LTER Databits

Information Management Newsletter of the Long Term Ecological Research Network

Spring 2010

The depth and breath of thought in the LTER IM community is once again made apparent in this latest issue of Databits. Within, you will find a discussion of the software tools being used in the cyberinfrastructure development effort by the LTER Network Office. Lynn Yarmey provides a description of the Panton Principles for open data. Two articles discuss the ideas behind, and the challenges in implementing, Drupal-based websites. Philip Tarrant shares his initial thoughts as a new LTER IM. There are reports and updates on conferences and working groups. Margaret O'Brien explains the use EML in site data catalogs. Eda Melendez-Colom reflects on the role information managers play in education outreach. Don Henshaw shares his thoughts on a recent paper on data management models. And finally there is a list of upcoming conferences and events of interest to this community.

Featured Articles

Openness and Transparency in Development of the LTER Network Information System

edit

Mark Servilla and Duane Costa (LNO)

The May 2010 start date is a watershed moment for LTER that begins a five-year cyberinfrastructure development effort to implement the Provenance Aware Synthesis Tracking Architecture (PASTA) as part of the nascent Network Information System (NIS). Making this a collaborative and open process between NIS developers and community stakeholders is paramount to the success of the project. In support of openness and transparency during this development effort, key software tools have been identified and put into operation. These include:

1. **Subversion** for software revision control and management;
2. Edgewall Software's **Trac** for task management;
3. **Drupal** for content management and providing a framework for information dissemination; and
4. California Institute of Technology's **EVO** collaboration environment for real-time audio and video communication.

Subversion

Subversion (http://subversion.apache.org/), informally known as SVN, is an open source version control system that has been in use at the LTER Network Office (LNO) for a number of years, succeeding the previously supported Concurrent Versions System (CVS). Subversion utilizes the notion of a host *repository* as a container of digital objects, namely software source code files, that may be modified locally and updated to the repository as a new revision. Individual revisions are tracked through associated metadata and may be *rolled-back* to a previous version, if necessary. The LNO supports multiple SVN repositories for various projects, both internally and Network-wide, hosted at https://svn.lternet.edu. All LNO SVN repositories are open to the public. As such, all files in the LNO SVN repositories may be *checked out* with unfettered access. Insertions and updates to files in the repository, however, require user authentication and access rights to the specific repository. Authentication is integrated with the LTER user registry, which is maintained in a local Lightweight Directory Access Protocol (LDAP)

server at the LNO. Browsing the LNO repositories may be performed by using the web-browser *WebSVN* application at https://svn.lternet.edu/ or by using one of many SVN clients and connecting directly to a repository at *https://svn.lternet.edu/svn/<repository>*. The NIS repository (https://svn.lternet.edu/svn/NIS) is divided into two main categories, `documents` and `software`, each represented as a separate directory in the repository. Artifacts from NIS project planning and software development may be found in either of the two directories.

Trac

The NIS development effort is utilizing the Trac web-based application (http://trac.edgewall.org/) for task management through its ticket and issue tracking capabilities. Trac supports a minimalist approach for both wiki page editing and ticket management, in addition to a long list of externally contributed plug-ins that extends its base functionality. Trac replaces the Bugzilla system that was previously used for defect tracking at the LNO. We have successfully modeled the NIS software development methodology, *OpenUP*, to the Trac ticket system, thereby allowing us to create and track OpenUP related tasks. Tickets may also be created for managing defects and requesting enhancements to any NIS related software component. The NIS Trac is hosted at https://trac.lternet.edu/trac/NIS and the content is open for viewing by the general public. Creating and modifying tickets does require users to login through the Trac site (see top of web page) by entering their LTER user credentials. As mentioned, the NIS Trac ticket model has been modified to adopt the OpenUP task requirements. This process was simplified by using the *TracCustomFieldAdmin* (Version 0.2.2) plug-in. We have also installed the *TracGanttCalendarPlugin* (Version 0.1) for graphically tracking progress on specific tasks by displaying them within a calendar view and Gantt charts. Trac also supports custom report queries. These queries are created through Trac's *Custom Query* interface from within the *View Tickets* web page and may be modified by any user to selectively filter report ticket information.

Drupal

The Drupal content management system (http://drupal.org/) provides the web-based framework for the NIS project website. Its goal is to facilitate the dissemination of information pertaining to NIS development and collaboration. It is hosted at https://nis.lternet.edu and, like Subversion and Trac, is open to the general public for viewing content. The NIS Drupal website is intentionally simple to navigate and to find meaningful content, and uses the *Abarre* theme (http://drupal.org/project/themes?text=abarre). Primary information concerning NIS development is located within the `About` block of the left column menu, including brief descriptions of the PASTA framework, the NIS software development methodology, and community engagement. More detailed information about the components of PASTA is located in the *Architecture* forums section of the website. There is also a *Frequently Asked Questions* forum for general information about NIS development. Authenticated users may append comments to any forum topic, thereby allowing threaded discussions about topics. This particular installation of Drupal takes advantage of a small number of add-on modules. For WYSIWYG editing and file management, we use the *FCKEDITOR* and *IMCE* modules, respectively. *Pretty print* formatting of content pages is possible with the *PRINT* module. As with Subversion and Trac, user authentication is integrated with the LTER user registry (LDAP). Drupal integration with LDAP is simplified with the *LDAP_INTEGRATION* module. Finally, the *WEBFORM* module is used to generate custom web-based *forms*, which supports the input fields used to capture information for volunteering as a NIS *Tiger Team* member and for providing feedback about NIS development and collaboration.

EVO

The final software tool that is being used to support design and development activities of the LTER NIS is EVO (http://evo.caltech.edu). EVO is a collaboration environment that supports multicast audio/video and desktop sharing and is compatible on all three of the common computer operating systems (Linux, Mac OSX, and Windows). EVO requires the latest SUN Java run-time environment (JRE) and uses the Java *web start* process to download and execute the most up-to-date EVO software client, *Koala*. EVO supports a host-based service for collaboration that routes online conferences through a set of load-balanced servers. Video and audio quality can be adjusted dynamically to cope with changing bandwidth. The EVO project leadership has graciously provided the LTER Network with their own *community* presence, thus making collaboration between NIS developers and stakeholders much easier. A brief description of how to use EVO can be found on the NIS project website (https://nis.lternet.edu/NIS/?q=node/9).

An Introduction to the Panton Principles for Open Data in Science

edit

Lynn Yarmey (CCE/PAL)

The Panton Principles for Open Data in Science (http://pantonprinciples.org/), released in February 2010, are relatively straightforward but powerful guidelines to data publication online. These principles come from the digital curation community, an increasingly active consortium of librarians, information scientists, technologists and a small number of open-access publishers. In the months since the American Naturalist announcement regarding mandated data publication (Whitlock et al, 2010), issues surrounding licensing will likely remain at the forefront of attention. While simply putting data online may imply permission for data copy, reuse and republication, the Panton Principles encourage stating these intents specifically through the use of licenses in order to formalize the permissions associated with accessible data.

The Panton Principles seek to formalize the nature of data available online, stating that: "data related to published science should be explicitly placed in the public domain." As a quick introduction, the Panton Principles condense to the following four points pulled from the above website:

- When publishing data make an explicit and robust statement of your wishes.
- Use a recognized waiver or license that is appropriate for data.
- If you want your data to be effectively used and added to by others it should be open as defined by the Open Knowledge/Data Definition - in particular non-commercial and other restrictive clauses should not be used.
- Explicit dedication of data underlying published science into the public domain via PDDL or CCZero is strongly recommended and ensures compliance with both the Science Commons Protocol for Implementing Open Access Data and the Open Knowledge/Data Definition.

The pantonprinciples.org website goes into more detail about each of the elements and offers supporting information about what specific licenses may or may not be appropriate, etc. Endorsed by a small but significant mix of international people from universities, research centers, corporations, publishers and Open Data groups, the Panton Principles offer an entry point to the legalities surrounding what it means for data to be truly "open."

John Porter (2010) provided a review this year of the LTER policy for data sharing. Costello (2009) recently suggested formalizing the notion of data sharing by developing the concept of 'data publication' together with associated mechanisms for data publication. Emerging from within the digital curation community, the Panton Principles continue this data publication discussion to include licensing of data in order to communicate intent and permissions with potential data re-users.

For more information:

http://cameronneylon.net/blog/the-panton-principles-finding-agreement-on...
http://wwmm.ch.cam.ac.uk/blogs/murrayrust/?p=1939
http://blog.okfn.org/2010/02/19/launch-of-the-panton-principles-for-open...

Costello, M.J. (2009). Motivating online publication of data. BioScience 59(5),418-427. http://www.bioone.org/doi/abs/10.1525/bio.2009.59.5.9

Porter, J.H. (2010). A Brief History of Data Sharing in the U.S. Long Term Ecological Research Network. Bulletin of the Ecological Society of America 91(1), 14-20. doi: 10.1890/0012-9623-91.1.14

Whitlock, M.C., McPeek, M.A., Rausher, M.D., Rieseberg, L. and A.J. Moore. (2010). Data Archiving. The American Naturalist 175(2), 145-146. http://www.journals.uchicago.edu/doi/full/10.1086/650340.

Using EML in Your Local Data Catalog

edit

Margaret O'Brien (SBC)

All LTER sites create EML to contribute to the LTER network, but it appears that only a few use their EML in their own data catalogs (see below for links). It is possible to use the EML you contributed to the network to provide content for your local data catalog. The benefits are obvious: your local data catalog will show the same information as the Network's, and it is less work to use an existing metadata resource than to create another. All modern web-scripting languages have libraries for reading and transforming XML for web page content. The process described here is similar to the display of projectDB documents using eXist (Sheldon, 2009).

In 2003, the PISCO project installed a Metacat data catalog at UC Santa Barbara, and SBC became an "early adopter." We have continued to use the PISCO Metacat as our primary data catalog, and our EML was replicated to the KNB and LTER Metacat servers. Needless to say, SBC enjoyed the control and learning environment that this relationship provided. But recently, to reduce load on the PISCO IM systems, SBC began employing the LTER Network Office Metacat installation as its primary metadata repository. As part of that process, we created our own EML-display application instead of using a pre-installed Metacat "skin".

In a nutshell, we needed three things:

- Access to EML documents through a URL
- Local XSL stylesheets to transform the XML into HTML
- A transformation routine, which could be PHP, Perl, ASP, JSP, Java, etc.

