages. In this capacity, it can be used for mirroring web sites. Options allow absolute web links to be adjusted to relative web links - thus allowing you to put a copy of a web site on your PC for on-the-road use. Capable of dealing with almost all the commonly used protocols (http, ftp etc.), it supports most of the features of a GUI browser, supporting HTML, XHTML, cookies, and proxies.

WGET is available at: http://www.gnu.org/software/wget/wget.html.

The second command-line tool, ImageMagick, is actually a set of 10 individual command-line utilities for the manipulation of images. I'll focus here on the two I use most frequently. CONVERT is a general-purpose image converter. In its most basic form it translates file formats. The command: convert image1.bmp image1.png will copy the image in the image1.bmp bit-mapped file to PNG format in the image1.png file. However, that only scratches the surface of what CONVERT is capable of. Options include resizing the output file (-size), rotating the output (-rotate), manipulating the image (-sharpen, -blur, -emboss). There are even options that use multiple input images to create a single output image. For example the command:

convert -loop 500 -delay 30 image1.jpg image2.gif image_all.gif

would create an animated .GIF file that when displayed in a web browser would first display image1 for 3 seconds, then

image2 for three seconds. The process would then repeat 500 times. I use this program to combine recent images from various web cameras into a single, small, animated .GIF file that we display on our site web page. Because the convert program can be put into a script or batch file, I can use cron (Unix) or Scheduler (Win) to create a new animated .GIF every 10 minutes. If you prefer to have images combined into a mosaic, the -mosaic option would use the input images to create a single combined image showing image1 and image2 side-by-side.

The features listed above are only a small subset of the CONVERT program options. Adding borders, manipulating image contrast and color schemes and even morphing one image into another in an animation are also available.

A second command-line program available in the ImageMagick package is COMPOSITE. This allows one image to be laid on top of another. It is particularly useful for adding grids to imagery used for metric purposes. Here is an example of combining the COMPOSITE and CONVERT commands to do a simple change detection analysis to help identify cryptic (but mobile!) fiddler crabs:

Example of Change Detection using 2 commands and 4 steps.

To do the change analysis, we identify two images of the site (here called the primary and secondary images). We subtract one image from the other by subtracting the "brightness value" (also known as radiometric intensity) of corresponding pixels from the two images and taking the absolute value using the command:

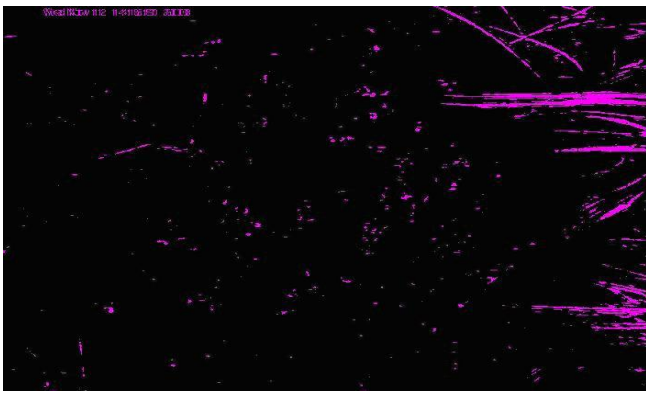composite -compose DIFFERENCE primaryimage.jpg secondaryimage.jpg out1.jpg

This produces an image (out1.jpg) that has bright values where the images were different and dark values where they were nearly the same. Below is a sample "raw" difference image. Note that this image contains lots of dark, but non-black pixels, indicating small changes in brightness.

The next step is to decide how large a difference is important for us to see by setting a "threshold". Pixels that are not bright enough to meet our threshold are set to black using the command:

convert -threshold 25% -colorize 0/100/0 out1.jpg out2.jpg

which applies a 25% threshold and colors things purple to create the new image out2.jpg. Below is a sample image after a threshold of 25% was used:

Next, just as we started off by subtracting the two images, we add together the original (primary) image and the change detection image using the command:

composite -compose PLUS primaryimage.jpg out2.jpg out3.jpg

Finally, we add an image that contains a picture of a grid with the command:

composite -compose PLUS out3.jpg crab_grid.gif finaloutputimage.jpg

Below is the resulting picture for a 25% threshold. The movement of the blades of grass is evident, but more important, the purple spots in the grid represent places where fiddler crabs moved or created new holes between the times the primary and secondary images were taken.
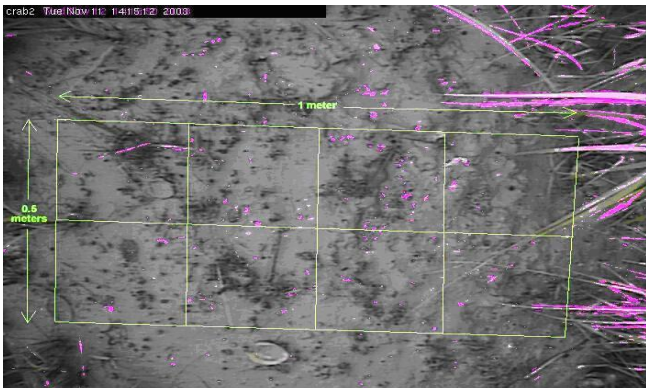
Transforming HTML into EML Text

- John Porter (VCR)

One of the persistent challenges in creating Ecological Metadata Language (EML) documents for the Virginia Coast Reserve LTER has been the need to transform HTML input by users into the XML DocBook subset required for incorporation in EML. DocBook is a Data Type Dictionary (DTD) that specifies the way textual data should be represented in XML (http://www.docbook.org). EML documents, however, use only a subset of the full DocBook elements (see the documentation on the eml-text module for the specific elements at: http://knb.ecoinformatics.org/software/eml/eml-2.0.1/eml-text.html).

