



01001100 01010100 01000101 01010010
LTER DataBits
Information Management Newsletter of
The Long Term Ecological Research Network
01001100 01010100 01000101 01010010

◆ **Feature Articles**

[About this Issue](#)

[Moving Toward Network Identity](#)

[Data Junction Integration Suite](#)

[Database Techniques for Creating Maintenance-free Web Pages](#)

[Special Report: Metadata Standards for Simulation Data / Models](#)

◆ **News Bits**

[2001 LTER Information Managers Meeting in Madison, Wisconsin](#)

[News from the Knowledge Network for Biocomplexity](#)

[International Information Management Outreach Activities](#)

[Metadiversity II Workshop](#)

[Detecting Environmental Change: Science and Society](#)

◆ **Good Reads**

[Ecology Through Time](#)

[Biodiversity Datadiversity](#)

◆ **Frequently Asked Questions**

[How can I check to see if my HTML and CSS code is valid?](#)

◆ **Calendar**

[Calendar Events](#)

DataBits: An electronic newsletter for Information Managers ----- Fall 2001 Issue (<http://intranet.lternet.edu/archives/documents/Newsletters/DataBits/01fall/>)

Featured in this issue:

Reports from information management workshops across the globe, a review of tools for importing structured text files into relational databases, and a comprehensive report on efforts to create a metadata standard for models and data sets from ecological simulation modeling efforts at Luquillo Experimental Forest.

DataBits continues as a semi-annual electronic publication of the Long Term Ecological Research Network. It is designed to provide a timely, online resource for research information managers and to incorporate rotating co-editorship. Availability is through web browsing as well as hardcopy output. LTER mail list IMplus will receive DataBits publication notification. Others may subscribe by sending email to majordomo@lternet.edu with two lines "subscribe databits" and "end" as the message body. To communicate suggestions, articles, and/or interest in co-editing, send email to databits-ed@lternet.edu.

----- Co-editors: Wade Sheldon (GCE), Linda Powell (FCE)

◆ Feature Articles

Moving Toward Network Identity

- *Karen Baker, Palmer Station (PAL) and James Brunt, LTER Network Office (NET)*

A research site's internet address includes its domain name and establishes a site's online identity. An appropriate web site address provides a measure of recognition and understanding as well as improved access when the path name is easily remembered. Two technologies used together create alternate identities for a web server - DNS aliases and virtual servers. If you have only one web server and manage it yourself then any number of DNS aliases can be added to that server. Virtual servers today can be name-based instead of IP-based so provide a method to create alternate web identities for a web server for sites that don't host their own web sites or that build their website as a sub-site of another organization. Although a Web site's success depends on a variety of elements including content, design, functionality and server power, issues of identity and memorability are equally important.

Currently, with the many Long-Term Ecological Research (LTER) sites, there are a variety of internet address forms reflecting each site's local context. For instance,

PAL: <http://www.icesb.ucsb.edu/lter>

SEV: <http://sevilleta.unm.edu>

SBC: <http://mistrallscf.ucsb.edu>

Since the LTER sites are partnered together into a research network supported by the LTER Network Office (lternet.edu) and early on established the tradition of identifying each site with a 3 letter code, there exists a natural naming schema for a set of virtual servers:

PAL: <http://pal.lternet.edu>

SEV: <http://sev.lternet.edu>

SBC: <http://sbc.lternet.edu>

These three are examples of sites that have already established their complementary identities. As a result, either of their above addresses brings up the site's web home page.

The LTER Network Office is ready to help each interested LTER site to establish their 'LTER network' identity. To establish name-based virtual servers where sites are managed as sub-site or by someone else, there are two sides to configure: the network server-side and the site host-side configuration. On the network server-side an entry is made to the DNS table with an alias for the virtual server in the form site.rootname.edu (i.e. pal.lternet.edu) pointing to the host-side server (guardian.icesb.ucsb.edu). Each alias (i.e. pal.lternet.edu) can have a virtual server entry associated with it in the host-side web server configuration file. Thus, site implementation is a straightforward procedure requiring an individual with system privileges to place a few lines of code in their site's web server configuration file.

When someone asks 'What's in a name?', the answer may be 'network identity'. When a site is part of several differing contextual domains, it need not be limited to a single web identity represented by a single domain name. The ability to establish virtual servers and hosts provides flexibility today in terms of network identity that can benefit loosely federated partnerships.

Data Junction Integration Suite

- *Ken Ramsey, Jornada Basin LTER (JRN)*

The Jornada Basin LTER (JRN) has been evaluating Data Junction (DJ) Integration Suite (DJ Suite) for parsing standardized, structured ASCII meta-data files and column-delimited ASCII data files to SQL 2000 database tables. JRN will be evaluating other uses for the DJ Suite such as dynamic EML generation.

DJ Suite consists of several components, of which JRN has only evaluated Content Extractor and Map Designer. Content Extractor parses text files (ASCII, HTML, XML, email, etc.) to an export format. The export format can be Content Extractor output format or a limited number of file formats such as Access, Lotus 1-2-3, Excel, Quattro Pro, dBase, and FoxPro. Map Designer allows fields to be mapped between numerous source and destination file formats to create or append to the chosen output format. Content Extractor's export format is one of the supported source file formats of Map Designer; another is XML. Another component of DJ Suite is DJ Engine, which provides dynamic data conversion. DJ Engine includes Java, COM, and C application programming interfaces (APIs) for integrating DJ Engine into custom applications. There are other components of DJ Suite besides DJ Engine that JRN has not yet evaluated, such as the software developers kits (SDKs) for DJ and DJ Engine. JRN will be evaluating DJ Engine very soon to see how well it can dynamically generate EML from SQL 2000 tables.

Potential uses of DJ at JRN:

1. Create, populate, and synchronize SQL 2000 tables from ASCII data and meta-data files.
2. Populate SQL 2000 tables from Adobe Acrobat 5 forms submitted via email.
3. Dynamically generate EML from SQL 2000 meta-data tables.
4. Allow dynamic data conversion for distributing data and meta-data files to data requestors in a file format other than SQL 2000 or ASCII.