In detail:

- **XML document**
  This is done for you, since, in addition to delivering HTML, Metacat also can deliver your EML as XML through a URL like this:
  http://metacat.lternet.edu/knb/metacat?action=read&qformat=xml&docid=knb-lter-sbc.5
  The URL contains parameter-value pairs to designate the document ("docid"), the requested Metacat action ("action") and the output format, in this case, XML ("qformat"). For the current purposes, the second and third parameters are fixed to "action=read" and "qformat=xml".
- **XSL transformation stylesheets**
  As part of our association with PISCO, SBC was already using a set of XSL stylesheets which were more modular than those shipped with Metacat and used by the LNO. As part of this project, SBC's stylesheets were further adapted to use a "tabbed" format for the major metadata sections. The XSL uses parameters ("<xsl:param>") to control which parts of the EML document are displayed. There is one main stylesheet (e.g., "eml-2.0.1.xsl") which handles the calls to templates for other modules. These stylesheets are fairly generic and easily portable (see below).
- **Transformer**
  A web scripting language is needed conduct the transform (passing on parameter values as necessary), and to send the resulting HTML to the browser. The basic steps are
  1. Read in the XML source
  2. Read in the XSL stylesheet
  3. Transform the XML with the XSL and send the results to the browser

Perl was used for SBC's project, but nearly all modern web-scripting languages include an XML library. Figure 1 shows sample code to process an XML document in Perl. This script operates on a URL which resembles: http://sbc.lternet.edu/cgi-bin/showDataset.cgi?docid=knb-lter-sbc.10

It should be noted that in practice, the program variables are likely to be configured outside of the script. Since EML documents can be quite long and complex, the XSL stylesheets also use several display parameters, and these will also need to be handled by the transforming script. Only one parameter is passed in this sample ("docid"). Your XML library will have instructions for passing parameters into the transformation.

Figure 2 is a screenshot of the output for an SBC EML dataset, which shows a view of the default "Summary" tab. The stylesheets also contain <xsl:include> statements for other page components, such as headers. As a test of portability, the MCR site has already installed both the Perl script and XSL stylesheets (Fig. 3, showing the "responsible-parties" modules). The configurations for application host and other parameters are set in a settings file called "eml-settings.xsl", also called by an <xsl:include>. MCR chose to further customize the display by editing the default CSS file.

Fig. 1. Sample code for reading XML and XSL documents, and transforming to HTML using Perl CGI. (click on figure to enlarge)

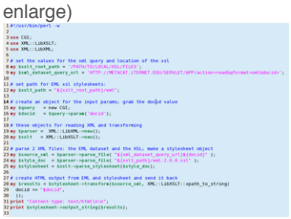


Fig. 2. Screenshot of SBC LTER's "Summary and Data Links" view of a dataset using new XSL stylesheets. (click on figure to enlarge)

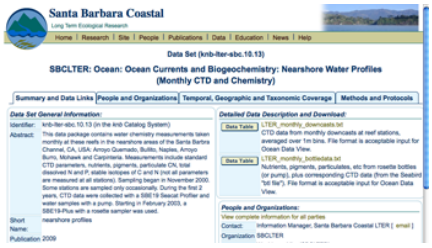


Fig. 3. Screenshot of a MCR LTER dataset with the "People and Organizations" view using SBC's XSL stylesheets and Perl CGI.

Since the script calls its XML source from the central LTER repository, any EML document in that repository can be transformed through this display. Simply replace the "docid" value in the URL above to view a dataset from your site through SBC's interface. During testing, non-SBC EML was viewed regularly in an effort to ensure that the stylesheets remained broadly applicable.

The current XSL stylesheets for EML that ship with Metacat are somewhat basic, and many of us have expressed a desire to see a more modular or tabbed display. PISCO and SBC intended to extend their stylesheet project, but lacked the resources to complete them.

Although the stylesheets are currently located in SBC's code repository, they could be further developed as a network resource and kept centrally. With input from our site scientists and other users, a group could streamline and standardize these according to recommended practices and contribute them to a broader community, and possibly offer another choice in the Metacat "skins".

A Sample of LTER site data catalogs which use EML:

- http://caplter.asu.edu/data (using eXist XML database)
- http://hfr.lternet.edu/data/archive.html (using eXist XML database)

- http://mcm.lternet.edu/data_home.htm (using LTER Metacat skin)
- http://mcr.lternet.edu/data (using this application)
- http://sbc.lternet.edu/data (using this application)

References

Sheldon, W. 2009. Getting started with eXist and XQuery. Databits, Spring 2009.

Acknowledgements

Chris Jones (PISCO) did more than 90% of the work on the original PISCO/SBC stylesheets in 2005, which are still available as a Metacat "skin".

Drupal developments in the LTER Network

edit

Corinna Gries (NTL), Inigo San Gil (NBII/LNO), Kristin Vanderbilt (SEV), and Hap Garrit (PIE)

Introduction

As we all know, maintaining a website that was created employing less than perfect design and coding principles is somewhat tedious and time consuming. General updates may have to be made to every page, adding a new section can be cumbersome, quickly adding a news item requires html knowledge, and rearranging things in response to new insights of how people would like to navigate is almost impossible. This is where content management systems (CMS) come in. With strict separation of content, organization, layout and design they overcome the above mentioned obstacles to a more dynamic and responsive approach of web site maintenance.

Most content management systems on the market do a good job in providing these aspects, but some are going further and really are hybrids between a content management system and a development framework, which is what is needed for the average LTER site web application. While most CMS can handle personnel profiles, publications, calendars, and images in addition to static content, the development framework aspect of the system enables handling of more specialized content types, like project descriptions and EML metadata. In the open source realm the most widely used CMS are Joomla, Wordpress, Plone and Drupal. These are all perfectly valid content management systems that offer similar functionality with Plone and Drupal providing the highest degree of flexibility. There are plenty of web accessible discussions about why some folks prefer one CMS over another. In these discussions, where you may draw a parallel to the Apple vs. Microsoft eternal and somewhat boring, but always passionate discussions, you may find arguments in favor and against a particular CMS. Some of these arguments seem to gain some community traction and are based on tangible evidence, but plenty of unfounded claims always prevent a conclusive analysis.

Several reasons have led to choosing Drupal over Plone: Plone's programming language is Python vs. PHP for Drupal and Plone requires a more sophisticated hosting environment than Drupal. Due to the flexibility they provide, the learning curves are high for both systems, but Plone is considered still more complicated in the initial site set up than Drupal while the content maintenance is more user friendly in Plone (Murrain et al. 2009). However, it is expected that Drupal 7 will provide a more intuitive user interface for the content administrator. Drupal's strongest points are the so called taxonomies or keywords. Every piece of information can be tagged with a keyword and displayed in groups based on these keywords. This allows for extreme flexibility in accessing information. A very simple example is used in the LTER IM website. Two different views of site based information are provided. One is by subject area (e.g. site bytes for 2008) and the other by site name (e.g. all site bytes and more for Andrews). Searches for information can also be made more successful. In another example from the LTER IM website, a person profile can be tagged with the same keywords that are used in the knowledgebase articles making people show up in search results as experts in a certain field.

The data model in Drupal

The most popular Drupal instances use MySQL to store all their content. At this point, it may be convenient to learn a bit of the Drupal community jargon. One of the Drupal claims to fame (and motto) is "Community Plumbing" (meaning community development and support, with well over half a million sites powered by Drupal). Although Drupal lingo is somewhat obscure to uninitiated we'll try to establish a map between known information management concepts and the Drupal language to understand better the documentation created by the Drupalistas.

The basic Drupal unit is the 'node' - a single record or an entry in a table of the database. A page, a story, an article, a blog post, a photo, etc. are all examples of nodes. Nodes are entries in the so called Content Types. Content Types are the Drupal categorizations of information. You can think of "content" as "information", and "types" as "tables" in the back end database. A better, more sensible, broader definition of a content type is a container for a specific category of information. Nodes stored in content types can be linked in one-to-many or many-to-many fashion. For example, a person may have authored several data sets, and a data set may have many contributors. A research site description, lat/longs, elevation and datum would constitute a 'node' of the content type 'research site'. Several research sites can then be linked to a project description or a data set.

This clearly is relational database design which can be directly inspected in the underlying MySQL database (RDBMS). And because it is a RDBMS the data can be manipulated outside of Drupal via third party administrative applications or custom code. However, access through third party applications is rarely used by the Drupal aficionado but is useful if you want to load information in bulk from a different database.

Another appealing Drupal aspect is that all you need to handle it is a web browser -- any web browser -- Safari, Firefox or Internet Explorer. Drupal functionality (like the other mentioned CMS) can be classified into a modest yet powerful core functionality and a vast extension of functionality, via "modules". The core functionality is preserved and updated by the original developer team (Dries Buytaert, 2010) and the extended functionality is provided by a large community of developers through a unified portal and conforming to a set of development guidelines, principles and open source philosophy. This dual character (core and extensions) resembles the Linux development model. All these custom extensions (modules and themes) are offered through the Drupal portal at http://drupal.org/project.

Currently, our LTER Drupal Group uses many extensions to the core and has developed custom content types for personnel, research site, research project, metadata for dataset with data table and attribute types. We also benefit from optimized modules that manage bibliographies, image galleries and videos developed by the Drupal community.

Using extended functionality it is possible to provide XML, (and with a bit of work, EML, FGDC and Darwin Core, see also developments by the U.S. Geoscience Information Network for ISO) as well as PDFs, Excel spreadsheets, Word documents, tabular views, charts and the like. Our group uses the powerful views API module extensively. The views module allows us to offer the content in many user friendly and intuitive layouts. The views module is nothing but a GUI into the creation of SQL queries, coupled with the actual final web layout. All views are of course managed by the same database, security, logs, user management, LDAP connectivity, SSL encryption and the list goes on.

In summary, the content types are specialized repositories of content. Creating a custom content type in Drupal is fairly simple and includes an associated, configurable input form to capture and edit the information. The simple web form used to create content types triggers a Drupal process to build the underlying database tables to hold the information. In other words we have developed a very simple EML, personnel, bibliography, media and project editor using the Drupal data model. The following figure shows a small subset of the metadata data model in Drupal. (Right-click and View-image to see full resolution.)

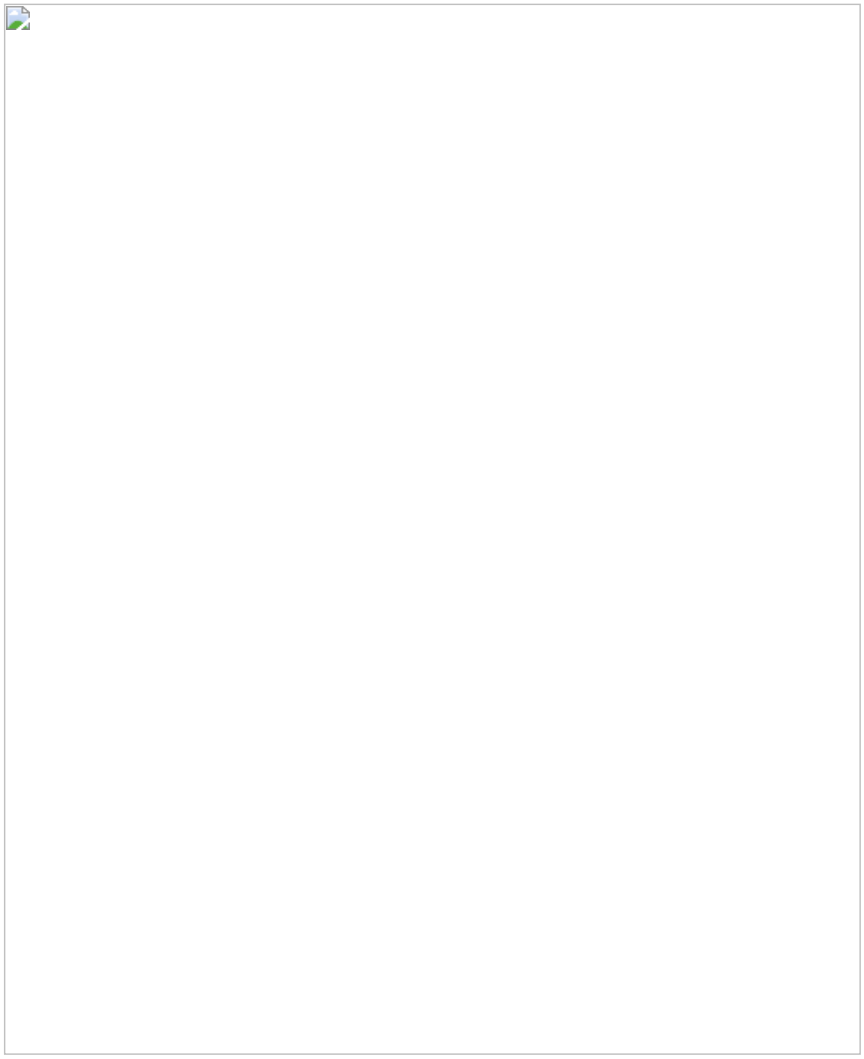


Figure 1. This is a simplified, metadata constrained, Drupal data model diagram.