My initial approach was to use simple transformations - thus the HTML paragraph marker <P> becomes the DocBook paragraph marker <para>. However, this approach promptly failed because while XML requires strict conformation to the rules of the schema, many (perhaps "most" or "all") hand-coded HTML documents are not rigorously structured. Typically, the lack of rigor in HTML documents causes no problems, because browsers are very tolerant of omissions. For example, all web browsers understand that if <P> is encountered the </P> (closing the previous paragraph) is understood. However, when that HTML document is transformed to go into an XML DocBook, lack of the </para> causes a fatal error during display or validation.

After failure of my initial approach, which used simple translation of HTML codes, I investigated other, more comprehensive options found on the http://wiki.docbook.org/topic/ConvertOtherFormatsToDocBook WIKI web page. Ultimately I ended up with a solution that involved three programs: the TIDY program (http://tidy.sourceforge.net/) that takes ill-formed HTML and cleans it up; the DBDOCLET JAVA application (http://www.michael-a-fuchs.de/projects/dbdoclet/en/index.html) which converts clean HTML code to XML DocBook, (but also adds some undesirable extras that can't be embedded in an EML document); and a locally-written PERL program which strips out those undesirable extras.

Here is a sample piece of HTML code that a user might have submitted as part of their documentation:

```
<H1>My Data</H1>
<P>This dataset addresses the species <B>Spartina alterniflora</B> at two
sites:
<UL>
<LI>North Hog Island
<LI>Phillip's Creek Marsh <2km from Brownsville
```

This file has two major problems: the <P> and <UL> tags are never closed, and the "<" sign has not been converted (and could be mistaken for the start of a tag). Despite these deficiencies, however the HTML displays correctly in a web browser.

This file has two major problems: the <P> and <UL> tags are never closed, and the "<" sign has not been converted (and could be mistaken for the start of a tag). Despite these deficiencies, however the HTML displays correctly in a web browser.

After running the TIDY program on the file we have:

```
<!DOCTYPE html PUBLIC "-//W3C//DTD HTML 3.2//EN">
<html>
<head>
<meta name="generator" content=
"HTML Tidy for Solaris (vers 12 April 2005), see www.w3.org">
<title></title>
</head>
<body>
<h1>My Data</h1>
<p>This dataset addresses the species <b>Spartina alterniflora</
b>
at two sites:</p>
<ul>
<li>North Hog Island</li>
<li>Phillip's Creek Marsh &lt;2km from Brownsville</li>
</ul>
</body>
</html>
```

Now all the tags are closed (and additional tags added. Additionally "<" has been transformed to < Using the JAVA program "dbdoclet" we transform the TIDY'ed file to:

```
<?xml version='1.0' encoding='UTF-8'?>
<!DOCTYPE article PUBLIC "-//OASIS//DTD DocBook XML V4.3//EN" "http://www.oasis-open.org/
docbook/xml/4.3/docbookx.dtd">
<article>
<title>Reference Handbook</title>
<!-- File: /tmp/dbtemp170047 -->
<sect1>
<title>/tmp/dbtemp170047</title>
<para>This paragraph was inserted, to avoid an empty section element.</para>
</sect1>
<sect1 remap="h1">
<title>My Data</title>
<para>This dataset addresses the species <emphasis role="bold">Spartina alterniflora</emphasis>
at two sites:</para>
<itemizedlist>
<listitem>
<para> North Hog Island</para>
</listitem>
<listitem>
<para> Phillip's Creek Marsh &lt;2km from Brownsville</para>
</listitem>
</itemizedlist>
</sect1>
<index/>
</article>
```

Now we have an XML file, but before it can be inserted into an EML document we still need to make some changes. These include stripping away <?xml and <!DOCTYPE lines, removing the extraneous title "Reference Handbook", removing the <sect1> markers (the eml-text schema allows <section> tags but not numbered section tags), removing the "role" attribute from the <emphasis> tag (again, the EML specification does not support a role attribute here) and removing unnecessary <article> and <index/> tags. This is accomplished by a short PERL program that removes lines containing undesired text strings (e.g., <?xml), make changes to tags to remove extra material (e.g., remove the "role=" attribute in an <emphasis> tag), and corrects problems with nesting (e.g., remove <emphasis> tags from inside <title>s). A copy is available at: http://www.evsc.virginia.edu/~jhp7e/html2docbook.pl.

The final product, ready for adding to an eml document is:

```
<section>
<title>My Data</title>
<para>This dataset addresses the species <emphasis>Spartina
alterniflora</emphasis>
at two sites:</para>
<para><itemizedlist>
<listitem>
<para> North Hog Island</para>
</listitem>
<listitem>
<para> Phillip&#8217;s Creek Marsh &lt;2km from Brownsville</para>
</listitem>
</itemizedlist></para>
</section>
```

There are doubtless alternative approaches using TIDY to generate well-formed XHTML documents that could then be acted on with a stylesheet. I tried one of these solutions during my testing, but encountered some difficulties and so went to a more programmatic approach. It seems to be working well, although because of the various programs that need to load, processing text for each EML document takes several seconds, so it may not be suitable for production applications.