So far, JRN has only evaluated using Content Extractor and Map Designer to complete use number 1 listed above. The approach I have taken is to search for and then concatenate all ASCII meta-data files using a Visual FoxPro program and then run the Content Extractor script that I designed to create a Content Extractor export format file. The export file is then used to populate the database table using Map Designer. JRN envisions using DJ Engine to automate this process to keep the database table synchronized with the master archival ASCII meta-data files, but we have not evaluated this yet. Another option to concatenating source ASCII files would be to run the Content Extractor script against each individual ASCII file.

DJ Suite appears to be a very useful suite of tools to support data migration, transformation, and integration. DJ Suite could potentially save hundreds of man-hours of custom programming and data entry for JRN. In the next few weeks, I will be evaluating Map Designer and DJ Engine for dynamically generating EML from SQL 2000 tables. I hope to finish my evaluations prior to the planned LTER EML Workshops.

Database Techniques for Creating Maintenance-free Web Pages

- *Wade Sheldon, Georgia Coastal Ecosystems LTER (GCE)*

It is widely recognized that web pages and web-based applications provide an effective means of displaying and updating information stored in relational databases. In fact, web database applications are used very effectively at the LTER Network office and many LTER sites. What may be less intuitive, though, is that the reverse is also true: databases can be used very effectively to update and control the content of web pages and navigation structures, when server-side scripting technologies such as ASP or JSP are employed. This article will describe how to combine several web database techniques -- stored virtual tables, dynamic hyperlinks, and dynamic navigation menus -- to ease web management and ensure that sites are automatically updated as information is added or changed in the database.

Most relational database management systems allow virtual tables to be created and saved as views, stored procedures or stored queries. Basing queries in web applications on these stored virtual tables offers many advantages over standard table queries. Most importantly, virtual tables centralize management of SQL code and provide a layer of abstraction, potentially insulating all dependent queries in web applications from changes in the underlying database structure. When changes need to be made to the schema of a database the corresponding views or queries can simply be edited and saved at the same time, automatically updating all queries that reference them on web pages and database applications. Virtual tables can also be used to shield web authors from complex SQL join syntax and permit more descriptive column names by aliasing, making it easier for people unfamiliar with database design to use tools such as Macromedia Dreamweaver UltraDev® or Microsoft FrontPage® to create database display pages. This technique often improves query speed as well, because database engines usually compile and cache views and stored procedures when they are first executed (unlike ad hoc queries, which must be parsed and evaluated each time they are run).

Like virtual tables, creating 'virtual hyperlinks' based on database queries is an effective way to insulate web applications from changes. Virtual hyperlinks are simply URLs that are generated dynamically by web scripts or database queries based on information stored in the database. If web content or directory structures need to be modified on the server, for example, updating corresponding file or path information stored in the database will automatically update all the query-based URLs in a single step. This approach is particularly effective if file and path information are contained in separate tables related by a path foreign key in the file table, thereby allowing changes to path names to be propagated to all corresponding files when the tables are joined (see example below).

The most obvious way to create virtual hyperlinks is to use a server-side script directive to embed directory and file names returned from a query inside an HTML anchor tag, e.g.

SQL query:

```
SELECT webpaths.pathname, webfiles.filename
FROM webfiles
INNER JOIN webpaths ON (webfiles.pathkey = webpaths.pathkey)
WHERE webpage = 'Site Description'
```

ASP/JSP script:

```
<A HREF="<%=pathname%><%=filename%>">Site Description</A>
```

Another approach is to use SQL string concatenation to create complete HTML hyperlinks as calculated fields in a view or stored query, e.g.

SQL view 'vwURL':

```
SELECT '<A HREF="' + webpaths.pathname + webfiles.filename + '"'>Site Description</A>' AS url, webpage
FROM webfiles
INNER JOIN webpaths ON (webfiles.pathkey = webpaths.pathkey)
```

SQL query:

```
SELECT url FROM vwURL WHERE webpage = 'Site Description'
```

ASP/JSP script:

```
<%=url%>
```

The latter technique is particularly effective for generating a column of links in an output table using one of the HTML editors mentioned above or simple custom scripts. In either case, changing content in one place - the database - updates the information globally.

In addition to 'virtualizing' information, databases can also be used to control web navigation, and therefore content, by including script code on web pages to build navigation menus based on database queries. Building HTML <SELECT> menus from queries is a fairly simple process, and can be achieved automatically using UltraDev or FrontPage. Adding a small client-side JavaScript function to the web page is then all that is needed to complete a database-derived navigation menu, e.g:

```
<SCRIPT language="JavaScript">
<!--
function lepto(form)
{
    var links = form.destination.selectedIndex;
    if (links > 0)
    { location = form.destination.options[links].value; }
}
//-->
</SCRIPT>
...
<FORM name="menu">
<SELECT size="1" name="destination" onchange="lepto(menu)">
<OPTION selected value="">Select a Page</OPTION>
```

```
<OPTION value="mypage.asp?site=mysite">My Site</OPTION>
<OPTION value="mypage.asp?site=othersite">Other Site</OPTION>
...
</SELECT>
</FORM>
```

As illustrated in the example, this technique is particularly effective for dynamically-generated web pages that accept input from HTML querystrings (e.g. `mypage.asp?site=mysite&map=1`). In fact, a single server-side script can be used to display an unlimited number of navigable web pages automatically when a database-generated web page incorporates a dynamic JavaScript jump menu containing the page URL with different querystrings (for an example of this approach, see the individual GCE sampling site pages at <http://gce-lter.marsci.uga.edu/lter/asp/studysites.htm>). In this scenario, simply adding or removing database entries in the corresponding menu table simultaneously updates both the website contents and navigation.

For added control over web navigation, a bit or Boolean field can be added to menu tables in the database and a corresponding restriction clause added to the view or query (e.g. `... WHERE displayonweb = 1 ...`). This field then functions as a toggle, determining whether records are displayed or not based on the field values for each record. This technique is very effective for preventing incomplete or outdated content from being displayed on the web site while preserving the information in the database.

While the software development and web server overhead incurred by scripted web pages with dynamic database access may not be justified for all web content, judicious use of these techniques can certainly help ease web site administration, freeing IM staff for more productive work.