Most of the core tables have been omitted, except for the table director "node", in the red area. The tables shown here are some of the tables related to the management of ecological metadata information. Five colors denote different categories of metadata -- Light green for the personnel directory, Pink for the basic metadata details, Teal for the information pertaining the structure of the tabular data and in yellow, the basic georeferences. Variable or attribute information is located on the orange area.

The main tables in the diagram above have a "content_type" prefix. Add to the prefix "data_set", "person", "research_site", "variable" or "data_file", and you have the main tables for the information categories, dataset, personnel, location, attribute or variable and file structure. The tables with prefix "content_field" could be of two types, but generally contain one-to-many indexes. One type contains placeholders with multiple values, such as "date values". Sometimes a dataset may have multiple dates associated with it, therefore the "content_field_beg_end_date" contains the multiple values, avoiding complete de-normalization. In this case, the key "delta" denotes a number of multiplicity in the multiple relation, that is if a dataset has three dates, the delta will have a value from 0 to 2. In the other case of "content_field" tables, we have the referential tables.

Experiences with migrating existing site information

NTL: We are still in the very early stages of migrating our website into Drupal. However, we already accomplished setting up two websites that are accessing the same underlying database, one for NTL LTER and one for the Microbial Observatory because the overlap in people, publications, projects, and datasets necessitates this. This approach allows us to add websites for the funding cycle of a project and then fold the information into the NTL website when funding ends. We received the content type definitions developed at the LTER Network Office (LNO) and were able to import them seamlessly into our application. Although we already have most of the information required to populate these content type, originally we used a slightly different data model. Queries are currently being developed to migrate the data into the new structure and the basic data model will have to be extended to accommodate everything we will need to replicate functionality from our original website.

LUQ: Luquillo is nearly complete in its migration process. Many new views are offered to the user. All the content is tied together minimizing the risk of running into a page without further suggestions or links. All the content is linked to the record or node level, and powered by a custom controlled vocabulary whose adequacy is being tested against the goal of unearthing related information (discovery functionality).

SEV: The new Drupal website will soon be unveiled. The old SEV website was implemented in PostNuke, a CMS that never gained much popularity. Because SEV content was already in the MySQL backend of PostNuke, Marshall White (LNO) was able to migrate much of it into the Drupal MySQL database with relative ease. The Drupal website also incorporates an EML editor, which thrills SEV IM Vanderbilt to pieces. Inigo San Gil (LNO) wrote a script that parsed all SEV EML files into five content types created to hold metadata:

- **Data Set** - contains discovery metadata (through Level 3)
- **Data File Structure**- details about the data containing entity (header lines, file type)
- **Variable** - captures info about variables (EML attributes)
- **Research Site** - captures info about plots or research sites
- **Project** - Higher level project information, which can encompass several DataSets

All of the content from the EML will have to be reviewed to ensure the parser didn't miss anything and because EML often contains peculiar formatting as a function of tags. Because each of the content types is essentially a form, new metadata can be entered via the website and stored in MySQL. A style sheet will be

created that will display the metadata as a text file or .pdf for user download.

PIE and ARC: The sites hosted at the Marine Biology Laboratory (MBL) are next to start the migration process. The IMC allocated some modest funds to train the information managers through the development of their new information management systems. This process is expected to start sometime during the summer of 2010. Other sites have shown interest in these developments, including CDR, BNZ, CWT and KNZ, and have not ruled out a possible adoption of this enabling data model. Generally, the process of adoption involves several steps:

1. Identifying the types of information being managed by the site. Typical categories include projects, metadata, data sets, data (tabular data, shape files, etc), personnel profiles, publications, news, stories, field protocols, photos, videos, research locations
2. Installing Drupal and several modules
3. Importing the predefined content types, which will generate the data tables in the database and input forms
4. Selecting a design theme
5. Customizing everything and migrating information

Synergistic efforts

Two groups outside of LTER have adopted our content models, University of Michigan Field Station and the Wind Energy group at the Oak Ridge National Laboratory. And several others are interested in learning from our experience: The National Phenology Network, The US Virtual Herbarium, The National Biological Information Infrastructure, Spain SLTER, FinSLTER, Israel LTER, Taiwan LTER. Other groups active in eco-informatics already have embraced Drupal as their development framework and are publishing modules for others to use: Encyclopedia of Life modules for biodiversity information management, U.S. Geoscience Information Network tools for metadata management in ISO, and VitalSigns modules for citizen science applications. The Biodiversity Informatics Group at the MBL is creating the software behind the Encyclopedia of Life (EOL - http://eol.org) - a single portal providing information on all 1.9 million known species. The infrastructure is seamlessly aggregating data from thousands of sites into species pages in the Encyclopedia using novel informatics tools to capture, organize, and reshape knowledge about biodiversity. The group collaborates with data providers. The information is then indexed and recombined for expert and non-expert users alike using aggregation technology to bring together different data elements from remote sites. All of the code related to the EOL is open source and available via Google Code (http://code.google.com/p/eol-website/) or GitHub (http://github.com/eol/eol). LifeDesks is an Encyclopedia of Life product also developed by the Biodiversity Informatics Group at MBL which uses Drupal as a platform (http://www.lifedesks.org/modules/).

Concerns and opportunities

Here is a list of common concerns heard. Some are Drupal-specific, and need to be carefully addressed. Other concerns, while valid, apply to any type of information management system. We address here most of the concerns expressed by the folks that discussed this Drupal based IMS with us, but note that the issues that apply to all systems. Some view the use of Drupal as "putting all your eggs in the same basket" or "locking ourselves into one system". This is a valid concern, however, it also is a major opportunity, easing collaboration among sites and providing the basics for web sites managing environmental information beyond the LTER network "out of the box". Although the system provides enormous functionality, the data are still in a database that can be accessed, managed and used by any other system. And this system provides tight integration giving a complete vision to all aspects of information management, from details to the large picture, but always connected and contextualized. Because Drupal is a 'framework' that defines certain aspects of program interfacing, all developments being done can be transferable. That is true not only for content types as described earlier, but also for modules containing programmed functionality and so called themes. Themes determine layout, design, i.e the look and feel of a site. Many Drupal themes from professional graphic designers are available for free download and can be modified to fit the particular needs or preferences. In addition to the plethora of modules available, custom functionality for LTER sites can be developed and widely used. For instance, at NTL we are intending to program a query application for data access based on the stored metadata (the same functionality that we currently have as a Java program). Any site that uses the same metadata content type (or an extension thereof) and stores its actual research data in the same MySQL database Drupal is using will be able to deploy that module.

Security

The issue of security plagues every computing application, even the most secure ones in our nation. Drupal does not seem to be more vulnerable than any other system. Zero-day threats related to PHP, Java, Apache or Javascript make any system, including Drupal, vulnerable. However, new releases in the core, and active policing of the extensions can prevent many of the malicious attacks, by following normal Drupal upkeep processes.

Some computer scientists think less of PHP. And they might be right. But PHP has come a long way and does well what it is intended for. Facebook, Flickr, Wikipedia are among some of the web powerhouses fueled by PHP. As for Drupal, add The Onion, MTV UK, the World Bank, the White House and the Economist to these sites. As mentioned above, developing a new content type basically is developing a data entry form in Drupal. And this process is fairly simple and fast. Therefore, at NTL we will be exploring this option for research data entry and hopefully reduce the number of spreadsheets that have to be uploaded to the database. Drupal supports this by providing sophisticated user access management, which allows us a fine-grained user management for different entry applications. Rapid evolution. As we write this, a team of Drupalistas is working on purging the last 100 critical bugs of the next big major release of Drupal, Drupal 7.

Some are concerned because it is not backward compatible. Drupal 7 comes with a number of usability issues addressed, and new features, including the integration of critical extensions into the Drupal core, such as CCK and 70 others. Our group relies heavily on the Custom Content Kit extension (CCK). While some applications will break, we will eventually embrace the evolution. Staying with older, deprecated systems eventually prevent innovation and foster security risks. Of course, this is not a Drupal-specific issue, but an IT issue. It just so happens that in our niche of specialty, the fundamental technologies advances condition severely any mid-term plan (3 years or more). Software projects that do not keep abreast of concurrent advances risk becoming irrelevant by the time the products are deployed.

Resources

- Drupal open source content management system: http://drupal.org/
- Encyclopedia of Life Drupal modules: http://www.lifedesks.org/modules/
- Joomla open source content management system: http://www.joomla.org/
- Murrain, M., L. Quinn and M. Starvish, 2009. Comparing open source content management systems: WordPress, Joomla, Drupal, and Plone.
- http://www.idealware.org/sites/idealware.org/files/idealware_comparing_o...
- Buytaert, D. 2010. State of Drupal. Keynote presentation at DrupalCon 2010, San Francisco, CA http://www.archive.org/details/Css3TheFutureIsNow
- LTER IM website: http://intranet.lternet.edu/im/
- Plone open source content management system: http://plone.org/
- U.S. Geoscience Information Network (USGIN),Creating ISO 19139 metadata through Drupal Views and Views Bonus Pack: http://lab.usgin.org/groups/drupal-development/creating-iso-19139-metada...
- VitalSigns: http://www.vitalsignsme.org/
- WordPress open source content management system: http://wordpress.org/
- Chronological discussion spanning 5 years on the Drupal vs. Plone topic http://drupal.org/node/13733

SIO Ocean Informatics Update: Growing Infrastructure in Support of Scientific Research

edit

Karen Baker (PAL, CCE) and Jerry Wanetick (CCE, PAL)

We report on the growth of an information infrastructure that began with the Ocean Informatics Initiative in 2002 at Scripps Institution of Oceanography. The initial aim to support the scientific research of two LTER sites, PAL and CCE, has expanded to include additional projects. In order to inform our future development, we look back on some of the milestones.

We define information infrastructure as encompassing both computational and informatics capacity. Two key points have guided the work of Ocean Informatics. First, we view the growth of infrastructure and its articulation as an ongoing process -- both conceptually and in practice. Second, we recognize the multi-dimensional nature of infrastructure. Working within the context of the intertwined social, organizational, and technical dimensions is sometimes called a 'sociotechnical' approach.