Site-based Interactions with Taiwan

- John Porter (VCR) (with help from Barbara Benson-NTL-)

A gathering of giant green snakes? A huge spider towering over us? Not really, just a tangled mass of roots and stems, each enshrouded in a thick layer of greenmoss, that surrounded us as we hiked down to Yuan Yang Lake (YYL) in north central Taiwan (see the Fall 2004 Databits and July 2005 BioScience cover for more photos). Surrounded by ancient cypress (some thousands of years old) and fed by cloud water and typhoons, the thick, verdant forest seems an odd place for a US LTER Information Manger to be! However, YYL is the site of one member of an increasing network of wirelessly-connected buoys studying lake metabolism (http://www.lakemetabolism.org; http://gleon.org), and my visit there was part of a growing collaboration between individual LTER sites and the Taiwan Ecological Research Network (TERN).

My part in this collaboration began when, at the 2003 LTER All Scientists Meeting, we were asked to host a day trip by Taiwanese researchers to discuss information management. They arrived in February 2004 as part of a whirlwind tour of LTER sites (also including CAP, NTL and SEV LTER sites and the Network Office). It was clear once they arrived that this was a group very interested in actually developing information management systems - not just talking about them. As a follow-up to that trip, TERN information managers Sheng-Shan Lu and Meei-ru Jeng each spent three months working at either North Temperate Lakes or Virginia Coast Reserve LTER sites during the fall of 2004 and the spring of 2005 learning the detailed practice of information management as conducted at these sites.

Sheng-Shan Lu spent three months during the fall of 2004 at the North Temperate Lakes (NTL) LTER site. His primary focus was to learn as much as possible about Ecological Metadata Language (EML) and the various methods and tools a site could use to implement EML. He input metadata from one of his projects from Taiwan into the NTL database in order to explore and understand the relational database model NTL uses for its metadata and how the metadata are used to drive dynamic database access on the NTL website. Sheng-Shan participated in the creation and population of the EML taxonomic module for NTL datasets --- a major undertaking. While at NTL, he learned about other ways of managing EML including Morpho, Metacat, and the FCE template. In addition, he took advantage of the opportunity to learn about many other aspects of the NTL information management system.

An important aspect of these extended visits was that it allowed time for real collaborations to develop. For example, one of the tasks Meei-ru Jeng worked on when she visited the VCR/LTER was to learn the PERL programming language. As she was learning, she created a program used for substituting names of investigators for their names/initials in the VCR/LTER publication database, providing her with practice writing PERL and the VCR/LTER with a needed piece of software that is still in use today.

Sheng-Shan Lu, Chau-Chin Lin, Chin-Yin Hwang, Meei-ru Jeng and Fu-Jing Yang (left to right) from the Taiwan Forestry Research Institute guide VCR/LTER IM John Porter (rear) to remote Yuan Yang Lake.

During her time at the VCR/LTER Ms. Jeng also was introduced to the PostNuke Content Management System, PHP, SQL and PERL programming, creation and use of web forms, linking application programs to databases, and developing interactive maps using Mapserver, along with two data collection trips to the Virginia Coast. Prior to arriving, she had already installed an Apache web server and MySQL database onto her laptop.

My trip to Taiwan in July 2005, which included the trips to YYL and the Fushan LTER sites, along with a trip to a forestry research station, had two major objectives. One was to help review their design of a plant disease database used to track occurrences of disease and to manage mitigation efforts. The second was to help develop further collaborations that could strengthen US, Taiwanese and East Asia Pacific ILTER information management.

I was very impressed with the organization and resources that I found in Taiwan. ILTER chair Hen-biau King of the Taiwan Forestry Research Institute (TRFI) has been a strong supporter of information management efforts. Taking the lead for TERN is Chau-Chin Lin. He has assembled a strong team for information management, including specialists in web, JAVA programming, spatial databases and relational databases. They were the first group worldwide to get Metacat software to operate on Windows-based server and have devoted substantial efforts to resolving language and character-set issues in MORPHO and Ecological Metadata Language (EML). We are currently working on several collaborations. Developing "best practices" guides for information management in TERN and the East Asia Pacific ILTER region, creating EML-based tools for researchers that facilitate analysis and templates for sharing spatial data via open-source web mapping programs are all on the table.

Subsequent to the trip, we also received a supplement from NSF INT Program Officer William Chang to support additional extended visits by TERN information managers to US LTER sites for additional training and development of collaborations. There will also be some limited funds to support travel by US LTER information mangers to travel to Taiwan.

An important lesson for me in these interactions is that several ILTER groups are developing world-class expertise in ecological information management. Although the US LTER program has a head start, the TERN group and others are developing at a rapid pace. Just as the US LTER network has benefited from intersite interactions, we are now at a point where we can similarly benefit from international collaborations. Many hands truly do make "light work!"

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. Commmunity Process in Implementing Standards