Special Report: Metadata Standards for Simulation Data / Models

- *Eda C. Meléndez-Colom, Luquillo Experimental Forest (LUQ)*

Introduction

Efforts to develop metadata standards are not unique to the ecological community, even to the scientific community. An Internet search for the string "metadata standards" results in more than 1,000 entries. These includes web sites prepared by metadata developers from a diversity of disciplines like Web Browser development, curriculum developments, government agencies National Archives, statistics, and many others ([see Appendix](#)).

Some efforts like the Dublin Core Initiative aim to develop interdisciplinary metadata standards. The Dublin Core is an international initiative that "has resulted in consensus concerning a base set of elements for descriptive metadata." Others focus their effort in looking for the common elements shared by all groups others ([see Appendix](#)).

In spite of the efforts to produce metadata standards that apply to all disciplines, academic or not, specific discipline developers search for the best set of standards that will meet the needs of their community. Some groups even define the concept of metadata in their own terms: "Metadata is machine understandable information for the web" (W3C Metadata Activity).

It is clear that a common objective to all disciplines is to display the information that lies implicitly in the shared set of data, system or product such that the user properly uses it. Another common objective is to make it publicly accessible on the WWW to allow for transfer to other platforms and set of tools. The common set of objectives is expanded when you examine specific groups' related activities. For example, the Association of American Publishers (AAP) recognizes the importance of the publication of metadata simply because their customers ask them to supply metadata as a requirement for doing business with them.

Discipline-specific needs are met by introducing specific metadata elements to the standards.

In this article I intend to present, in general terms, the metadata elements typical of a model metadata standard and those elements common to any metadata standards. For this purpose I selected two model standards from the Internet. I will compare their objectives and design in order to give different perspective and methods to develop model standards, and will compare it to the Ecological Metadata Standards (EML) from the NCEAS presented at the last LTER Information Managers group meeting at Madison, Wisconsin (August, 2001) whenever possible.

The Model Metadata Standards

The community of modelers represents a wide variety of disciplines both from the private and academic sectors. There are two types of models: those that are computer encoded, and those that are mathematical equations. In terms of its components, a computer model has been defined as a process that runs on algorithms and differential equations, which receive data in a variety of forms, and gives some specific output. In terms of its functionality, a computer model is defined as a collection or compilation on what we know about a natural system. Mathematical models have been identified as representations of a large collection of scientific knowledge and experience about structures and functions of ecosystems. Ecology models use more on mathematical equations.

The efforts for creating model metadata standards are taken by the academic community (Crosier, 2001). Although the need for metadata has been recognized, the same issues responsible for the lack of good documentation of ecological data are also present within the community of modelers. In general, the identified causes are: the amount of work required, that scientists are neither punished nor rewarded to comply with documentation requirements, the lack of consistent documentation layouts, and the concern about standardization hampering the creativity and the freedom to formulate specific, individual and optimal adapted models, and that it slows down the introduction of innovative concepts.

Objectives in Developing Models Metadata Standards

I examined two sources of metadata standards development, one for computer models (Crosier, 2001, Metadata for Models, 2001) and another for ecological models (Benz, 2000, Gabele et al., 1999).

Basically, “the need to describe something to someone who does not know anything about it” (Crosier, 2001) is the main objective in creating metadata. In practice, developers confront the difficulties of developing metadata that meets the need of a large community with a variety of situations and resources. Thus the need for specification and portability become part of their objectives. Most importantly, the modelers need the scientists to understand what they have formulated in/with the model so they can decide if the model is applicable under certain conditions and even tell them what they have done wrong. In addition, they want to link documentation of mathematics with a documentation of the environment to minimize misuse of mathematical coding.

Creating metadata standards for models has its particular motivation. The need to catalog computer models in a situation where there are an increasing number of digital libraries, registries, and clearinghouses has been identified as the driving force to develop computer model metadata standards. The need to specify a set of more or less independent process-modules rather than large ones in order to facilitate model reuse and modularity is another objective in developing these standards.

The ECOBAS project (Benz, 2000, Gabele et al., 1999). further identifies the needs for standardized metadata file format for making accessible a WWW database to document mathematical formulations in modules, a WWW-based front end documentation program, and several converters and code generators that allow transfer of the model modules to other processing tools, simulation systems, and algebra packages.

Metadata Elements

Design and metadata elements

The structure or design characteristics of the metadata standards reflect the objectives that were considered in developing the standards.

The Alexandria prototype (Crosier, 2001, Metadata for Models, 2001) bases the core of its compound data elements in the definition of a computer system (Input, computational, and output processes), This results in a modularity that responds to the requirements of computer programming (an Input compound elements: data requirements, data extent and resolution, modeling construct description, and data set description; Data Processing conglomeration of metadata elements, and Output compound elements: Model output, output representation, and output modeling construction description).

ECOBAS, on the other hand, segregates the data elements in objects or processes that describe real-world objects and link all documentation of mathematical objects with a documentation of the environment. To distinguish between mathematical and ecological objects, it introduces three components: (1) the “type” component which contains the definitions of the mathematics and variable declarations, (2) the “specification” component which contains parameters values and ranges making the module specific for a given ecological process, and (3) the “domain” component which contains a declaration of the valid environmental context. It separates the model by its components and each component or module describes a system that can be treated independently from the whole system for scientific, historical or practical

reasons (e.g., the process photosynthesis in a forest ecosystem model).

Common metadata elements.

There are general metadata elements that are common to any data documentation standards. Table 1 summarizes these elements. Each data element maximum occurrence and its obligation or condition under which it must be entered is part of the design of the metadata standards studied herewith.