Infrastructure comes center stage as a result of expanding expectations associated with data. Today, there is an increase in data and information management work due to the increasing quantity of data and its diversity, the increasing sophistication of data analysis and visualization, and the increasing number of standards and exchange requirements. There is concurrent expansion of scientific work as researchers assume responsibility for not only data use in support of traditional local scientific work but also for preparation of data for reuse by disciplines outside of the traditional boundaries of the intended domain. Data reuse is stimulated by new, integrative, science approaches and goals.

Tasks related to instrument platforms and computational hardware have evolved, moving from a single central computer to distributed computational arrangements and hybrid storage solutions as well as from independent applications to distributed information systems, online data repositories, and web service architectures. And the next steps are in sight: machine virtualization and cloud computing. In addition, analytic work has exploded, now including collaborative efforts and community processes particularly those relating to semantic issues and standards.

Four elements of the Ocean Informatics (OI) infrastructure are summarized in Table 1. Each element is described in three dimensions: social, organizational and technical.

Table 1. Multiple Dimensions of some Ocean Informatics infrastructural elements

Infrastructure	Dimensions		
Elements	Social	Organizational	Technical
Foundation	Roles	IM strategy	Design process
Collaboration	Teams	Shared resources	Shared solutions
Networking	Communities and communication	Policy, personnel, resources, and identity	Online connectivity and applications
Environments	Learning	Information	Distributed collaboration

1. Foundation

The foundational elements of our informatics approach are the roles that delineate the distribution of work, the information management strategies that frame the work, and the design approach that defines the ongoing work process. With individual roles within information management evolving as rapidly as computer applications and data requirements, the work of OI participants includes mediation, translation, and articulation. We anticipate mediation requirements increasing over time even as workflow tasks become more automated. We foresee information expertise requirements expanding and diversifying from need for a programmer, a system administrator, and a data manager to include the expertise of information managers, systems architects, infrastructure designers, informaticians, and data scientists. Organizationally, we recognize the importance of developing an information management strategy and to do so in multiple arenas at multiple levels - group, project, community, institutional, and domain. First and foremost this involves understanding informatics as a scientific discipline replete with theoretical and empirical concerns. Our understanding of the design process has changed as computational capabilities have moved from ftp to gopher to WWW and most recently to emergent immersive environments framed by Web 2.0 and Google Earth. The OI design approach is integrative, keeping the long-term firmly in mind while undertaking everyday work. We've moved from focus on individual file systems to relational information systems to a vision of distributed systems, abstraction layers, and web services.

2. Collaboration

New types of interdisciplinary collaboration are key to imagining and planning for contemporary connectivity. For Ocean Informatics this has taken the form of multi-project, collaborative teams that bring together different disciplines (e.g. biology, physics, and science studies), different projects (e.g. LTER, CalCOFI, and SCCOOS) and different organizations (university and government labs). In order to support needed expertise and facilitate shared resources, a recharge facility, The Computational Infrastructure Services (CIS), was established in 2008 at a division level. Shared solutions have expanded with CIS to include desktop support, a ticketed help line, shared printers, and augmented storage. On the horizon are virtualization of servers and participation in a collocation facility to address physical platform location at an institutional level.

3. Networking

As the roles associated with data and information continue to change and diversify, communities and communications become important to formation of working groups. Within the LTER an information management committee represents an active community of coworkers that is drawn upon. Organizationally, there are policy, personnel, resources and identity to consider. At the institutional level, we have established a co-directorship of OI that reports at the division level. The technical dimension of networking involves developing and maintaining online connectivity and applications. Technological services and tools enable bridging individual repository arrangements but creating a coordinated web remains a grand challenge. Networks require extensive infrastructure in place at the individual, institutional, and national levels.

4. Environment

In planning and carrying out the work of Ocean Informatics, we emphasize learning and a learning environment by continually forming and reforming reading groups and creating professional development and leadership opportunities in addition to training in new technology use. Organizationally, we focus on infrastructure as central to creating an information environment. Such arrangements ensure data and information are able to travel among laboratories, crossing projects and data repository boundaries. A distributed, collaborative environment is the very public, online digital realm within which we operate, and the substrate for much of our work.

Table 2. Ocean Informatics Timeline

Time frame	Milestone	Implementation Details
2002	Need for new approach to shared infrastructure recognized	Conceptualized Ocean Informatics as infrastructure to support PAL LTER
2003	Ocean Informatics Initiative launched	Baker, Wanetick, Jackson, 2003 paper
2004	Shared servers & backups expanded	iOcean, server online
2004	Multi-project effort launched	Began funded support with CCE LTER

2005	Staff expanded and informatics reading group initiated	Discussion groups launch with Berners-Lee and The Semantic Web
2005	Dictionary work launched	Unit Dictionary prototype followed by implementation of units in local system
2006	Open Directory implemented	Federated authentication and authorization; LDAP and Kerberos
2006	Collaboration server and software instantiated	iSurf, collaboration server online
2006	Design studio established	Physical manifestation of OI with design table to facilitate co-design activities
2007	Initiated collaboration with WHOI IM	Began joint discussions; resulted in paper by Baker and Chandler, 2008, DSR
2007	Digital storage infrastructure expanded	From physical disks to RAID configuration
2007	DataZoo information system launch	Development of multi project architecture
2007	Added two new projects	Began funded support with CalCOFI
2008	Recharge Facility established	Desktop support initiated with centralized remote-office; shared printing established
2008	Semantic dimension added to DataZoo	Launch of qualifier system with units and attributes as DataZoo semantic integration
2008	DataZoo information environment launch	Multi-component approach for differing data collections
2009	Ocean Informatics institutional integration	Ocean Informatics directors report at division level
2009	DataZoo web service architecture launch	LTER Unit Registry, OI Controlled Vocabularies, OI Plot Server
2010	Facility Services expanded	Help request implemented; cishelp@sio.ucsd.edu
2010	Collaboration server updated	vSurf, server virtualization online

Toward an Information Environment

Table 2 provides a timeline of Ocean Informatics milestone events.

Co-founders Karen Baker and Jerry Wanetick summarized the initial Ocean Informatics vision in two early publications highlighting design, collaborative care, informatics and the concept of an information environment (Jackson and Baker, 2004. Ecological Design, Collaborative Care, and Ocean Informatics in Proceedings of the Participatory Design Conference, 27-31 July, Toronto; Baker, Jackson, and Wanetick, 2005. Strategies Supporting Heterogeneous Data and Interdisciplinary Collaboration: Towards an Ocean Informatics Environment in Proceedings of the 38th Hawaii International Conference on System Sciences (HICSS), 3-6 January, Big Island, Hawaii, pp. 1-10, IEEE, New Brunswick, NJ).

It's a long way from the time of individual punchcard deck submissions and centralized computer centers to working on shared infrastructure with distributed servers in collaborative arrangements. Ocean Informatics strives to maintain sensitivity to the new understandings and transformational aspects that emerge from the interplay of traditional computational arrangements with new concepts featuring design processes and collaboration enabled by new technologies. Attention to development of multi-dimensional infrastructure facilitates the move from individual data services and independent systems to new types of information environments.

Commentary

LTER site Information managers balance site and network demands

edit

Don Henshaw (AND) and Margaret O'Brien (SBC)

In listening to LTER Information Managers (IMs) describe potential uses of supplemental information management (IM) funding, the most obvious observation one could make is that there is no shortage of ideas on how this funding could be used. NSF has presented an opportunity to LTER sites to propose information management efforts related to better standardizing data, metadata, and approaches to facilitate the use of site data in synthesis projects. The IMs shared proposal ideas and explored potential collaboration in two April video teleconferences, which were cumulatively attended by 26 representatives of 20 LTER sites and the Network Office. It is clear that any new resources dedicated to information management are welcome (a nice way of saying desperately needed), and all IMs are challenged to balance focus between their site concerns and the capacity to participate in network-level activities.

While the proposal ideas ranged from specific tasks to generalized solutions, several common themes were identified that could become the basis for future collaboration, production workshops, or topics for our annual meetings. Some common themes:

- The use of a common data model to manage and store metadata to allow common tools to uniformly deliver metadata as EML or generate queries of the data (e.g., using the Drupal content management system)
- The need for tools to facilitate the capture of metadata at the front end of experiments such as web user interfaces
- The use of best practices as a guide for improving and completing metadata elements
- The need to develop or generate EML for geodatabases
- The installation of sensor networks and the need for tools to facilitate the quality assurance of data, and generally the need to improve the flow and processing of collected data into managed information systems
- The incorporation of standardized approaches such as the controlled vocabulary or unit dictionary into site information systems

These themes represent the spectrum of necessary site work required to begin paving the way for more seamless and interoperable access of site data throughout the LTER network. Our community must now address the best use of funding opportunities to meet this daunting challenge.

This NSF supplemental funding opportunity is likely offered to remedy recent LTER data access issues, to help position LTER for future funding similarly to other environmental observatories, and to provide resources to sites in parallel with development of the Network Information System (NIS) at the LNO. The complexity of integrating the many diverse and legacy data sets of the LTER sites has limited the development of cross-site data resources, but the success of LTER synthesis initiatives and system development in creating synthetic data products is paramount. To this end the LNO is embarking on a 5-year operational plan to address the broad integration of LTER site data, and LTER site IMs will be asked to engage at several levels to assure accomplishment of this plan. Here again, the IM community must consider the nature of their network-level involvement to most effectively take advantage of this opportunity to meet this challenge.

The LNO operational budget provides significant funding for meeting and workshops of varying types. The following list of IMC workshop and activities demonstrates the increasing number of community activities and opportunities for individual involvement.

Planned or in progress for 2010:

- LTERMaps post-ASM workshop: LTER site maps with connections to the SiteDB network database
- Unit Dictionary post-ASM workshop: Management tools for standardized units
- EML Best Practices production workshop: Revise and extend the existing best practices document
- Drupal training workshop: Installing and configuring the Drupal CMS for sharing common tools
- Network databases redesign/web services: Redesign NIS administrative databases and plan web services
- IMExec participation at the Science Council meeting: Strategic planning for the NIS, legacy databases, communication
- IMC Annual Meeting at Kellogg: IMC governance, EML best practices and evaluation metrics, Network databases redesign, GIS, controlled vocabulary, units dictionary
- IMC Annual Meeting at Kellogg (extra 3rd day potential topics): LTERMaps, EML Best Practices follow-up, EML Dataset Congruency Checker (EML Metrics), Relational data model comparisons, initial planning for attribute standardization and dictionary, Kepler workflow software workshop, sensor networks and quality assurance, EML for Geodatabases

Planned or proposed for 2011:

- ClimDB production workshop: integrate ClimDB as a module of the NIS
- EML metrics and congruency checker 'Tiger team': development of the NIS data loader
- Training module on web services for consumers: instruction for site information managers on using content from Network databases at sites
- EIMC meeting: The annual IMC meeting will be conference-style with participants from the entire IM community, and a theme centered on data integration for synthesis

The success of the LTER Network Information System will be largely dependent on site efforts to produce quality data and metadata.

Equally important may be the active involvement of the IM community in building the NIS and in standardizing approaches for network-level synthesis of site data resources. The challenge is on. The IMC co-Chairs intend to provide ongoing commentary in future issues of DataBits to feature planning and progress in this crucial effort.

