Published on *LTER Information Management* (http://im.lternet.edu)

LTER IM Practices

LTER Information Management Practices
**Note: Many of the documents here are out of date. More up-to-date resources can be found at the Environmental Data Initiative (https://environmentaldatainitiative.org/resources [1]) and DataOne (https://www.dataone.org/resources [2]).

Bibliography management

This is a section for practices relating to managing bibliographies

All-site Bibliography

The U.S. Long-Term Ecological Reseach (LTER) Network All-site Bibliography serves to account for the scientific contributions of the LTER Program, facilitate cross-site synthesis and synthetic studies, and generate new interest in LTER sites and LTER research. The infrastructure supporting the bibliographic records and the functionality of it's interfaces continue to evolve as described in Brunt and Maddux (2002). There have been some major jumps forward – each time producing a functional new product. However, until now, none of these products have been maintainable and thus quickly became obsolete. The first successful attempt used unique delivery scripts for each site that quickly became obsolete. This solution also relied on indexing software that, you guessed it, quickly became obsolete. The task of building a distributed all-site bibliography was intractable because of the heterogeneity of the way sites were storing and managing their bibliography data and the frequency at which these methods changed. More recently, we standardized on a particular end-user software package, EndNote®, to take advantage of a proprietary web-publishing solution, Reference Web Poster®. This solution was successful because of the standardization but extremely limited in the way information could be retrieved and used. Neither of these previous solutions had the power of an open relational database management system behind them. This was the most important requirement in attempting to provide any value-added components to this very useful data set.

The LTER All-Site Bibliography has now hopefully completed its last platform migration. Now, finally, implemented in a highly normalized relational database model, it should serve the LTER Network for years to come. Moving the bibliography database out of proprietary software and into an open rdbms framework has allowed us to now focus on generating standards-based outputs, easing the burden of updates for sites, providing useful searching and reporting, and providing virtual bibliographies.

General Guidelines for LTER Bibliographic Entries:
Provide Year, title, author, reference type, and accession number for all bibliographic entries. Year should be either a 4-digit year format or “in-press”. Other dates, such as conference dates should go in the date field.

Enter All authors (including the first one) as Lastname, (Firstname Middle Initial) or (Initials).

If you are generating bibliographic entries from a database use the EndNote controlled vocabulary for “reference type” (e.g., “Journal Article”, “Book Section”, “Edited Book”, “Conference Proceedings”, “Thesis”). These drive the automatic display and editing interfaces.

Use full journal names as registered with the Library of Congress. Do not use society specific Journal abbreviations in the Journal/secondary_title field. EndNote can easily convert these names for you. If you really need them, put them in one of the custom fields.

Changes in the current version:
1. The database now recognizes only LTER funded publications (however you determine that). There is no longer a checkbox for this attribute.
2. Create individual author entries. EndNote now supports individual author entries and those entries that have them are ready for EML. This level of granularity is necessary to be compliant with EML and to allow the linking of individuals to publications in the database.
3. Accession number is now required and is represented by %M – accession numbers must begin with the site code but may be followed by any reasonable numbering scheme. Email me when you’ve successfully incorporated accession number and I will delete your old entries if necessary.
4. The database now supports digital object identifiers (DOI) in use by most commercial publishers. These numbers should be represented by the %R tag.
5. Each entry in a record must be on a single line. This is the most common error that I encounter. Endnote will let you put in carriage returns in some fields but I can’t deal with them. You’ll have to filter them out. The new validate script is sensitive to this error.
6. For published abstracts use type “Conference Paper” - do not use type “Conference Proceedings”.

The site management interface can be reached from the “Administration” tab on http://intranet.lternet.edu [3] or via: http://savanna.lternet.edu/biblio/index.php [4] Either approach requires a login - successful authentication should then provide you with authorization for managing your sites bibliography entries.

You have three options for approaches to take for entering bibliographic entries:
1. Add Record |Enter the the bibliographic entry one record at a time directly into the web interface. (This feature is primarily for the standard methods and obfs bibliography groups).
2. Bulk Upload | Upload an EndNote Export file.
3. URL Harvest | Harvest an EndNote Export generating URL.

The process to follow for the 2 or 3 is to upload or harvest first – you can verify this to be successful by clicking on the filename or “site Registered URL” link and viewing your file. Then, select “validate” from the record for the file you just uploaded. This will produce a log of information that will help you to debug your file (“view log”). Unfortunately it’s not yet as friendly as it could be and it’s based on records in the file and not the accession numbers but you relate the two and usually find the problems. Once the file validates successfully there will be an “ingest” menu option available. Ingestion completes 99% of the time now because of the more detailed validation step and the accession numbers. You can view your records by returning to Return to Site Admin Page |. From this page you can you can see a brief report of your current bibliographic records (and personnel records) in the database at
View Database Report |.

Note: Endnote export format is no longer the default in EndNote 7 and 8. You have to go and select it from the style list and make it available and then check it off in the pull-down before it works. If you have a URL that you want harvested on a regular basis for this purpose please email it to me and I will incorporate it into the harvester configuration once it's complete. There is an EndNote Tag reference available at:

http://savanna.lternet.edu/reports/endnotetags.php [5]

As always, Email tech_support@LTERnet.edu [6] and/or call me directly (505.277.2535) for help or bug reports.