Table 1. Common metadata elements

Common category of element definition	Specific corresponding element definitions	Included elements
Identifiers	Model long name or ID, model short name or acronym	Usually compound or complex including model's title, date of creation, version number, citation, and identification number
Responsible parties	Model creator or author, metadata creator, contact person, Institution	Usually compound or complex including name, regular and e-mail address elements, phone numbers, etc.
Descriptors	Conceptual, keywords, temporal coverage, geographic coverage, cross reference to other datasets or models, additional information source	Abstract, list of keywords, start and ending dates of temporal range covered by the model, overall locations and sub-locations covered by the model, URL for related models and additional documentation
Access or availability	Constraints ¹ , availability, ordering procedures ¹ , cost ² , software requirement	Access or use constraints or conditions, Usually compound or complex including comment and contact availability, all the information regarding to the contact person in charge of distributing the model, hardware and software requirement, operating system, and expertise required ³
Metadata source ⁴	Metadata related information	Usually compound or complex including metadata source name and version, date of creation, information of person completing the metadata, modification date of metadata ²
Variable description		Usually compound or complex including name, acronym, definition, max and min values, range, units, and data type
Literature		Usually compound or complex including titles (article and book's), year, journal or publisher, pages, volume, issue, editor, ISBN, ISSN, and URL

¹common to the EML and Alexandria, ECOBAS does not include this elements or compounds

²only in Alexandria metadata standards

³ECOBAS provides only a free text metadata element for the description of this requirements

⁴Not part of EML and COCAS

Unique model metadata elements.

The principal concern of a modeler when sharing his/her model, does not differ with that of the scientific community in regard to their data: they want to make sure that the model is not misinterpreted or misused. Furthermore, the modelers want that

the user knows “the environmental conditions for which the model was developed or validated in order to make the modeling efforts reproducible and thereby scientifically valuable to others.” (Gabele et al., 1999).

The characteristics intrinsic to models, based on computers and mathematics coding determines the need of a special design for metadata standards. As mentioned before, the metadata elements from the two standards studied were clustered by their computer usage or by modules that responds to real-world systems. In addition to new and different metadata elements in model metadata standards, the type of aggregation of metadata elements of both standards are substantially different to the EML standards.

The EML standards modules clearly depicts the way we manage information. The modules reflect the different kind of information we deal with (eml-entity, eml-dataset, eml-literature,eml-attribute), the different types of resources (eml-resources, eml-physical, eml-software, eml-party) and even activities (eml-access, eml-distribution, eml-protocol, eml-research, eml-project).

The Alexandria standards reflect the importance given to computer processes (Input, data process, and output conglomerations of metadata elements) and a great importance given to the geographic coverage (contains 29 related metadata elements). Table 2 contains a list of metadata elements and/or compound elements that are unique (as compared to the other two discussed herewith) to the Alexandria’s standards.

Table 2. Alexandria’s unique metadata elements

Common category of element definition	Specific corresponding element definitions	Included elements
Additional Geographic coverage elements	Name of the planet, basis for the geospatial values, place or event that the model is about	Planetary body covered, Geodetic Reference System, place or event name, source, ID ,and URL
Bounding box	Generalized geographic coverage footprint for the model as a whole in the form of a bounding rectangle	Coordinates (W, E, S, N) of the coverage extent ¹ , source, ID, URL, description, accuracy estimate, and vertical dimension, vert. base level, vert. min, and vert. max
Detailed Geometry	Footprints(s) for the sub-locations or more detailed footprint(s) for the overall location.	Category, Number of points, point order, long. and lat. values, and name, ID, URL, source, accuracy estimate and vertical dimension of the gazetteer or other source that documents the source of the detailed geometry
Related Model	Description of the related model sufficient to locate it	Citation contact information, text, and URL relating to the model
Expertise Required	The level of expertise required to download and install the model	Description of the level to obtain (download and install), run the model once is installed, and interpret (understand the model and its result)
Input data Requirements, Modeling Construct	Variability and format of the input data required to run the model	Data file URL, Input model and data set descriptions, an external file containing this information (name classification, description, input source, dataset, type and units)

Data Processing Iterative cycles		Description of the use of the output of one run of the model as the input for a following run of this model
Model Output, Output Representation, and Output Modeling Construct	Data or visualization produced by the model	Output representation, post-processing requirements, an external file containing this information (name classification, description, output source, dataset, type, units, etc.)
Calibration Efforts and Validation	Efforts taken to justify the model as an accurate representation of real world events or situations	Confirmation data set, calibration efforts, model experiments and/or case studies, current use or application, level of uncertainty, known errors, strengths and weaknesses

¹ present in the eml-coverage module of EML

The ECOBAS standards show the importance given to ecological processes by separating the metadata elements in modules corresponding to real world processes, and designing an entering tool that forces the user to enter the information about the modules separately and then linked to the overall model. ECOBAS also emphasizes the importance of the environment for which the model is to be applied by containing metadata elements related to the type of soil, climate, ecosystem, and taxonomy. For the first three, it provides complete lists of possible types. Table 3 presents unique metadata elements for ECOBAS.

Table 3. ECOBAS' unique metadata elements

Common category of element definition	Specific corresponding element definitions	Included elements
Aggregate Modules (vector of sub-models) ¹	Hierarchy of modules that encapsulate other modules/objects and become an object	Vector of components (list of sub-models associated to a component vector element (e.g., component[1] = soillayer_1;), connections (assignments of each sub-component to an element of the component vector (e.g., component[10].var2 -> component[1].var1), and interface (assignment of global to local variable if module type is static)
Procedures, Functions and Equations	Equations and Functions return a single value, procedures return several	Equation contain real mathematical equations, Identifier, Collection of input variables declarations, collection of output variables definitions, and body of procedure; Identifier, list of inputs declarations, and definition of function
Linear algebra objects		Value, vector and matrices declaration
Control flow statements in the declaration, procedure, etc.		Loops, If-then-else, piecewise defined equations

Domain or ecological environment		Lists of appropriate soil classes, soil texture, climate class, and ecosystem type, taxonomical nomenclature ² , and free text description of the domain ²
----------------------------------	--	--

¹ Similar to the concept of package in EML

² present in the eml-coverage module of EML

Conclusion:

The metadata standards respond to the objectives of the standards, which in turn appears to be determined by the community that creates the standard. The Alexandria example is computer system oriented and appears to serve a community of spatial modelers. It forces the modeler to enter the documentation in terms of the Input, Output, and Processing activities undergone in the use of the model. On the other hand, the ECOBAS effort is ecological-system oriented and forces the modeler to separate the model in sub-components or modules which represent real-world systems. When searching for the online models (<http://eco.wiz.uni-kassel.de/ecobas.html>) one can search by subject or name of the model. Both standards have interfaces that export to XML. ECOBAS provides an online downloadable package, that provides the user with a model entry tool that debugs the entry and does not allow the user to save an incorrect coding. The XML coding is accessible at all times, and editing can be done while entering more metadata elements.