Community Process: Standards Implementation

Baker/Benson 28Jul05
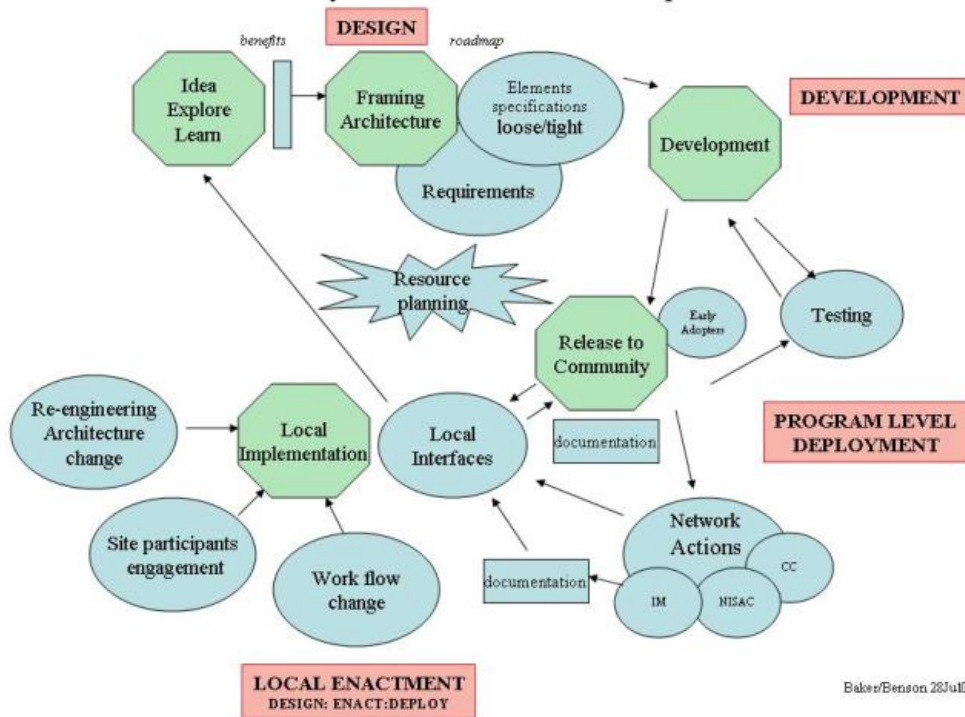
Survey results

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 Publications.

Web Communication Strategies in a Collaborative Environment: Lessons Learned

- Shaun Haber (PAL/CCE) and Karen Baker (PAL/CCE)

A year ago we wrote an article (http://intranet.lternet.edu/archives/documents/Newsletters/DataBits/04spring/#3fa)

about PostNuke (http://www.postnuke.com/), an open-source content management system (CMS), deployed locally as an experiment with community web tools. Our focus included blogs and wikis as part of a collaborative Ocean Informatics Environment (PAL LTER, CCE LTER, and others) initiated in 2004 at Scripps Institute of Oceanography. Though rich with features, we ultimately concluded that PostNuke, and other similar content management systems, are too burdensome for addressing our local and basic collaborative needs. In addition, with rapid development in this type of application arena, the administration interface for PostNuke seemed clunky, creating difficulties in maintaining something as simple as a blog.

During an initial period of CMS prototyping, we gained valuable experience with features useful to our work, including making use of the blog as a community mechanism for communication and content capture. additionally, we created categories for organizing posts, enabled email notifications of new entries, and structured a three-tier administrator/group/public user system. Our group use of the blog is not typical; blogs traditionally are a single-user journal genre. As our PostNuke usage waned over a period of months, we shifted focus to consider some more agile alternatives. Although we investigated other open-source projects including Mambo (http://www.mamboserver.com/) and Xoops (http://www.xoops.org/), both with similar approaches to blogs, wikis and forums, we made the design choice that CMS's were not compatible with our local infrastructure development.

Though CMS's became the first in a series of experiments, our quest to find collaborative web applications continued, particularly in the blog and wiki realms. We worked with Blogger (http://www.blogger.com/), a free and popular blogging service offered by Google. Blogger has some very nice features, most noticeably a stunning user interface. We set up a group blog and ported over previous entries from PostNuke. This exercise served to address migration issues, moreover, the need to capture older material archived in an abandoned application. Despite having a fresh start, Blogger quickly met the same fate as PostNuke, declining in use and, over a period of months, becoming a peripheral tool rather than a core part of our social infrastructure. Perhaps it was the fragmented nature of our project work intervals, or the lack of features that Blogger offers. One significant feature lacking in Blogger is the ability to tag or categorize individual entries, a feature we had used frequently with PostNuke. Further, although Blogger is a free service, it is not an open-source project. Posting a new blog entry requires that you login to Blogger's central server. This removes the overhead of installing and processing the blog engine on a local server, but it also removes the flexibility of extending the blog's functionality and integrating its user-base with other distributed local applications.

Another web application we experimented with was MediaWiki (http://www.mediawiki.org), an open-source project used by popular sites including Wikipedia (http://wikipedia.org). On a side note, wikis differ from blogs in that users create, edit, and link together pages in a wiki, whereas a blog maintains an archived, sequential listing of individual posts. Both tools, however, may be used for collaborative purposes. We initially found the wiki concept cumbersome, particularly the various editing conventions that provide layouts and structure for content. (These conventions are meant to simplify content generation, particularly for users unfamiliar with HTML). In addition, the wiki was generally tricky to use when porting over content and logging page changes. Because of these seemingly minor inconveniences, we were ultimately unable to incorporate it into our work practices.

Despite these shortcomings, why should the applications receive all the blame? The notion of a perfect collaborative infrastructure is more a journey that a destination. A line must be drawn at some point to distinguish an inadequate tool from an inflexible community. With thousands of open-source projects floating around cyberspace, it is difficult to settle on one without researching others. In the end, it is up to the community to let the communication flow, regardless of the media.

We recently installed a new blogging platform, the open-source project WordPress (http://wordpress.org/) . With a simple and intuitive interface, it contains many useful features (e.g. categories), some of them critical to what we've learned to be important to local practices (e.g. email notifications), and others that are of interest for future plans (e.g. rss feeds). Could this be the perfect solution for us? Probably not, but it is fairly ideal for now, providing an outlet for experimenting with communication while communicating. Only time will tell if we will choose to sustain it. If it does fail, perhaps we will entertain the possibility that we are indeed not just a unique community, but also a finicky community incapable of engaging in the greater technological realms of communication.

The only potential obstacle for WordPress seems our willingness to embrace it as an integrating communication medium. We remain hopeful, however, that as we avoid certain tools that create a closed box of constraints detrimental to collaborative work, we will continue to nudge each other to use and develop best practices for collaborative tools like our WordPress blog. That is, of course, until we find something better!