Compatibility with EML
1. See attached file below for a stylesheet to convert documents written against EML 2.1.0 schema to Endnote text format appropirate for this database. The file extension is set to ".txt" to comply with drupal requirements, and should be edited to ".xsl" after download. The stylesheet was contributed by SBC, and is simple to adapt for other systems (see notes in the file). See sbc.lternet.edu/info_management for other material on using EML for citations.

2. export from the DB as EML 2.1 is available from: TBA

Related Links

- Status Query of bibliography (517) [7]
- 2005 Databits article describing bibliography database (350) [8]
- EndNote Tag Tool (389) [9]
- In-Press Journal Article Query (435) [10]
- Link to bibliography admin pages (357) [11]
- Model for URL based input from GCE (413) [12]
- 2002 Databits article (330) [13]

Attachment	Size
bibliography_guidelines.pdf [14]	17.13 KB
emlCitation2endnoteText_xsl.txt [15]	12.48 KB

NSF reports and BibTeX files

The website for submitting your annual report (research.gov) now accepts bibliographic citations in BibTeX format. Be warned, though, that they constrain both the types and the fields somewhat. Below is the list of BibTeX fields that research.gov accepts, as of August 2014:

Below is an explicit list of the bibtex entries and fields that will be accepted when submitting bibtex (.bib) files to RPPR
Also see: http://www.research.gov/common/robohelp/public/WebHelp/Project_Reports.htm [16]
Find the string 'bibtex' on the page, and there is a table of allowed fields, which does not quite match this one. This list includes optional fields, and at least one that (de facto) can be uploaded via bibtex files.

For general bibtex examples, see https://verbosus.com/bibtex-style-examples.html [17]

THis list was originally supplied by help-at-research.gov, with the statement "Any other bibtex entities, aside from the ones listed above, are not supported by RPPR at this time and will not be accepted via BibTeX file submission.". This is not true. We have already found other fields that are accepted, and some on the original list were wrong. If you learn something new that would benefit someone else, please add it to this page, or tell Margaret.

Key:
RPPR: the acronym for the NSF code ingesting citations
BIBTEX-ENTITY [equivalent rppr product]
* required field
+ optional field

ARTICLE [Journal]

* author
* title
* journal
* year [if Status of Publication is "Published"]
+ volume
+ number [Issue]
+ pages
+ DOI (reported as usable by Sven)

BOOK/BOOKLET [Book]
* author
* title
* year [if Status of Publication is "Published"]
+ edition
+ volume
+ editor
+ publisher
+ address
+ ISBN, ISSN (reported by margaret)

INBOOK/INCOLLECTION [Book Chapter]
* title (chapter title)
* author
* booktitle
* year [if Status of Publication is "Published"]
+ editor (research.gov will add this list of names to the chapter's authors. but they don't appear to be re-parsed, just 2 sequential lists. reported by mob)
+ edition
+ pages
+ volume
+ publisher
+ address

MASTERTHESIS/PHDTHESIS [Thesis or Dissertation]
* title
* author
* institution
* year

CONFERENCE/PROCEEDINGS/INPROCEEDINGS [Conference or Presentation]
* author
* title
* booktitle [Conference Name] (not 'organization', as one would expect)
* year [if Status of Publication is "Published"]
+ address

Notes on RPPR Behavior:
1. Rejected entries:
If a product is rejected in the results after adding it via a file upload, this either means that a required field is missing from the entry (rejected – red x will appear next to it) or a product type is not supported (result will be greyed out).
Also: some invalid entries do not generate a red-x, instead they don't cause failure till the final commit. Bummer, that!. I think this behavior might only apply to the year-field, where it's presence is dependent on another field which is manually-entered. see below.

2. ancillary data
NSF requests/requires ancillary data. Can I label these in bibtex fields so that they can be ingested automatically along with the citation itself?
Unfortunately, it is not possible for ancillary data to be included in the citation file and ingested automatically by RPPR. The user will need to manually input the ancillary data through the UI. The bibtex ingestion process is handled by a 3rd party application which we have no control over, and therefore it is unlikely that ancillary data will be supported anytime in the near future.

3. year field:
RPPR seems to read in the first 4 characters of the string in the year-field, and it seems to be able to recognize legitimate years. No fooling RPPR! The error message is cryptic, however. You will not see the red x. you'll get a green error message at the top of the page that ingestion failed, after you've tried to finalize the upload -- that is, AFTER you gone through the menus for the other required fields for every entry. (status, federal-support-acknowlegement).

4. In-press pubs:
Use an empty string for the year field (eg, year={}). year is not required if status is not 'published', ie, for in press pubs. You cannot put 'in press' in the year field (you'll get an error when you try to finalize the upload, but no reason why it occurred). You can include in-press citations in the bibtex upload. However, see below under "status"

5. status
On all citations, you have to go back later and select a status with the dropdowns. The default is 'published'. It's easiest of you group your citations into bibtex-files by status, so you can set the status of the whole batch to be the same. If you mix status-types in your upload, you'll have to figure each one out individually from the tiny, grayed-out citation-text that research.gov constructs.