The EML contains many of the metadata elements present in these two standards. In this article the intention was to give an overview of the metadata elements characteristic of model data. Some comments were done in regard to EML, model metadata elements should be selected and a more exhaustive mapping should onto the EML standard should be done in order to incorporate a model module into the EML.

References:

Benz, J. 2000. Ecobas-Model Interchange-Format (ECOBAS_MIF 3.0) Reference manual. Department of Grassland Ecology and Forage Protection, University of Kassel in Witenhausen, Germany. (http://www.wiz.uni-kassel.de/ecobas/article_ecobas/node4.html)

Crosier, S.J. 2001. An Introduction to Metadata for Computer Models. Geography Department of the University of California in Santa Barbara. (<http://www.geog.ucsb.edu/~scott/metadata/intro/index.html>)

Metadata for Models Workgroup of the Alexandria Digital Earth Prototype Project. July, 2001. Content Standards for Computational Models Version 1.1. University of California, Santa Barbara. (Accessible from <http://www.geog.ucsb.edu/~scott/metadata/standard/index.html>)

Gabele, T., J. Benz, and R. Hoch. 1999. Standardization of model documentation, Part II: Usage of ECOBAS model documentation system - a short introductory manual. ECOMOD, the Newsletter of the International Society for Ecological Modeling. Perspectives. Department of Grassland Ecology and Forage Protection, University of Kassel in Witenhausen, Germany. (<http://eco.wiz.uni-kassel.de/ecobas.html>)

Benz, J. and R. Hoch. 1999. Modeling and documentation - Description of the ECOBAS system: Pros and cons of standardization. Department of Grassland Ecology and Forage Protection, University of Kassel in Witenhausen, Germany. (<http://eco.wiz.uni-kassel.de/ecobas.html>)

Appendix -- Some Web sites of Metadata Developers:

Dublin Core Metadata Initiative - <http://dublincore.org/>

XML & Metadata Developer Central Netscape - <http://developer.netscape.com/tech/metadata>

United Kingdom National Curriculum online - <http://www.nc.uk.net/metadata>

Record-keeping Metadata Standard for Commonwealth Agencies - <http://www.naa.gov.au/recordkeeping/control/rkms/>

[contents.html](#)

Access to official and other statistical data - <http://www.faster-data.org>

W3C Metadata and Resource Description - <http://www.w3.org/Metadata/>

The Association of American Publishers (AAP) Open Ebook Publishing Standards Initiative- <http://www.doi.org/ebooks.html>

For more comprehensive lists visit:

An Introduction to the World of Metadata and Models, Metadata and Cataloging: Methods & Ideas <http://www.geog.ucsb.edu/~scott/metadata/intro/index.html#methods>

Visit Mapping between metadata formats - <http://www.ukoln.ac.uk/metadata/interoperability/>

◆ News Bits

2001 LTER Information Managers Meeting in Madison, Wisconsin

- ***Barbara Benson, North Temperate Lakes (NTL)***

Representatives from all 24 LTER sites attended and actively participated in the 2001 LTER Information Managers Meeting in Madison, WI. Other attendees included James Brunt, David Blankman, Owen Eddins, Troy Maddux, Bill Michener, and John VandeCastle from the Network Office, Judy Cushing and Erik Ordway from the Forest Canopy project, Matt Jones from NCEAS, Dick Olson from the Oakridge National Laboratory, Jens Schumacher from the German Biodiversity project at the University of Jena, and Phyllis Adams from the National Park Service.

The major themes of this meeting were:

1. Network Information System
2. Metadata standards and implementation
3. Information management support for cross-site synthesis
4. Mentoring of site information managers
5. Outreach
6. IMExec reorganization

Status reports were presented on the LTER Network Information System (NIS) including the following components: the LTER Intranet page, personnel database, SiteDB, Data Table of Contents (DTC), All Site Bibliography, ClimDB/HydroDB. The NET staff plans to have the personnel database, DTC, and All Site Bibliography updated for all sites this fall. Sites were asked for their cooperation in these efforts. Don Henshaw (AND) is undertaking the further development of the intersite climate database (ClimDB) and a parallel database for hydrologic data (HydroDB).

Matt Jones (NCEAS) presented an overview of the Knowledge Network for Biocomplexity (KNB) project. KNB is a collaboration among NCEAS, LTER, and SDSC to develop the technological infrastructure to promote data accessibility, synthesis and analysis, and data preservation. Three products under development were highlighted: Metacat (a metadata catalog that supports storage, search, and presentation), Ecological Metadata Language (EML; an XML schema for ecological metadata), and Morpho (a software tool to create and manage data and metadata).

The participants spent a significant portion of the meeting discussing metadata standards and implementation. Matt Jones (NCEAS) and Peter McCartney (CAP, metadata working group leader) reported on the status of the Ecological Metadata Language (EML) 2.0 which is nearing completion and is currently available in a beta version. The two main tasks for LTER sites are

1. The restructuring of existing metadata content to EML (A site's metadata do not need to be managed in EML format but must be easily translatable into EML.)
2. The development of new metadata content to be compliant

We identified 3 metadata working groups based on the form of existing metadata: non-parsable, parsable, or stored in a relational database. These groups are charged with documenting the different kinds of needs across sites, the tools needed, and the costs for each metadata type (both for restructuring content to EML and developing new content) of producing a body of metadata.

Recognizing the need for information management support for cross-site synthetic research, we discussed ways to coordinate with the principal investigators in these efforts. During June 14-16, 2001, investigators interested in cross-site synthesis for net primary productivity met with some information managers to review what leads to successful cross-site synthesis (Advancing the Sharing and Synthesis of Ecological Data: Guidelines for Data Sharing and Integration workshop, Benson (NTL) et al.). A draft of a guidelines document is in preparation along with a manuscript ("Synthesis in Ecology: Approaches, Principles, and Procedures") based on case studies from within LTER and the broader ecological community. We were briefed on two upcoming intersite research projects related to species invasions and biogeochemistry. Concern was expressed that information managers be included in the planning of these projects and that these information managers be liaisons for the IM committee.