International Workshop on Information Management, Beijing, China July 20-22, 2005

-Barbara Benson (NTL)

An International LTER (ILTER) workshop on information management (IM) was held in Beijing, China on July 20-22, 2005.

The main goal of this workshop was to share progress on information management systems by the different networks involved in the East Asia/Pacific region of ILTER and other regional networks. Barbara Benson (NTL) and John Porter (VCR) were invited to give presentations on the US LTER information management systems and to participate in discussions of future meetings and planning. The meeting was attended by 80 people including representatives and information managers from the networks.

Opening ceremony of the International Workshop on Information Management, Beijing, China. July 2005. (left to right: Dr. Zhao Shidong, Dr. Barbara Benson, Dr. Fu Bojie, Dr. Liu Jiyuan, Dr. Yue-Joe Hsia, Dr. Eun-Shik Kim)

Representatives of the member networks (including the Australia LTER Network; CERN, the Chinese Ecological Research Network; Korea ILTER Network; JERN, the Japanese Ecological Research Network; Philippines LTER; TERN, the Taiwan Ecological Research Network; Thailand ILTER; Vietnam LTER; Mexico LTER Network; South Africa LTER Network and the US LTER Network) gave presentations on the status of long-term ecological research programs in their countries and progress on building information management systems. While some networks have made substantial progress in information management, many are very challenged in terms of resources for long-term ecological research programs. The second day continued the presentations from member networks as well as providing information and demonstrations related to Ecological Metadata Language (EML) and other IM tools. Barbara Benson presented the US LTER information management system at the Network level, and John Porter contributed the site-level perspective. Both the US LTER and TERN representatives were involved in the EML presentations and demonstrations.

The meeting provided a great opportunity to interact with people from the international community engaged in developing information management. During the third day, a smaller group of representatives from the participating networks planned future activities. Two follow-up workshops were proposed. In 2006 TERN will host a meeting in Taiwan that will provide training in EML tools and other technology, and the Korean LTER will host a workshop on using EML databases in collaborative research. A website for the East Asia/Pacific ILTER region was discussed and is under development by CERN.

The local hosts went to great efforts to provide hospitality. Many of the participants were able to visit some of China's wonderful historic sites such as the Forbidden City and the Great Wall. We all enjoyed great meals featuring delicious Chinese cuisine. The meeting was a successful exchange in multiple ways.

Baltimore Ecosystem Study Data Document, Management, and Sharing: Preliminary Findings of a Site-Specific Study

- Shawn E. Dalton (BES), Jonathan Walsh (BES)

September 13, 2005

The LTER network places strong emphasis on data documentation, management and sharing, and relies on information managers at all project sites to standardize and implement protocols for doing so. Our sites are all composed of a variety of researchers with varying levels of expertise in information management, and understanding of its importance to the LTER network and beyond. Thus, compliance with standardized protocols may or may not be in place, and levels of compliance may vary with the nature of research, types of data collected, and the background/expertise of the researchers.

Part of the job of LTER information managers is to know the levels of compliance within the research communities and to identify means by which fully standardized data management protocols can be disseminated and adopted throughout the LTER network and beyond. However, because information managers may or may not be aware of all datasets collected by their participant researchers, it is difficult to know whether or not data management protocols are uniformly and fully implemented.

Thus, we undertook a project to document data management among Baltimore Ecosystem Study (BES) researchers, the ultimate goal of which was to determine whether and what level of data documentation, management, and sharing training is needed among BES researchers, their technicians, students, and other staff. To meet this goal, we designed and disseminated, via a questionnaire, a survey to all BES Co-Principal Investigators. The survey was designed to: 1) address the level of understanding of and compliance with BES standardized data management protocols among BES researchers; 2) assess risk of data loss or decay through lack of adequate management procedures; 3) evaluate the efficacy and extent of data sharing within and between BES researchers and other potential data users.

The survey was administered initially in October of 2004, but received low response rates. It was redistributed on January 1, 2005; and respondents were given until February 15, 2005 to reply. Two reminders to return the completed questionnaire were sent to respondents in the intervening weeks. Of the 42 Co-PIs who received the questionnaire, 15 returned completed forms, a response rate of 35.7%. Preliminary findings indicate high levels of understanding of and compliance with BES data management protocols among respondents. A majority of respondents document and back up their data regularly; some have quality assurance protocols in place; some requested training for these and other data

management standards. Most respondents (73%) share their data via email; some online (47%). Four respondents indicated that they do not share their data, while eight indicated that they share their data with a variety of BES researchers, and one respondent indicated that their data are publicly available through a federal website. Relatively few researchers share their data with state, county, or municipal agencies, or with non-government organizations. Five respondents indicated that they share their data with schools. Fifty-three percent of respondents manage 0-9 separate datasets; 27% management 10-25 datasets; 13% manage 26-50 datasets; and 20% manage 50 or more datasets.

Given the value of these publicly-funded data sets, both in terms of dollar investments and their contribution to understanding long-term trends in US ecosystems, it is critical that they be protected from data decay and loss. We intend to complete a more detailed analysis of the results of this survey, and to build a data management training module for use in training BES the research community, and which may also have the potential to serve the needs of other LTER sites as well.

On the LTER Ecological Metadata Language implementation

- Inigo San Gil (LNO)

The following article summarizes the great work done by the LTER community in adopting a common metadata standard in order to facilitate data access and cross-site synthesis. This article has three small sections; first, a summary of where the LTER sites are in regarding to the EML standard implementation; second, a brief history of the process and lessons learned; and third, where we are heading.
What could be more boring than filling out a long report about a data set? This is one of the questions that may pop into the mind of the busy LTER information manager (IM) when facing the unavoidable task of filling forms to make the site's metadata EML compliant. Despite the daunting prospect of gathering, compiling info and filling out long forms, the LTER IM community has pulled together, and managed to standardize most of the metadata collected by the LTER sites ever since its inception some decades ago.