First Impressions

edit

Philip Tarrant (CAP)

Joining CAP LTER as the new Information Manager in January was a return to familiar surroundings. I worked here as a research technician and I pursued my graduate studies with assistance from CAP. The project is based in a new building now and is part of an expanded Global Institute of Sustainability at Arizona State University, but many of the old faces are still there and the good work goes on.

Now that I have had a few months to digest my new role, I am excited by the tremendous additional value that can be realized by providing public access to these long term datasets. However, a lot of the work I do is at the back end of the process, manipulating data that arrive (if at all!) in disparate forms and sometimes of questionable quality. This begs the question: how can the Information Manager of an LTER encourage the adoption of standards and quality management in the body of researchers with which he/she interacts?

Coming from a background of running business improvement projects, I have been left with the strong conviction that 'process' is everything. Well defined, simple to follow processes, supported by effective productivity tools and clear, sensible standards, have a way of making sense to the people that use those processes. If individuals can see the value of a process and they do not have to 'tussle' with it in order to complete their objectives, they can buy-in to that process as an aid to getting the job done. But it is also true that our researchers have responsibilities in this matter. When they accept funding from CAP they also commit to supply their data for wider distribution. In their role as project managers they should be monitoring the quality of their data and the adherence of the project to any organization defined standards.

Clearly, it would be wonderful to have researchers knocking down my door to get my input and support at the front end of their project and some do ask for help. However, I suspect that I will also have to strongly remind others of their responsibilities. I hope that as a new IM I will be able to generate enough perceived value that my colleagues see me as someone who can benefit their project rather than as the 'old nag' that lives in the office around the corner.

Either way, I have started documenting our processes; aiming to understand how these processes operate today and considering how they should operate in the future. I found the information on the LTER IM website to be very helpful in building this picture and I hope to use this existing material as we continue developing our information management processes. I plan to work with the researchers to see how we can directly support their data collection efforts. My goal is to encourage CAP researchers to deploy the LTER data standards and prepare quality metadata as an integral part of the data collection phase of their research rather than as an afterthought at the end of the project.

Now as I remember it, that's the hard part.

Developing a Drupal "website-IMS" for Luquillo LTER while learning Drupal

edit

Eda Melendez-Colom (LUQ)

LUQ's Drupal New Website under construction (http://gorilla.ites.upr.edu/) is the answer that, for years, LUQ IM had searched for in order to have an interactive-database web site. What we were not aware at the beginning of this ordeal was that the same system that will serve as a website will also hold all LUQ metadata and data. As a content management system as well a "content management framework" (CMF) (http://drupal.org/getting-started/before/overview), Drupal has all the characteristics LUQ needs to finally develop its long-time anticipated and promised "information management common framework"

(http://gorilla.ites.upr.edu/node/definition-common-management-framework). In addition, the LNO staff has added some functionality to the Drupal system that will facilitate the entry of metadata into the system and the generation of EML packages from the system.

In other words, LUQ will have an integrated system that will serve as a website, a file system, a relational database management system (provided by MYSQL), a metadata and data depository, and EML package generator, where all the information is or has the potential to be interrelated. I like to call this system a "website-IMS", just to make it short.

The two major and innovative characteristics of this new system are that it serves both as a website and as an Information Management System (IMS) and that it has the potential to interconnect all its information. A graphical representation of the latter would include sets of keywords joining and interconnecting all the information in such a system. Such a diagram would clearly depict the central role that keywords (Taxonomies, as called in Drupal) have in this system. Taxonomies determine the website-IMS capability to ultimately connect all types of information (Content Types, as called in Drupal). Nevertheless, simplicity of the representation of these concepts would preclude the complexity of the implementation of such a system.

First, in Drupal all individual information becomes an entry in a database (a node, as called in Drupal). The soon-to-be old LUQ web site contains almost 4GB in data, metadata, photos, presentations and descriptive html files, setting aside its remote sensing documents. All this information is displayed statically by the current web server using HTML. Every single document of the old web site will become a node in Drupal; it might be story, a page, a blog entry, one of the custom LTER content types designed by Inigo San Gil and Marsh White at the LNO (a Data Set, Data File Structure, Research Site, Research Project, an Institution Links or Variable -EML Attributes-), a Biblio, or a person. This only gives us an idea of the complexity of migrating LUQ's website to Drupal.

Second, all nodes must be assigned a set of keywords. The complexity of this process goes beyond entering a set of values in a table. The following is a list of content types and their relation to the set of keywords or Taxonomy:

- **Data Set** - Set assigned by the owner scientist; One data set will be related to one or many Data Set keywords (Data Set keywords (Drupal Taxonomy)
- **Data File Structure** - Indirect: receives keywords from the Data sets they are related to in content type "Data Set" (Data Set keywords (Drupal Taxonomy))
- **Variable** - Each becomes a keyword itself that will be related to one or many Data Files [Variable's labels keywords (Drupal Content Type)]
- **Research Site** - Each becomes a keyword itself that will be related to one or many Data Sets; Each Data Set will be related to one or more Research Site [Research Site Titles keywords (Drupal Content Type)]
- **Research Project** - Each becomes a keyword itself that will be related to one or many Data Sets. A Data Set will only be related to one Research Project. [Research Project Titles keywords (Drupal Content Type)]
- **Publications** - Extracted from the set of keywords in the publication; Old publications (1980' and some of 1990's) did not have keyword assignment (Publication keywords)
- **People** - Extracted from all information entered in the LNO's personnel database profile for each person (Several sets of keyword types: habitat, organism, etc.)

The assignment and implementation in Drupal for the Data sets' keywords was completed during my 3-week "micro-sabbatical" in the LNO early this year. The process started around 12 months before that when the sets of keywords for each of the almost 150 LUQ data sets were extracted from the old-format metadata forms. The last part of this process represents a teamed, structured coordination effort that lasted almost 9 months.

The following are the steps taken and the iterations of each step to complete this process.

- Extract a set of keywords assigned in metadata by scientists - every time list was edited or created
- Keep a list of the keywords assigned to data sets identifiers - every time keywords were edited or corrected this had to be done again.
- Create a relational database table of the keywords to do QC on the list
    - Eliminate typos that give two or more instances of the same keyword, including undesired spaces and capitalization - preformed three times or more
    - Decide which version of keyword to use when grammatically different but representing the same concept (eg., rainfall and precipitation)
- Build a hierarchy to the set such that the main list can be narrowed to a maximum of ten terms, but still related to their children keyword and to the children of their children, etc. - performed by LUQ Principal Investigator three times
- Revise existing taxonomy adding, collapsing or deleting terms - three times by one of the LUQ IMC members; once by other 3 LUQ scientists
- Export the hierarchy of keywords into a specially structured Excel spreadsheet that can be imported into Drupal - as many times as there is a new version of the Taxonomy (done when the updates included addition of terms only; otherwise corrections were made in Drupal)
- Import taxonomy into Drupal (after installing Drupal module allowing this)

The keyword assignment for Data File Structures is achieved by including the related data files in the Data Sets content type. This is done by adding the Data File Structure as a Field in this node.

In regard to Research Sites, Research Projects, or Variables the effort is trivial since they actually represent a keyword to the data set they are assigned to.

The real challenge is presented by the set of variable labels. There are many examples where the same variable has different labels. For example the variable "year" can have be labeled as "YEAR", "year of measurement" and many others. One format should be selected and all data files with the same column or variables should be edited to show the same label as the one selected. This Drupal-based system makes it easier to create the views to spot variable redundancy and merging.

As for the sets of keywords related to the publications, ways should be developed to assign keywords to old references lacking them. The same thing happens with the People's keyword with some people with no profile in the LNO personnel database.

Next Steps

There are 3 sets of keywords (Data Sets, People, and publications) which need to be synchronized. The Data Set Drupal taxonomy will be the model for the other two sets. We expect all to have some keywords in common and a mutually exclusive subset with the other two.

The Research Projects and Research sites are being developed in such a way that all nodes for those content types are standardized. Once entered into the system, they are defined as fields in the Data Set content type. This type of information will be standard to all data sets.

In the Data Sets' nodes, the Research Projects are configured as what is known in Drupal as a "Node Referrer", simply meaning that as soon as the corresponding Research Project is entered in the system listing a specific data set as one of its "Related data sets", the fields selected of the Research Projects will show automatically in the "View" of the corresponding Data Set node. Node Referrer is a mechanism that can be used to implement many-to-many relationships in Drupal.

The Research site is configured as a "Node Reference" field in the Data sets nodes. This means that the selection of the sites related to that data set is made within the creation of the data set and the default View of the data set will display it when saved. Each data set is comprised of data gathered at one or more sites, thus, a data set is related to many research sites or locations (a one-to-many relation).

Standardizing the variables (units, attributes, names, date formats) will be a detailed manual process that will require the collaboration of PIs and information managers, that will serve as keywords to the Data Files Structures. If this standardization is not performed, then the functionality of displaying related data files when searching for specific files or data sets will not be as effective. Redundancy in the use of variables at the site level hinders effective integration across data sets and across sites.

All these steps are related to the synchronization and standardization of keywords and Taxonomy within the system only. There are other levels of synchronization that could and should be done in order to foster integration of information with other sources. For instance, LUQ scientists are developing similar Taxonomies in other non-LTER Projects that have many scientific themes and data in common with the LUQ LTER data. Having a common taxonomy not only will

make integration and comparison of data easier but will eventually simplify the job for the LUQ scientist community in the generation of documentation, and other data-related documents. The quality of this functionality will further benefit if, in addition, we synchronize our set of keywords with the Keyword Vocabulary developed by our LTER Network of IMs.

Furthermore, the LTER Network has developed a unit dictionary that we may want to incorporate into the system to streamline the process of documentation and prepare the Luquillo data for future integration using PASTA-driven mechanisms.

Closing Remarks

The complexity of the migration of the LUQ web site and IMS into Drupal is due more to the complexity of our LTER system than to Drupal itself. The complexity really lies in the sum of all the standards, best practices and guidelines that may well be a reflection of the complexity of the science we are trying to document. At this moment, I do not know of a better system to host the kind of system LUQ needs to complete its common information management framework. After all, Drupal is not only a content management system but a content framework as well.

References:

1. The Drupal overview - http://drupal.org/getting-started/before/overview.

2. Definition of an "information management common framework" - http://gorilla.ites.upr.edu/node/definition-common-management-framework

LUQ LTER Information Management Education Exchange, Communication & Collaboration with Scientists and other Community Members

edit

Eda Melendez-Colom (LUQ)

Editor's Note: Some links in this article may not function properly at the time of publication due to site technical difficulties

Reflections on the LTER information managers' roles and responsibilities.

In many occasions, the information manager of a site is the first to be contacted by other communities' members to request information from the site. This way we become the liaison between the rest of the community and the scientists. Our history shows that we have been ambassadors of the LTER program in other countries. Asia, South America, Africa, Indonesia, and Europe all provide very good examples of the kinds of activities many of us have been involved with.

Whoever placed the 'manager' denomination to our LTER professional title surely knew what we do. Information and all its related hardware we have to administer are the principal resources that we organize but are certainly not the only ones. A manager plans, organizes, directs or leads, and supervises. In the LTER community, we also design, develop, and implement databases and systems. We could keep adding to the list of roles of an LTER information manager.