We established a new Site Mentoring and Training subcommittee (chairs are Susan Stafford (SGS) and John Anderson (JRN)) after discussion highlighted the needs in this area. In addition to the 3 new LTER sites that joined the network recently, several of the older sites have had IM personnel changes in 2001. In some cases, incoming IMs did not have the benefit of any "on the job" training because the previous IMs had already left the position at the time of their hire. There was a general consensus that there is inadequate guidance or training available to new IMs. Some suggested improvements included 1) the NET Office taking more responsibility in centralizing information such as a list of tools (databases, software site licenses, site surveys, and white papers) available to the IMs, 2) an "Orientation" for new IMs (e.g., each new manager would visit the NET Office where they would be given an orientation packet of information that would include a checklist of tasks; each new manager would also be encouraged to visit another LTER site).

Several LTER IMs (Peter McCartney (CAP), John Porter (VCR), Kristin Vanderbilt (SEV)) have been involved in international training workshops for information management. During the past year workshops were held in Hungary (with participants from Hungary, Poland, Romania, Slovakia, and the Czech Republic), Mongolia (with participants from Taiwan and Mongolia), and at the Seville field station for Israeli and Palestinian participants.

Kristen Vanderbilt (SEV), Emery Boose (HFR), and Don Henshaw (AND) were elected to serve on the IMExec (along with existing members Barbara Benson (NTL), Peter McCartney (CAP), John Anderson (JRN), and Susan Stafford (SGS, chair) and James Brunt (NET, ex-officio). Karen Baker (PAL) and Ned Gardiner (CWT) rotated off IMExec. Karen Baker (PAL) led a discussion of various critical tasks to be performed by IMExec and the assignment of responsibility for these tasks to individual IMExec members.

Presentation files, meeting agendas, LTER Site Bytes, notes and other information pertaining to the 2001 LTER IM Meeting are available at: <http://www.vcrlter.virginia.edu/nis/im2001/>.

News from the Knowledge Network for Biocomplexity

- Matthew Jones, National Center for Ecological Analysis and Synthesis (NCEAS)

This year's information manager's meeting in Madison saw a lot of discussion of the products and promise of the Knowledge Network for Biocomplexity (KNB). The KNB is designed to facilitate and promote collaboration and synthesis in ecology by building a national infrastructure to enhance data sharing and data preservation. It is a collaborative project among researchers at the National Center for Ecological Analysis and Synthesis (NCEAS), the Long Term Ecological Research Network (LTER), San Diego Supercomputer Center, and Texas Tech University. This article is a brief update on the status of the technology development efforts at KNB and is targeted at information managers and informatics researchers.

Ecoinformatics development workshop

In October we will be hosting a workshop at NCEAS for ecoinformatics developers to get together and explore ways

to collaborate on software development. This meeting developed out of the realization that there is tremendous overlap in the efforts occurring at NCEAS, LTER Net, ASU, and the Forest Canopy project. Thus, developers from these groups will meet for two days to discuss areas of common interest and overlap, and find mechanisms to increase the efficiency and effectiveness of the projects through collaboration. The project is jointly sponsored by the KNB and the "Networking our Research Legacy" project at Arizona State University.

Software development

The KNB is developing a number of software products for data management. Each of the software products discussed below can be downloaded from the KNB project web site (<http://knb.ecoinformatics.org>).

The Ecological Metadata Language (EML) is an XML-based syntax for encoding documentation about ecological data in a highly structured format. We are continuing our revisions of EML, and have fixed a number of minor bugs and problems that have been pointed out, some by LTER information managers. The major outstanding issue that is left to be resolved is how to combine the various EML modules together when distributing metadata. This "packaging" issue will be discussed in detail at the Ecoinformatics workshop mentioned above. Our target is to have a candidate release of EML 2.0 by December 1, 2001.

Metacat is our java-based metadata catalog. It supports storage and search of any XML-encoded metadata without prior knowledge of the metadata schema. This makes it robust in the heterogeneous world of ecological data management. The major improvements to Metacat include the addition of namespace support, and support for easily configured alternative user interface presentations. We call this feature "style-sets", and it has allowed us to develop the UC Natural Reserve System data registry and the Organization of Biological Field Stations Data registry without changing any of the Java code associated with Metacat. These new data registries will be demonstrated at the OBFS annual meeting this year, and will be available on the internet after that point.

Morpho is our Java-based client application for ecological data management. We have been working on a number of new features for Morpho, including a reverse-engineering module that parses a data file and generates metadata about the data file automatically, as well as a data browsing feature. Our release plans for Morpho are to finish the current targeted set of features by November 1, 2001 (Beta 2 release), and then spend a solid month writing documentation and fixing any reported bugs, for a final release of Version 1 around December 1, 2001.

One of the features of Morpho is taxonomic searching. To support this, we have developed and released "Itislib", a Java-based library that can be used to query the ITIS Canada (ITIS*ca) XML interface. The Itislib library is a developers SDK that allows you to query ITIS for a taxonomic phrase, parse the list of species that match that phrase, and for any of those species download and process all of the information about those species, including their place in the taxonomic hierarchy, their synonyms, and any children taxa. For example, you can query for "Psychotria" and find out all of the parent taxa, child taxa, and synonyms of "Psychotria nervosa". We use Itislib in Morpho, but decided to release it separately from Morpho because of its broad applicability in ecoinformatics development.

Finally, we've started the implementation and further development of some of our tools to support advanced services. We've been working on generalization and extension of our metadata-based quality assurance processor, and hope to have an initial prototype available early next year. We've been researching tools for ontology building so that we can begin the process of building an ontology for the biodiversity research domain to support our data integration engine. In addition, developers at SDSC have made substantial progress on our Hypothesis Modeling Engine. You can get more details about these projects from the KNB web site.