The following is a brief report of the progress made by the LTER sites in adopting the EML standard as of the time of writing.

The take-home:

- More than 90% of the LTER sites have implemented the EML standard.
- About 85% have made EML metadata available at centralized servers (Metacat). These servers have harvested over 3,000 EML documents from LTER sites. "Harvested" here is loosely defined as: a site has at least a few metadata sets of low content placed in the server, and the site has a good plan to place all legacy data in, including specific plans to enrich EML, if appropriate.
- How about the remaining 15% ? (excluding the two new sites, CCE and MCR)

    HFR : Has implemented EML level 3ish. (yet to complete: Entity table)
    PAL : Focused on site reorganization process.
    CDR : Working on EML implementation and harvest process.
    BNZ : A database-to-EML implementation plan will start in Sept. 2005.
    JRN : On the verge of posting level 5 EML for 70% of datasets

A complete tabular view of the LTER EML implementation status follows

Site	TIER level	# Harvested sets	% of EML implemented	% Harvested	Harvesting since
AND	5	124	100	100	June 2005
ARC	2 ½	1585	100	100	April 2005
BES	1	3	100	1	April 2005
BNZ		0	0	0	N / A
CAP	5	30	100	25	August 2004

Site	TIER	New	Site		Date
CCE		New	Site		N / A
CDR		0	0	0	N / A
CWT	2 ½	190	100	97	May 2005
FCE	5	102	100	100	August 2005
GCE	5	249	100	100	April 2004
HFR	3 ½	0	100	0	N / A
HBR	4	112	100	100	July 2004
JRN		0	0	0	N / A
KBS	3	40	100	100	August 2004
KNZ	5	42	100	96	August 2005
LNO	2	360	100	100	Jan. 2005
LUQ	3 ½	96	100	100	May 2005
MCM	2	20			N / A
MCR		New	Site		
NTL	3 ½	43	80	80	April 2005
NWT	3	139	100	100	June 2005
PAL		0	0	0	N / A
PIE	2 ½	118	100	100	July 2005
SBC	5		100	100	2004
SEV	5	66	80	80	July 2005
SGS	2	9	10	10	June 2005
VCR	5	104	100	100	July 2005

The source of the harvested set numbers us the the LNO metacat as of September 2005. The percentages are ballpark numbers, that is, rounded estimates that help us understand where the sites are best, since the total number of documents depends heavily on whether the site is a data lumper or a splitter. The TIER levels (0-5) reflect the richness of the EML documents. These are based on the EML Best Practices document, whose last revision can be found at the LTER CVS repository. The noted site TIER levels are also to be taken as good faith estimates.

Perhaps a pie chart of the EML richness of the LTER metadata sites gives a good snapshot of the EML implementation state.
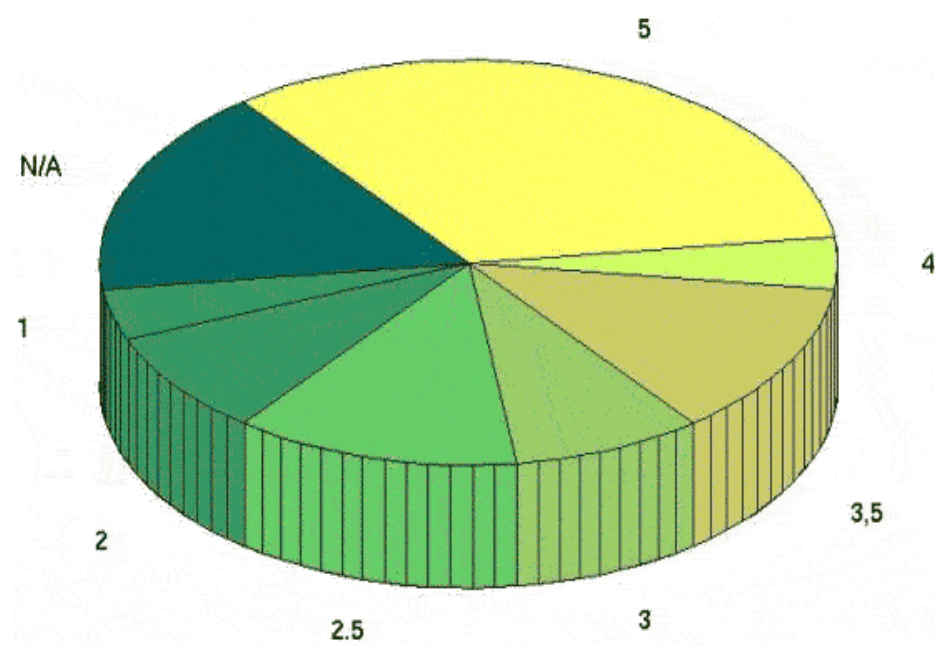
The chart shows the estimated EML TIER levels. The TIER levels are the EML implementation richness where 0 corresponds to no EML implementation and 5 corresponds to about the richest level of metadata. Data estimates are made as of September 2005.

A brief look at the EML implementation timeline

EML was first implemented by the GCE LTER site. By April 2004 the first datasets were stored in the central Metacat servers. A few months later, in August 2004 a couple more sites (CAP and KBS) added themselves to the harvesting process. It was in 2005 when the LNO Metacat server's storage capacity was maxed out that the majority of the LTER sites started harvesting. Recently, the LNO upgraded the Metacat server allowing the harvesting process to continue without disruption. The commitment from the LTER sites, and in particular their information managers, has made the EML implementation a successful project. Having made such great progress in the adoption and implementation of the LTER network metadata standard, we can look forward into the future with optimism.