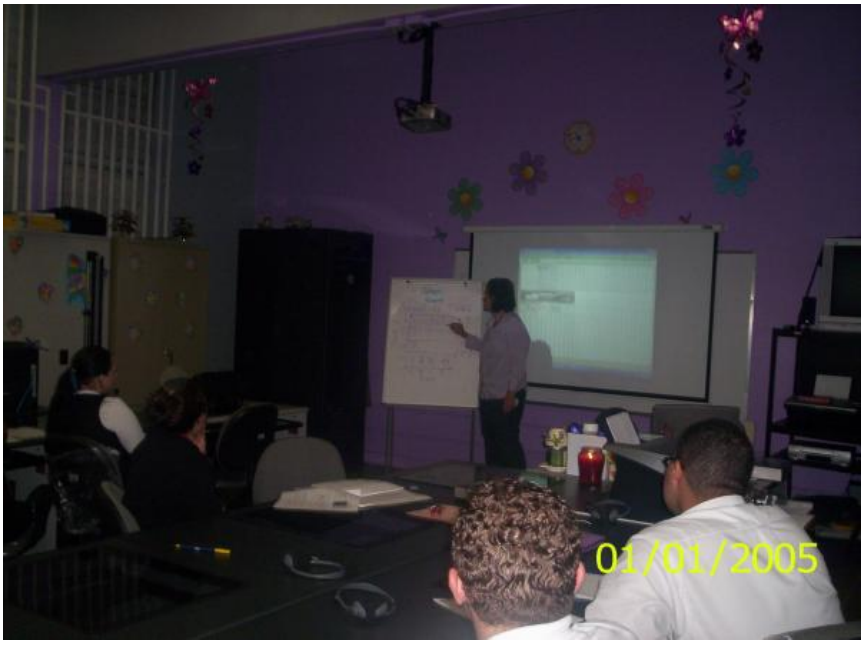
In this article, I am going to concentrate on the 'exchanging education' and 'enhancing communication and collaboration' additions to that list. Educating is definitively not only related to students but to a wider part of the LTER community and beyond. In relation to the communication role, we certainly act as translators between technical people and scientists, like we do with information technologists and ecologists, but we also assume teaching roles when we teach information management concepts and methods to our students and scientists.

These roles do not happen in one direction only, which makes our responsibilities even more complex. In many ways we become students and in many instances researchers. If we do not already have it, we develop a scientific mentality to be able to communicate with our scientific community, scientists and students as well. We need to be ready and open to learn about scientific methods and concepts as we learn algorithms, programs and systems that allow us to do a better and faster job in managing, documenting, analyzing and web-publishing all the information we handle.

Moreover, very often information managers coordinate and organize research activities for students, science teachers and other professionals in the community outside the LTERs.

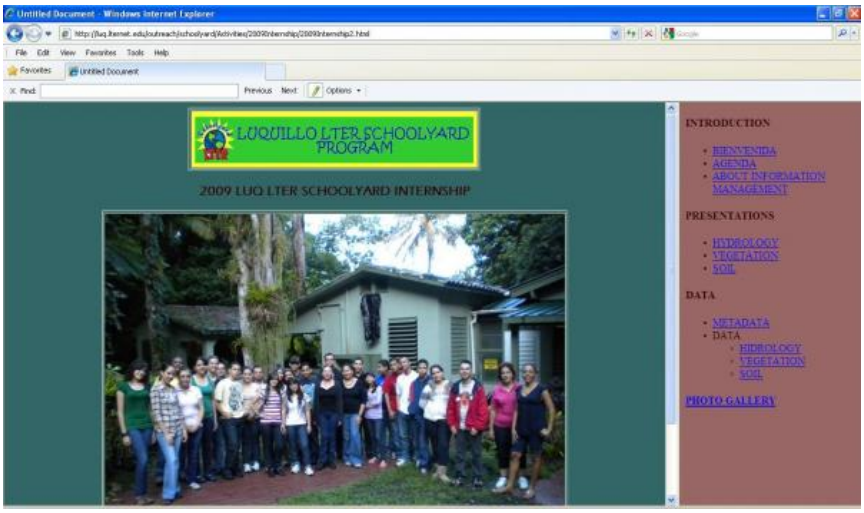
LUQ LTER IM role in Education.

The LUQ IM staff has been participating regularly in LUQ LTER Schoolyard's activities for more than a decade.



Information Management Workshop in Naranjito (See more photos)

The information manager has visited the participant schools to train students and teachers in the entering and managing of data and offers the IM workshop in every summer internship that have been held almost yearly since 2006 (http://luq.lternet.edu/outreach/schoolyard/Activities/2009Internship/200...).



2009 Schoolyard Summer Internship Web page

These activities always include a talk given by the information manager about the LTER program in general such that the teachers and students learn about LTER methods.

LUQ designed and currently maintains a web site for its Schoolyard program (http://luq.lternet.edu/outreach/schoolyard/index.html) and helps the investigators in charge to fill out our metadata standards forms to document the long-term data generated at the participating schools.

LUQ IM is assisting graduate students assigned to each school in completing metadata and enhancing the already developed data bases in each school with the idea of publishing their data in their web pages and preparing charts of their analyzed data for presentations given by students in local and national symposiums (http://crest-catec.hpcf.upr.edu/news/1er-simposio-luquillo-schoolyard-lt...). Collaborating scientists assist the schools in the analyses and publication of their data (Reference: http://luq.lternet.edu/outreach/schoolyard/Publication1999.html).

LUQ LTER IM Projects with other community's members. LUQ IM has participated in field data gathering, data manipulation, presentations of the LTER Information Management Methods, and professional talks to students of all grade ranges: from elementary (5 and 6 graders), Middle School (7 and 8 graders), High School (Juniors and Seniors), to undergraduate (REU students) to assist Schoolyard's teachers in introducing their students to the scientific research process.

LUQ IM activities with students have not been circumscribed within the LTER Schoolyard community. In several occasions, teachers from the local community, other than the LUQ LTER Schoolyard's, have requested special activities to introduce their students to the real world of research. Also, a presentation on Methods of the LUQ LTER Information Management was given to the staff of the LUQ LTER's institution, the Institute for Tropical Ecosystem Studies of the University of Puerto Rico.

We have documented all these activities on our web site, sometimes creating a sub-web site for a particular activity containing data, metadata, graphs, manuals, and photos. The following is a list of these activities and they are also presented on a LUQ IM web page: (http://luq.lternet.edu/datamng/index.html).

- IM and Remote Sensing activity at a local elementary school: (Activities and Photos, Metadata and Data) GPS of sites locations (an ITES' Remote sensing Project) (http://luq.lternet.edu/outreach/IMRemoteSensingandSchoolComm.htm)
- A presentation on data management and the LTER Program at a local DOD middle school (Buchanan). (http://luq.lternet.edu/outreach/IMandAMS.htm)
- Information Management Talk at ITES: "INFORMATION MANAGEMENT: a path less taken. A PRESENTATION TO THE ITES PERSONNEL"
- El Verde Field Station (EVFS), Rio Grande Puerto Rico Field Trip Activity for the Nashua High School South and Campbell HS Students of New Hampshire (http://luq.lternet.edu/outreach/SCHOOLS/NashuaMain.html)

LUQ LTER IM Role in Research.

One of the sectors of the LUQ LTER scientific community that LUQ IM interacts more with is the graduate students'. For years, the LUQ graduate students' representative has been collaborating with the LUQ information manager to maintain the LTER Network Office (LNO)'s personnel database. The selected student updates LUQ LTER graduate students' information in this database and reports these updates to the information manager. In 2009, the graduate students requested a special web site where they can share more specific information, graphs and photos about their research and any other additional information not included in the LNO's personnel database. A static version for this web site was published on the LUQ Web site (http://luq.lternet.edu/people/StudentdPers/index.html). The idea is to develop a site where each student can update and upload their documents. Students collaborate with LUQ IM in other ways. Before graduation, they complete the metadata for their theses dissertations which they file along with their data in the LUQ LTER Archives. Their data are published on the LUQ LTER Web site following the LUQ Data Management policy (http://luq.lternet.edu/datamng/imdocs/dmpolicy.htm). Some students also collaborate in gathering data for LUQ scientist and are also in charge to document these data following the LUQ LTER Metadata Standards. In addition, several students have collaborated with great photo galleries of the species they study and more. These photos are published in our web site:

- Diatom Population Dynamics (Spatial and temporal dynamics of freshwater diatoms in the Rio Mameyes, Puerto Rico) / (http://luq.lternet.edu/data/lterdb114/metadata/lterdb114.htm)
- Original Project: Fresh Water Shrimp Population Dynamics ; Additional: The LUQ LTER Animal Gallery / (http://luq.lternet.edu/images/LEFAnimals/LuquilloFauna.htm))

Communication with the Scientific Community.

The state of an Information Management System (IMS) in a site is a reflection of the communication status and/or professional relationship between the site's information manager and the site's scientists. A respectful and trustful relationship between the site's information manager and its principal investigator (PI) is essential such that the former is able to make proper decisions to do the job. The best situation is where this kind of relationship also exists with the rest of the site's scientists in addition to an effective communication between these two sectors of the LTER community.

LUQ LTER has experienced a progress in the communication between their scientists and the information manager along the almost 21 years of the information manager in the job. Collaboration in research had its peak when the information manager prepared a document on the relation between air and soil temperature and elevation along an elevation gradient in the Luquillo Experimental Forest (LEF). This paper (http://luq.lternet.edu/data/lterdb90/data/bistempdata/Bis-temp.htm) was a requirement of a graduate course taken by the information manager under the supervision of one of the LUQ scientists.

Real collaboration fostering communication between the two parties started in 2001 when a new web site (http://luq.lternet.edu) was designed to meet the recommendations of the 2000 Proposal reviewers. (An additional benefit of this probation time was that LUQ IM started to collaborate with the LNO staff to develop its IMS. This collaboration has continued until the present).

In 2001, a new Information Management committee (LUQ-IMC) was formed in LUQ including the site's PI, two LUQ LTER scientists and the information manager as chair. It was formed to assist the information manager in making decisions about the web site design and the generation of its content. In 2009 the LUQ GIS scientist was added as a member of this committee.

Since then, collaboration between LUQ scientists and the information manager has evolved to collaborating in a scientific-informatics project regarding the generation of a set of keywords that will help in the new Drupal LUQ IMS and web site (LUQ 'website-IMS'). Again, LUQ is involved in a project where collaboration exist between the LUQ scientists, LUQ information manager and LNO' staff. This time the LUQ IMC and other two members of the scientist community develop taxonomy for all the LUQ data sets online. Indirectly, all the LUQ scientific community participated in this project since the raw set of keywords where assigned by their data sets owners.

Closing reflections on LTER information manager's roles.

Our activities can have an impact in the way different sectors of the community communicate and collaborate with each other. In LUQ, these activities are included in the LUQ IM regular plans and reports as outreach activities. They are a result of continuous communication occurring between LUQ IM and its scientific community. It is the product of an awareness of the importance of IM from key LUQ's scientists that came from NSF directives and from LUQ information manager's long time efforts to educate the LUQ community on the critical and important role of IM in research and education.

In general, it is my opinion that the site information manager's team role and efforts to educate their community plays a crucial role in the importance the site gives to the site's IM.