Participation

I'd like to remind everyone that the KNB project welcomes collaboration. If you are interested in contributing to the types of informatics development work that I have described here, please contact me (jones@nceas.ucsb.edu). We're also looking for feedback, so please take a look at our software, and provide us with any comments, issues, problems, or critiques that you think will be useful to the project. These comments can be made in our bug tracking system (<http://bugzilla.ecoinformatics.org>) or by sending email to knb-software@nceas.ucsb.edu.

International Information Management Outreach Activities

- *Kristin Vanderbilt, Sevilleta LTER (SEV)*

EAPR-ILTER Information Management Workshop

In July 2001, Peter McCartney (CAP) and myself traveled to Ulaanbaatar, Mongolia to teach a three-day Information Management Workshop for scientists in the East Asian and Pacific Region International Long-Term Ecological Research Network (EAPR-ILTER). Travel and materials funds for the workshop were provided by a grant from NSF to Tony Fountain or the San Diego Supercomputer Center (SDSC). The workshop was conducted immediately following the regional ILTER meeting that was held at Lake Hovsgol, Mongolia to encourage scientists attending the meeting to stay for the workshop. Workshop participants were from Mongolia, Taiwan, Korea and Thailand. The Mongolian Technical University provided a computer lab with Internet access for the workshop. Organizational support was supplied by Dr. Tsogtbaatar and Dr. Amarsaikhan of the Mongolian Academy of Sciences.

The objective of the workshop was to give participants some training in ecological information management theory and techniques. Lecture and laboratory topics included Basic Concepts of Information Management, the Relationship Between Data and Research, Database Design and Modeling, SQL and Data Query, Ecological Metadata, Connecting Databases to the WWW, HTML and Web Page Design, XML, Data Archives, and Quality Control and Quality Assurance. Tony Fountain also gave a lecture on data mining. Students practiced designing and querying databases using Microsoft Access, and the more advanced students successfully used ASP scripts to query their databases from the web.

Several positive things have occurred as a result of the workshop. Shortly after the workshop, the Mongolian participants met with Dr. Galbaatar, the General Secretary of the Mongolian Academy of Sciences (MAS), and formed the Organizing Group for Databases in the MAS. The Mongolians have also resolved to begin hosting the Mongolian LTER web site within Mongolia, rather than have it hosted by the Taiwanese as it has been. As a follow-up to the workshop, Tony Fountain plans to host a Mongolian scientist from the workshop at the SDSC for a year to study data mining.

Information Management Workshop for Scientists from the Middle East

The Cooperative Monitoring Center (CMC) at Sandia National Laboratories funded LTER personnel through a grant to UNM to lead a seven-day workshop (September 9-15, 2001) for scientists from the Middle East. Participating scientists are all involved in the "The Regional Initiative for Drylands Management," a program designed to build regional cooperation among the governments of Egypt, Israel, Jordan, Tunisia, and the Palestinian Authority to address both the environmental and sustainable production issues related to desertification.

The workshop was held in Bonn, Germany at the Gustav-Stresemann-Institute, home to microscopic guestrooms and deliciously monotonous cafeteria food. David Blankman, database administrator for the LTER Network Office, grappled tenaciously with the many obstacles surrounding transporting computers into Germany and getting them connected to the Internet for the workshop. We were surprised to learn that a \$2000 deposit was required to get German customs to release the ten laptops, so the computers languished at the Frankfurt airport for the first two days of the workshop until the cash was obtained.

The workshop included a healthy dose of lectures and demonstrations on Remote Sensing and GIS concepts and software, led by John Vande Castle (NET). Other lectures and laboratory exercises were expanded versions of the topics covered in the Mongolia workshop. Students created web pages first using HTML and then more elaborate pages using Dreamweaver. They designed and queried databases using Microsoft Access, and David Blankman demonstrated how to use Dreamweaver UltraDev to connect databases to the web. Bill Michener (NET) lectured on metadata and demonstrated MORPHO. He also lectured and led exercises on data archives and data synthesis. John Porter (VCR) was to have taught database modeling and SQL during the last few days of the workshop, but the events of September 11 precluded his trip to Germany. David and I carried on using the presentations and advice that he emailed us.

It was a tough week for us to be out of the United States, but evaluation of the workshop was positive. Our ten students reported that they had learned a great deal and would be able to implement much of what they had been taught. Their goal is to use technology to build bridges between scientists in the region studying the common issue of desertification, and they felt that this workshop brought them a step closer to cooperation through data sharing.

Metadiversity II Workshop

- John Porter, Virginia Coast Reserve LTER (VCR)

The "Metadiversity II" workshop was held in Charleston SC, June 25-26, 2001. The workshop brought together members of the library, computer science, museum specialists, taxonomists and ecoinformatics communities for presentations and discussions aimed at identifying "gaps" in the information bases provided to researchers and other users of biodiversity information. The conference followed up on a 1998 workshop held at Natural Bridge, VA. The meeting was sponsored by NFAIS (National Federation of Abstracting and Information Services), an organization associated with the library community. A WWW site for the conference is at: <http://www.pa.utulsa.edu/nfais.html>.

Presentations were diverse, ranging from studies of how scientists access and use information, to ecological information systems to publishers' perspectives on data and electronic journal publication. Organized working groups that combined individuals representing each of the disciplines and stakeholders developed recommendations regarding identifying and filling information gaps and improving communications between the different groups.

ILTER representation at the meeting included James Brunt (NET), who provided a presentation on the LTER Network Information System and the "Knowledge Network for Biocomplexity," and John Porter (VCR) who served as a peer-review panelist.

The keynote presentation was by Librarians Carol Tenopir and Donald King on "What do we know about scientists' use of information?" which provided some findings from their new book "Towards Electronic Journals: Realities for Scientists, Librarians and Publishers" that focused on understanding how scientists access and use information. One of the more interesting findings is that although academic scientists account for the writing of 70% of peer-reviewed articles, most of the reading of the articles was by a (larger number) of non-academic researchers. Reading by scientists has gone from about 100 articles per year to 130 per year, but this differs widely among disciplines with medical researchers reading an average of 322 articles vs. only 72 for engineers (there were no specific figures for life sciences or ecology in specific).