What is next for EML?

I have some ideas about where to go from here. But it is up to the LTER community to figure out where to go in the long term; whether to put emphasis in cross-site synthesis applications or grid cyberinfrastructure, etc. However, there are a couple of immediate actions in this project that I can outline.
First, we would like to enrich further the EML documents with metadata. Most of the benefits of EML rely on good quality and complete metadata sets. It is not just a matter of including the dreaded attribute info, but to make sure that it is good quality and that we use the EML in the way we think it should be used (again, let me refer to the EML Best Practices document hyperlinked above).
We can leverage from existing applications that use the EML standard (or the NBII standard via crosswalk), whether is Kepler, the LTER Grid Project, or the NBII web-services (for example, enrichment of EML document contents for controlled vocabulary).

◆ Editorials

Tacit Knowledge Acquisition: Approaching Replicability

- James Brunt (LNO)

One of the goals of the LTER Network is to produce a legacy of well-documented experiments - experiments that could conceivably be replicated in 20 or even 100 years in the future. With the advent of Ecological Metadata Language (EML) we are now producing a lot of well-documented data but having reviewed many of the metadata documents currently being harvested I would hesitate to say that we are producing well-documented experiments. Although there are some shining examples of extensive methods sections, the majority of metadata documents are inadequate for replicating an experiment. The alternative, the methods section of published papers, is an abstraction of reality that is neatly folded into a specific format but still doesn't provide enough information to reach the target of replicability 20, even 100 years from now.

Much has been done in the area of documenting procedures in the knowledge management (KM) community that can guide us in achieving the goal of replicability. The instructional or "how-to" aspects of experimental procedures are a form of explicit knowledge and EML lends itself well to documenting explicit knowledge. However, I make the assertion here that procedural knowledge in ecological studies has elements of both explicit knowledge and tacit knowledge (Box 1) thus making it complex and difficult to adequately capture using just procedural, "how-to", techniques.

Box 1. Explicit vs. Tacit Knowledge

Explicit knowledge - can be expressed in words and numbers and shared in the form of data, scientific formula, product specifications, procedural manuals, etc. This kind of knowledge can be readily transmitted across individuals formally and systematically. It can, theoretically, easily be processed by a computer, transmitted electronically, or stored in databases.

Tacit knowledge - is highly subjective and hard to formalize, thus making it difficult to communicate or share with others. Personal insights and intuitions fall into this category of knowledge. The subjective and intuitive nature of tacit knowledge makes it difficult to capture or transmit the acquired knowledge in any systematic or logical manner. For tacit knowledge to be communicated, it must be captured and converted into words.

We have experience in documenting procedures in LTER, however, that experience is limited in the area of tacit knowledge. The following example will hopefully illuminate the point.

A real example demonstrating the importance of tacit knowledge

Here's a real example that I use in Ecoinformatics training classes. The scenario is a 50-meter plant intercept transect established for the purpose of training technicians. Each of the two technician trainees is given a procedural "how-to" document describing the process developed by the investigator. Figure 1 shows the results of the two trainees compared to the baseline produced by the lead technician working with the investigator. Series 1 followed the procedure using explicit "how-to" knowledge only and did not produce results close to the baseline. Series 2 shows the results of the same two trainees after tacit knowledge was imparted by the lead technician.
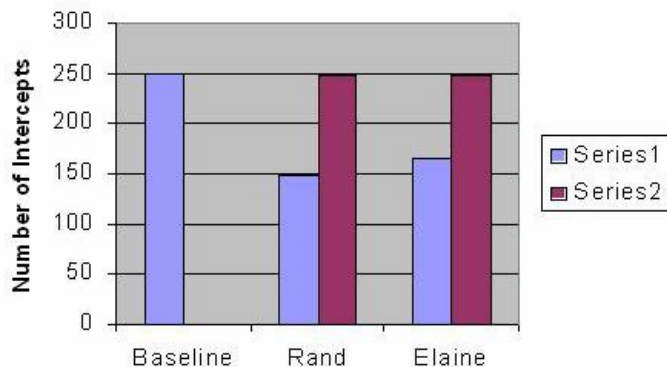
Documenting procedures works best when the people who generate the data, are the same people who generate the metadata, explain it to others, and coach them as they try to implement it. The "how-to" type procedure, however, fails to capture the tacit knowledge imparted in the "explaining" and "coaching" that is needed to achieve replicability.

Tacit knowledge acquisition

Many techniques have been developed in the KM community to help elicit tacit knowledge from an expert to capture the " who, what, when, how, and why" of tasks and events. These are referred to as knowledge acquisition (KA) techniques. These include various types of interviews (unstructured, semi-structured and structured), reporting techniques (such as self-report and shadowing), observational techniques, protocol analysis techniques, and the generation and use of concept maps, event diagrams and process maps.

To address this goal of producing a legacy of well-documented experiments will require engaging the knowledge management community. This community, however, is often divided into two camps with differing perspectives:

- Knowledge viewed as an object which relies upon concepts from Information Science in the understanding of knowledge. These researchers and practitioners are normally involved in the construction of information management systems, AI, etc. This group builds knowledge systems.
- Knowledge viewed as a process which relies upon the concepts from sociology and psychology. These researchers and practitioners are normally involved in education, psychology, sociology applied to communities of practice and are primarily involved in assessing, changing and improving human individual skills and behavior.