News Bits

A Controlled Vocabulary for LTER Datasets

edit

John Porter (VCR)

Currently most keywords used to characterize datasets at most LTER sites are uncontrolled, meaning that they are selected entirely by the data creator. One of the challenges facing LTER and external researchers in discovering data from LTER sites is inconsistent application of keywords. A researcher interested in carbon dioxide measurements must search on both "Carbon Dioxide and "CO_2." Moreover, the existing set of keywords is highly diverse. For example, in a 2006 survey of EML documents in the LTER Data Catalog, over half (1,616 of 3,206) the keywords were used in only a single dataset, and only 104 (3%) of the keywords were used at five or more different LTER sites (Porter 2006).

To address this problem, in 2005 the LTER Information Management Committee established an ad hoc "Controlled Vocabulary Working Group" and charged it with studying the problem and proposing solutions. To that end the group compiled and analyzed keywords found in LTER datasets and documents, and identified external lexographic resources, such as controlled vocabularies, thesauri and ontologies, that might be applied to the problem (Porter, 2006). Initially the working group attempted to identify existing resources, such as the National Biological Information Infrastructure (NBII) Thesaurus, that LTER might be able to adopt wholesale. Unfortunately, using widely-used LTER keywords as a metric, none of the external resources proved to be suitable. Too many keywords commonly used in LTER datasets were absent from the existing lexographic resources. So, starting in 2008 the working group focused on developing a LTER-specific controlled vocabulary, ultimately identifying a list of ~600 keywords that were either used by two or more LTER sites, or were found in one of the external resources (NBII Thesaurus and Global Change Master Directory Keyword List), and conformed to the recommendations of the international standard for controlled vocabularies (NISO 2005). This draft list was then circulated to members of the LTER Information Management for suggested additions and deletions, which were then voted upon (Porter, 2009). The final list consists of 640 keywords (http://intranet.lternet.edu/im/files/im/LTER_Keywords_V0.9.xls).

The final list was presented to the Information Management Committee (IMC) at the 2009 All-Scientists' Meeting. The sense of the meeting the keyword list was sufficiently evolved to form the basis of an LTER Controlled Vocabulary, but that adoption of an official LTER controlled vocabulary was beyond the powers of the IMC, and that a system of procedures needed to be developed for managing LTER-specific lexographic resources.

Earlier this year the LTER Information Management Executive Committee requested guidance from the LTER Executive Board regarding:

- Should there be an official LTER dataset keyword list that sites would be encouraged to integrate into their datasets?
- Who should determine what the contents of a keyword list should be, and who should manage revisions to the list?
- What resources might be available for creating tools and databases that will help sites integrate the keywords into their datasets, and help data users discover relevant datasets?
- Are there additional steps that are needed to further improve the discoverability of LTER datasets so that they have the maximum value in promoting scientific research?

The general response was positive, and in early 2010 the LTER Executive Board committed to helping to locate some domain scientists to work with the Information Management Committee on future activities, and endorsed the use of the list by LTER sites. Duane Costa with the LTER Network Office has already been working on some tools and databases to support access to the list via web services.

Next steps for the process include:

1. **Getting the keywords integrated into existing and future LTER Metadata.** Some of this may be automated, because of the synonym ring created as the list was compiled that includes the forms of words actually found in LTER metadata. However, some additions will necessarily be manual. This process should be enabled through the use of tools that suggest possible words based on free-text searches of the metadata and through type-ahead drop down lists, similar to the one used on the LTER Metacat now.
2. **Creating taxonomies that provide browsable and searchable structures for use in LTER data catalogs.** More than one taxonomy (a polytaxonomy) will be needed. For example, we might have one taxonomy for ecosystems (e.g., forest, stream, and grassland) and another for ecological processes (e.g., productivity with net and gross productivity as sub-categories). Each of these taxonomies will include the keywords from the list, so that they can be linked to the datasets. Steps in the creation of the taxonomies include:
    a. Identifying the taxonomies to be created (e.g., ecosystems, processes and objects measured)
    b. Examine existing lexographic resources (NBII Thesaurus, GCMD) to see if there are existing structures there that we can adopt
    c. Develop the taxonomies, assuring that each of the keywords falls into at least one of the taxonomies and adding modifiers to the keywords to help prevent them from being ambiguous (e.g., "head" can have both hydrological [pressure] and anatomical uses).
3. **Develop software tools that will use the taxonomies for browsing and searching.**

These developments will require active participation by LTER Information Managers and ecological researchers to assure that the resulting products will well serve the ecological research community. Throughout the process and into the future, the keyword list and taxonomies will need to be revised and improved. However, before we can improve them, we need to create them!

References:

National Biological Information Infrastructure (NBII) Thesaurus. http://www.nbii.gov/portal/server.pt/community/biocomplexity_thesaurus/578

NISO. 2005. Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies. ANSI/NISO Z39.19. http://www.niso.org/kst/reports/standards?step=2&gid=&project_key=7cc9b5...

Olsen, L.M., G. Major, K. Shein, J. Scialdone, R. Vogel, S. Leicester, H. Weir, S. Ritz, T. Stevens, M. Meaux, C.Solomon, R. Bilodeau, M. Holland, T. Northcutt, R. A. Restrepo, 2007. NASA/Global Change Master Directory (GCMD) Earth Science Keywords. Version 6.0.0.0.0 http://gcmd.nasa.gov/Resources/valids/archives/keyword_list.html

Porter, J. H. 2006. Improving Data Queries through use of a Controlled Vocabulary. LTER Databits Spring 2006. http://intranet.lternet.edu/archives/documents/Newsletters/DataBits/06sp...

Porter, J.H. 2009. Developing a Controlled Vocabulary for LTER Data. LTER Databits Fall 2009. http://databits.lternet.edu/node/70

UAF Researchers Propose a Climate Data Center for the Arctic

edit

Jason Downing (BNZ)

Researchers from the International Arctic Research Center and the Arctic Region Supercomputing Center at the University of Alaska Fairbanks (UAF) hosted a meeting early this year to discuss the development of a Climate Data Center for the Arctic that could build upon and enhance the visibility of data archives without duplication of current activities.

The purpose was to gather input to formulate a development plan that would identify and address a range of needs by various existing and potential stakeholders, both at UAF and beyond. The proposed center would promote additional synergistic activities and alleviate barriers to data utilization. Communication among the science researchers and computing specialists is critical at this time to map the development of the necessary applications and services to provide customizable product generation that could be most efficiently utilized for collaboration efforts. Included in these discussions were representatives from two LTER sites (Bonanza Creek LTER and Arctic LTER) and organizers interested in co-operation with the LTER network and its local sites.

While this meeting was preliminary, there is substantial momentum and ongoing progress... So stay tuned.

An update on LTERMapS

edit

Adam Skibbe (KNZ), Jonathan Walsh (BAL), Theresa Valentine (HJA), Jamie Hollingsworth (BNZ), John Carpenter (GCE), John Vande Castle (LNO) and Marshall White (LNO)

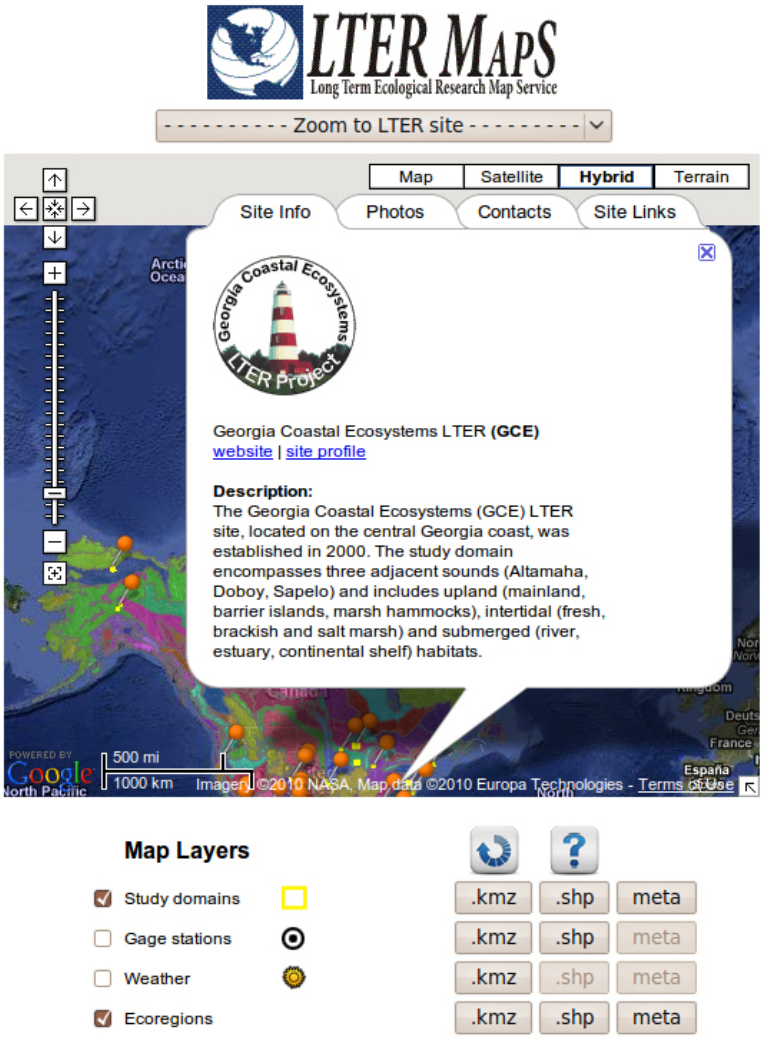
As the Information Managers GIS committee originally conceived it, LTERMapS (LTER Map Services) was envisioned as a means for creating a focused set of mapping tools to be made available across the LTER network. This has spawned a two-phase approach to accommodate both the needs of the general populace as well as those of researchers interested in a more robust system.

Phase one of LTERMapS provides a valuable information portal for the LTER community and can be accessed at http://www.lternet.edu/map/. This phase was built using the Google Maps API and offers a network-wide mapping solution targeted at a broad user group. The map dynamically harvests information from the LTER Network's SiteDb to reflect real time changes in the system. Information from SiteDb is displayed via a popup balloon for each site, with clickable links for various outside resources, viewable and downloadable KML and Shapefile geographic data (for site locations [view only], boundaries, ecoregions, gauging stations and weather) and a shadowbox popup photo utility.

The first phase of the project concluded this spring with a successful migration to the LNO during a funded post All Scientist Meeting Working Group meeting in Albuquerque. This migration centralizes the resources of the application at the LNO and should provide a better experience to the user in both speed and usability. An additional outcome of this meeting was the further development of requirements for Phase two. Two videoconferences were held with the IM community during the meeting to continue planning and receive feedback on the current product.

Phase two of the project is focused on site level data with an emphasis on creating a more robust set of tools, both cartographic and analytical, aimed at the scientific community. Phase two is in the testing phase, using three ESRI ArcGIS Server APIs: .NET, JavaScript and Flex.

The final platform will ultimately depend on performance tests, functionality, ease of development, and the user experience. This product is intended to employ a standardized set of data and tools for all LTER sites (DEM, infrastructure, hydrography, structures, and high resolution aerial photography), as well as be modifiable to fit each site's specific needs. In addition to development of analytical tools (as yet to be determined), phase two of LTERMapS will also allow for user submitted queries to harvest information and data and will integrate closely with the Network Information System (NIS) modules.