Presentations of special interest to LTER Information Managers included:

- John Pickering - www.discoverlife.org - using WWW databases to support citizen and student participation in creating databases of species distributions. Online, graphical taxonomic keys are linked with a system for uploading images of the organisms in question. Although this system is currently aimed primarily at the flora and fauna of the Smokies, new keys with a nation-wide perspective are in the process of being created. These systems might be of special interest to Schoolyard LTER efforts.
- Robert Collwell - the BIOTA database program for specimen data.
- Jorge Soberon - CONABIO, an impressive system for biodiversity information for Mexico that integrates information from taxonomic collections to help meet the information needs of decision makers, managers and scientists. WWW site (mostly in Spanish) <http://www.conabio.gob.mx>.
- Janet Gorman - the Integrated Taxonomic Information System (www.itis.usda.gov), a source for standardized taxonomic names that include animals, plants, fungi, protists, and bacteria.

Other talks focused on some of the issues regarding copyrights/ownership of data and trends in the publishing industry, which is wrestling with the issues surrounding electronic publications.

A formal report on the meeting is pending and, when available, will be posted along with the Metadiversity I report, on the <http://www.nfais.org> WWW site.

Detecting Environmental Change: Science and Society

- *John Porter, Virginia Coast Reserve LTER (VCR)*

The Detecting Environmental Change: Science and Society conference was held in London on July 17-20, 2001, in conjunction with an International LTER (ILTER) meeting. A WWW site for the conference is <http://www.nmw.ac.uk/change2001>.

The conference featured a large number of presentations aimed at understanding the methodological and information needs for detecting environmental change. It had special emphases on Environmental Data, Indicators of Change, Using Models, Global Networks, Data on Past Environmental Conditions, Early Warning Systems, Involving the Public in Monitoring and Better Communication.

ILTER activities at the conference included the annual ILTER meeting (held prior to the other sessions), posters, presentations and a special electronic display/demonstration of the LTER Network Information System. Participants from the U.

S. LTER included Bob Waide, Patty Sprott and John Vande Castle (NET) and John Porter (VCR).

The Environmental Data section of the conference provided presentations that focused on: climate change, atmospheric pollution, biodiversity conservation, land management, land use change, marine ecosystems, pest management, catchment management, urban management, water quality and fisheries management. Recommendations dealt with the need to have more data from human impacted landscapes, issues regarding the integration of data and the incorporation of social scientists into research, and the needs for additional baseline information.

Key issues and associated abstracts are available at: http://www.nmw.ac.uk/change2001/dec2001_keymessages.htm

Abstracts of the talks are available at: http://www.nmw.ac.uk/change2001/Abstracts/dec_abstracts.pdf.

◆ Good Reads

Good Read: Ecology Through Time

- *Karen Baker, Palmer Station (PAL)*

Kaiser, J, 2001. Ecology Through Time: An Experiment for all Seasons. Science 293(5530):624-627.

In providing an overview of concepts fundamental to the Long-Term Ecological Research network, one can place the recent issue of Science (July 2001) on the reference shelf right next to the Jul/Aug 1990 issue of BioScience which contains that original introductory trio of articles (Swanson and Sparks: Long-Term ecological research and the invisible place; Magnuson: Long-term ecological research and the invisible present; Franklin, Bledsoe, Callahan: Contributions of the long-term ecological research program). This article begins "the NSF's LTER network has proved a smashing success", continues with illustration of the ramifications of network science, and summarizes with site-specific scenarios as well as cross-site tables. Aspects of the spirit of the LTER's practice of science are captured along with examples from the broader community, rounding out this Science issue dedicated to ecosystem science.

Good Read: Biodiversity Datadiversity

- *Karen Baker, Palmer Station (PAL)*

Bowker, GC, 2000. Biodiversity Datadiversity. Social Studies of Science 30(5): 643-683.

One might consider this article by Bowker an ecological continuation of a conversational thread developed by Brooks in his book *The Mythical Man Month* (1975). Acknowledging the difficulties arising from the multiplicative factor inherent in team project communications, Bowker strides on into the territory of well-counted megafauna and under-valued microscopic entities to discuss the diversity in databases and transitions in frameworks. This work is a call for social scientists to consider potential contributions given their experience with naming and classifying, context and integration, organizational practice and scientific history. Attention to infrastructure layering impacts first the movement from data discovery to data management and ultimately the communication of knowledge through semantic synthesis. The challenges are confronted as Bowker articulates the nontrivial nature of the informatics task and summarizes: "The information collection effort that is being mounted worldwide is indeed heroic."

◆ Frequently Asked Questions

How can I check to see if my HTML and CSS code is valid?

- *Wade Sheldon, Georgia Coastal Ecosystems LTER (GCE)*

As the WWW has matured and become a major outlet for scientific information exchange, information managers have been increasingly expected to serve as web masters and web authors for their LTER sites. During the same time the diversity of WWW standards, HTML editing tools, and web browser versions in use has increased exponentially. New and even veteran web authors need help deciphering these standards and ensuring the code they write is valid and will be interpreted correctly by end users' software.

Fortunately, there are many online resources that can help. Tools available at AnyBrowser.Com (<http://www.anybrowser.com>), for example, can validate HTML and CSS code against various W3C standards, show you how your pages would be displayed in various graphical and non-graphical web browsers, and assist you in writing appropriate meta tags to improve search engine listings for your pages.

Other useful tools in this category are:

- W3C HTML Validation Service (<http://validator.w3.org/>)
- W3C CSS Validation Service (<http://jigsaw.w3.org/css-validator/>)
- Netscape's Web Site Garage (<http://websitegarage.netscape.com/O=wsg/start.html>)

◆ Calendar

Feb 9, 2002: Distributed Collective Practices, a Distributed Knowledge Research Collaborative (DKRC) conference, UCSD, <http://www.limsi.fr/WkG/PCD2000/indexeng.html>

April 18-20, 2002: LTER Coordinating Committee Meeting at Sevilleta, New Mexico, <http://intranet.lternet.edu/meetings/future/ScienceSymposiumTheme.html>