It's not hard to see that the two perspectives line up along the boundary between explicit and tacit knowledge. There is much to gain from engaging and drawing from both these groups in finding the tools and techniques that we need to manage the documentation of our ecological experiments. But, understanding the nature and value of tacit knowledge in documenting ecological experiments, at individual scientist and project levels, is critical to formulating a strategy for capturing and structuring such knowledge, and picking appropriate tools for their management.

In addition, describing categories, making distinctions, assigning names, sharing meaning, concepts and experiences, promoting understanding and making sense of the ecological world are fundamental to community inquiry and collaboration. To have a successful engagement with KM scientists, we need to devote time and resources to building a *lingua-franca*, a common, documented language at least for LTER, if not all ecology, to make it possible for KM scientists to work with us and for future generations to understand what we've done.

Finally, the time has come for us to stop ignoring our ecological legacy and instead to force ourselves to attack that other metadata element that's vastly more complex and much more difficult to figure out than the "unit type definition" - human behavior, expressed in scientific processes.

◆ Good Reads

Incorporating Semantics in Scientific Workflow Authoring

- David Ribes (PAL/CCE)

C. Berkley, S. Bowers, M.B. Jones, B. Ludascher, M. Schildhauer, J. Tao. Incorporating Semantics in Scientific Workflow Authoring. Proceedings of Scientific and Statistical Database Management, SSDBM'05, Santa Barbara, CA.

This article presented at the SSDBM (http://2005.ssdbm.org/program.html) and available on the LTER Information Manager Meeting website (http://gce-lter.marsci.uga.edu/lter_im/2005/app/resources.asp?webpage=references) describes an instance of the merging of ontologies with workflow systems in the with a particular workflow system called Kepler being used by the SEEK (Science Environment for Ecological Knowledge; http://seek.ecoinformatics.org/) together with a number of partners. Kepler is a promising research avenue, a tool designed with a generalized support for scientific workflows which are locally tailored for specific tasks through ontology-enabled, domain-specific customization. Workflows are the automation of scientific processes related to data manipulation and representation, while ontologies are system accessible scientific terminologies and computable linkages between them. This article describes how in combining ontologies and workflows a system for data discovery, manipulation and representation was created which is both domain specific and cross-disciplinarily configurable. The authors briefly describe problems in creating workflows and ontologies, and the future possibilities of automated data integration through re-use of semantically annotated workflows.

The Meaning of Everything

- Lynn Yarmey (PAL/CCE), Karen Baker (PAL/CCE)

Winchester, Simon. The Meaning of Everything: The Story of the Oxford English Dictionary. Oxford: Oxford University Press, 2003.

Pearl S. Buck noted that 'One faces the future with one's past.' As the LTER community plans for the future of data accessibility and interoperability, we may draw upon the experiences provided by an example from our recent past. In his book, The Meaning of Everything: The Story of the Oxford English Dictionary, Simon Winchester sends a quiet word of encouragement to those on the designing edge of data organization requirements. Our own metadata processes seem to have so much in common with the story he tells of the creation of the infamous English dictionary.

The OED began as an attempt by the Unregistered Words Committee of the Philological Society (founded in 1842) to mend and improve upon the lesser dictionaries of the age, though not long into the endeavor the need for a complete summation of the language became apparent. With supporting institutions and the editors themselves repeatedly underestimating the enormity and daunting complexity of the project, funding was a constant issue despite the collaborative network of volunteers numbering in the hundreds. The philosophical balances - between short and long-term goals, between meeting a budget and maintaining excellence, between highly specific definitions and the very general, and between creating a fixed standard while acknowledging the flexibility of the language itself - along with the practical consequences of each, were continually fine-tuned and adjusted over the 71 year project (the original estimate for creation of the OED was "no longer than 10 years"). For those of us looking to categorize and catalog our own site information, don't these balances sound familiar? As the LTER community moves with the tide of informatics to meet the present and future needs in the formidable realm of standards, we have the distinct advantage of facing the future armed with our past.

Gung Ho! Turn On the People in Any Organization

- Eda C. Meléndez-Colom (LUQ)

Blanchard, Ken . Gung Ho! Turn On the People in Any Organization. William Morrow and Co., Inc.

What has a book on increasing profits and productivity to do with Information Management? Maybe nothing, maybe all. When I first saw the cover of this book, I did not relate to the subject in any form, but it was recommended to me by a fellow worker, who has seen the work I have done and how I do it for 15 years. As I read it, I was amazed to learn how much we, the LTER Information Managers, apply the three main Gung Ho principles as we do our job: the spirit of the squirrel, the way of the beaver, and the gift of the goose. As odd as this may sound to you, the LTER Information Managers use these principles to do their work.

They are the principles learned by an executive from a General Manager of a plant, that has been set up to "save" the plant of its bankruptcy in 6 months, a year, the most. As she learns how the situation is in the plant, she also learn that while all the departments were completely unsuccessful in maintaining a high production, there was this single department, ran by a Native American whose productivity was outstanding. Learn with the Manager how this finishing Department was run using the principles thought by Andy Longclaw's ancestor, the Gung-Ho principles.

◆ Frequently Asked Questions

Data release policy interpretation in regards to the accessibility of Type I data

- K. Vanderbilt (SEV), J. Brunt (LNO)

Question: How should I interpet the data release policy in regards to when Type I data should be accessible to the broader community?

Answer: Investigators and projects should submit their data for release to the broader community within two years of collecting the data or after the main findings are published whichever comes FIRST. The spirit and intention of this policy as it was approved is that "2 years from collection", means 2 years after a single data collection event or field season - not 2 years after the study is "complete".