In the coming months we will be soliciting each site for further input (both content and data) to help us fine tune our specifications for phase two of LTERMapS. We encourage feedback on phase one and help guiding us in our efforts with phase two. Please look for future emails regarding LTERMapS and what you would like to see.

Information Manager Extraordinary Teleconferencing: An ET Moment

edit

Karen Baker (PAL, CCE)

The use of video-teleconferencing (VTC), a relatively new coordination mechanism called the 'Virtual Watercooler' by the LTER information managers, was reported on earlier by Linda Powell, Nicole Kaplan and James Williams (Databits Fall 2009). Use of VTC expanded last month with scheduling by IMExec of an 'Extraordinary VTC'. Organizationally, the VTC arrangements followed what is now a community tradition: the use of 'doodle' for sign-up and the conduct of two calls on separate days/times to accommodate the rather large group with a variety of schedules and geographic locations. The calls were chaired by IMExec co-chairs Don Henshaw and Margaret O'Brien.

The April Virtual Watercooler was extraordinary for two reasons. First, it was in addition to the now quasi-routine monthly calls. Second, the call was scheduled as a rapid response to notification that funds were available for site supplements. The supplement itself was remarkable in that it specifically delineated options to apply for funds targeting information management totaling up to $30,000. NSF program managers worked with IMExec to detail a range of topics and tasks that provided a descriptive account of information tasks. Such description is often left unexpressed or frequently lacks a vocabulary for describing the diverse data management requirements of a site that is a member of a scientific network. The VTC and collaborative planning provide an example of what it means to network. The usefulness of these funds was evident; a list of potential activities and/or specific projects appeared for discussion immediately. At short notice, every site had tentative plans in mind by the time of the VTCs.

The process that unfolded involved collaboration, task packaging, and timely dialogue. It resulted not only in individual site engagement but also in identification of joint interests early in the planning process. In terms of site engagement, the open forum discussions, the subsequent email exchanges, and wiki postings of text provided examples to participants who had heretofore not been part of site project formulation or preparation of proposals and budgets. The VTCs opened up discussions for those not aware of particular activities or approaches. Topics of shared interest arose naturally. Informality and inclusiveness for the watercooler was established by a 'round-robin' site reports format introduced and facilitated by meeting hosts. Each participant occupying one of the "LTER-Hollywood Squares" summarized tentative plans for the IM portion of their site supplement. For both VTCs this involved fourteen sites including several individuals participating by phone, from 'Virtual Squares' so to speak. The order of report giving was self-organized, guided by an initial suggestion to volunteer to give your

report when an activity at your site matched an activity reported by another site.

From left-to-right: C. Gries (NTL), D. Henshaw (AND), M. Gastil-Buhl (MCR), H. Garritt (PIE), I. SanGil (LNO), K. Baker (CCE/PAL - taking photo), J. Brunt (LNO), J. Porter (VCR), K. Ramsey (JRN), K. Vanderbilt (SEV), M. O'Brien (SBC), N. Kaplan (SGS), S. Remillard (AND), Y. (LNO)

Topics of shared interest that emerged related to the development of web-enabled metadata forms, unit registry, GIS, workflows, and meta-data enabled websites using Drupal as well as the development of new data models with an emphasis on making the development work transparent. Another item of interest to many site participants was the idea of support for travel for professional development opportunities as well as for participation in working groups in a variety of pertinent venues, i.e. site-site, site-network, and multi-site working group meetings.

The rapid response making use of the Virtual Watercooler served to highlight two aspects of the LTER IMC: the community ability to coordinate and its recognition of the benefits of a collaborative planning approach with the advent of new resources. The combination of a well-framed supplement call and the Extraordinary-Teleconferencing created a noteworthy ET moment that demonstrated how LTER Information Management community participants carry out the 'work' in network.

Good Reads

Webs of users and developers in the development process of a technical standard

edit

Don Henshaw (AND)

Review: *"Who are the users? Who are the developers? Webs of users and developers in the development process of a technical standard" by Florence Millerand and Karen Baker, published online in Information Systems Journal, Volume 20, Issue 2, pages 137-161, 24 July 2009, doi:10.1111/j.1365-2575.2009.00338.x*

The authors describe the inherent complexity in developing technical standards for a scientific community, and highlight the development of the Ecological Metadata Language (EML) for the ecological research community as an empirical case study. Preconceived notions of the roles of the informatics specialists, site information managers, and scientists were transformed as the development process moved forward. Specifically, "the roles of user and developer emerge as dynamic webs of users and developers within the development process of a standard". The authors dispel the simplistic view that 'system developers' and 'system users' are independent and homogeneous groups, and demonstrate how this boundary tends to blur when considering information system development. The role of the information manager is better represented through their emergent roles as both 'co-users' and 'co-developers'. Three models are presented describing the development process, and the final Integrative Design Model represents "an understanding of the roles – multiple and evolving – that the actors played in practice".

A few interesting nuggets paraphrased from the article:

The notion of local site 'implementation' of the standard is considered insufficient to describe the "redesign and redevelopment", "readjusting of pre-existing practices", or "adoption of new conventions" that actually occur. This implementation phase is termed 'enactment' as a process representing both "technical and organizational as well as social and political" dimensions.

The information managers "made visible and explicit the inherent difficulties in enacting the standard". Their role in the "elaboration of the standards development process", that is highlighted in this enactment stage, "holds potential benefit for other research communities" by contributing to "a better understanding and planning of such processes".

"Delays and unexpected challenges" in the enactment of the EML standard "may perhaps be better understood as symptoms of collaborative work being in early phases of development", misunderstandings of the new and multiple roles of actors, and the distribution of resources".

And finally, "an overarching insight of this research is the benefit of interdisciplinary research bridging information systems and social science perspectives in a research action framework."

This paper is timely in reminding the LTER and ecological community of the important lessons learned as we move forward with new standard development efforts.

Calendar

Events 2010 Summer & Fall

edit

Event: LTER Science Council, Executive Board and Communications Committee

Location: PIE-LTER, Peabody, MA, USA
Dates: May 11-14, 2010
Web: http://intranet2.lternet.edu/content/2010-science-council-plum-island-ecosystem-lter

Several site information managers are also attending the Science Council meeting, including all of IMExec. IMExec will also hold its regular monthly meeting in person on the evening of May 11.

Event: International Conference on Scientific and Statistical Database Management

Location: Heidelberg Germany
Dates: June 30-July 2, 2010
Web: http://www.ssdbm2010.org/

SSDBM provides a forum for original research contributions and practical system design, implementation and evaluation. Individual themes differ year to year with the main focus remaining on databases and application in the scientific and statistical fields. Recent themes included geospatial and sensor databases, bioinformatics (genomics, biodiversity informatics including biological databases), geological databases, data mining and analysis, metadata management, conceptual models, data integration, information visualization, scientific workflows, and system architectures.

Event: MULTICONF-1

Location: Orlando, Florida, USA
Dates: July 12-14, 2010
Web: http://www.promoteresearch.org/

The primary goal of MULTICONF is to promote research and developmental activities in computer science, information technology, control engineering, and related fields. Another goal is to promote the dissemination of research to a multidisciplinary audience and to facilitate communication among researchers, developers, practitioners in different fields.

Event: BIOCOMP'10

Location: Las Vegas, Nevada, USA
Dates: July 12-15, 2010
Web: http://www.world-academy-of-science.org/worldcomp10/ws/conferences/biocomp10

Event: East-Asia Pacific Large Forest Plot Workshop

Location: Malaysia
Dates: July 19-23, 2010

Scientists and IMs from Malaysia, Taiwan, Japan and the US will meet to develop tools to augment the system developed by the Center for Tropical Forest Science (CTFS) for managing data about the growth and survival of approximately 3.5 million trees and 7,500 species located on large forest plots around the world.

Event: Ecological Society of America Meeting

Location: Pittsburgh, PA, USA
Dates: August 1-6, 2010
Web: http://www.esa.org/pittsburgh

The Ecological Society of America will place global warming at center stage to draw a critical combination of scientists, policy makers, and concerned citizens to understand further its causes and consequences and to elucidate a clear scenario for addressing what is perhaps the most serious environmental threat facing the biosphere.

Event: Annual LTER IMC Meeting

Location: KBS-LTER, Hickory Corners, Michigan, USA
Dates: September 22-25, 2010
Web: http://intranet.lternet.edu/im/news/meetings/2010

The 3-year cycle dictates that this year's participants are only LTER personnel. Every LTER site is required to send a representative to the annual meeting. Two

days (Sept 23-24) are mandatory. A third optional day (Sept 25) has been added to advance product-oriented working groups. Optional dinner activities are planned on Sept 22.

Event: 2nd Annual Argonne National Lab Soils Workshop to discuss the next generation of ecologically meaningful soil metagenomics.

Location: Argonne, Illinois, USA
Dates: October 6-8, 2010
Web: http://www.mcs.anl.gov/events/workshops/soils/

Event: American Society for Information Science and Technology (ASIS&T)

Location: Pittsburg, Pennsylvania, USA
Dates: October 22-27, 2010
Web: http://www.asis.org/Conferences/AM10/am10cfp.html

Description: Since 1937, the American Society for Information Science and Technology (ASIS&T) has been the society for information professionals leading the search for new and better theories, techniques, and technologies to improve access to information. ASIS&T brings together diverse streams of knowledge, focusing what might be disparate approaches into novel solutions to common problems. ASIS&T bridges the gaps not only between disciplines but also between the research that drives and the practices that sustain new developments. ASIS&T members share a common interest in improving the ways society stores, retrieves, analyzes, manages, archives and disseminates information, coming together for mutual benefit.

Event: Participatory Design Conference (PDC)

Location: Sydney, Australia
Dates: November 29-December 3
Web: http://www.pdc2010.org/

Participatory Design (PD) is a diverse collection of principles and practices aimed at making technologies, tools, environments, businesses, and social institutions more responsive to human needs. A central tenet of PD is the direct involvement of people in the co-design of things and technologies they use. Participatory Design Conferences have been held every two years since 1990 and have formed an important venue for international discussion of the collaborative, social, and political dimensions of technology innovation and use.

Event: Digital Curation Conference (DCC)

Location: Chicago, Illinois, USA
Dates: December 6-8, 2010
Web: http://www.dcc.ac.uk/events/conferences/6th-international-digital-curation-conference

Scientists, researchers and scholars generate increasingly vast amounts of digital data, with further investment in digitization and purchase of digital content and information. The scientific record and the documentary heritage created in digital form are at risk from technology obsolescence, from the fragility of digital media, and from lack of the basics of good practice, such as adequate documentation for the data.

Event: Hawaii International Conference on System Sciences (HICSS)

Location: Kauai, Hawaii, USA
Dates: January 4-7, 2011
Web: http://www.hicss.hawaii.edu/

Since 1968 the Hawaii International Conference on System Sciences (HICSS) has become a forum for the substantive interchange of ideas in all areas of information systems and technology. The objective of the conference is to provide a unique environment in which researchers and practitioners in the information, computer and system sciences can frankly exchange and discuss their research ideas, techniques and applications. Registration is limited.