The three groups of citations SBC used in 2014 were:
1. status = 'awaiting publication": these were our currently "in press" pubs. we deliberately leave off the year. research.gov will highlight these as needing more info next year
2. status = "published": these are all non-presentations whose citations are complete.
3. status = "other": presentations. we felt these should not be classified as published, although that is still an option.

Testing
You can test your bibtex uploads (and other report submissions) at: http://demo.research.gov [18]

Data Management

LTER Data Management Practices.

Related Links

- Hook et al. Best Practices for Preparing Environme ntal Data Sets to Share and Archive (1522) [19]
- ORNL DAAC data management for data providers (496) [20]

Accessing Data in the NIS

Best practices documentation developed during a 2011 NIS Workflow Development Workshop at LNO is attached. Participants included Corinna Gries (NTL), John Porter (VCR), Ben Ruddell (CAP), Mark Servilla (LNO), Wade Sheldon (GCE) and Jonathan Walsh (BES). More information about the workshop is available here [21].

Screen casts of the different workflows [22]

See also next chapter for more detailed instructions [23]

Attachment	Size
NISdataworkflowsbestpractices0.2.pdf [24]	504.51 KB
NISdataworkflowsbestpractices0.1.pdf [25]	463.2 KB

PASTAprog - generating R, Matlab, SPSS programs based on EML

The PASTAprog web service allows you to rapidly create and run statistical programs for the analysis of Long-Term Ecological Research (LTER) data from the LTER PASTA Data Portal. PASTAprog uses Ecological Metadata Language (EML) metadata to generate statistical programs that:

1. Download and ingest tabular data in LTER Datasets
2. Run simple statistical summaries on data columns

PASTAprog is designed to use the information from EML metadata documents to streamline the “routine” tasks associated with ingesting LTER data. Users are expected to be knowledgeable about the statistical tools they are using so they can amend and extend the programs created by PASTAprog to support sophisticated and problem-specific analyses. Statistical packages currently supported by PASTAprog are:

1. MATLAB (stylesheet from Wade Sheldon, GCE)
2. R
3. Statistical Analysis System (SAS), and
4. Statistical Package for the Social Sciences (SPSS)

Use of PASTAprog
The PASTAprog REST-based web service is invoked using a URL consisting of at least 3 elements:

- The base URL – typically: http://www.vcrlter.virginia.edu/webservice/PASTAprog/ [26]
- An LTER dataset identifier or package ID consisting of a scope, identification number and a revision number.
  For example: “knb-lter-vcr.26.14”. These package IDs can be found by searching the PASTA Data Portal for datasets.
- A suffix that indicates the type of program you wish to generate:
  - .r for a R statistical program
  - .sas for a SAS statistical program
  - .spss for a SPSS statistical program
  - .m for a MATLAB program (see below for special notes regarding use with MATLAB)

Thus a complete URL for invocation would look like:
http://www.vcrlter.virginia.edu/webservice/PASTAprog/knb-lter-vcr.26.14.r [27]
If used in a web browser, a program will be generated that can be copied and pasted into a text editor or into command interfaces for

statistical packages. However, the URL can also be included directly in statistical programs using functions that retrieve code from a URL and run it without using a web browser. For example, the R “source()” function causes R to read a program from the URL specified and run it. As in:
source(“http://www.vcrlter.virginia.edu/webservice/PASTAprog/knb-lter-vcr.26.14.r”, echo=TRUE)

Using PASTAprog with Other Metadata Repositories
By default, PASTAprog searches the PASTA repository for the metadata document to transform. However that default can be overridden to point to any web-accessible metadata document by appending a “emlurl” attribute to the URL as in:
http://www.vcrlter.virginia.edu/webservice/PASTAprog/myprog.sas?emlurl=h... [28]
Where http://myserver.edu/myEMLdoc.xml [29] is the web address of an EML metadata document and myprog.sas is the name of the SAS program you wish to generate. The easiest way to get the URL for an EML metadata document in another repository, such as the LTER Metacat, is to use a web browser and use “copy link” (or similar) to copy the URL for pasting into the web service URL.

Datasets with Multiple Data Tables
PASTAprog generates programs that read all of the data tables documented in the metadata for a specific dataset, adding sequence numbers to the names of the constructed data structures. Thus a dataset with three dataTables will generate R code for creating three data frames, dataTable1, dataTable2 and dataTable3. SAS creates working datasets named datafile1, datafile2 etc. Of special note, the program for generating SPSS creates statements that sequentially read each of the data tables into the active system file (overwriting previous data tables), therefore it is necessary to add SAVE statements to save each data table to a separate file.

PASTAprog and Data Retrieval
When run, the programs created by PASTAprog should automatically retrieve the data from within the statistical package without additional user actions. However, this feature depends on the EML metadata file correctly specifying the URLs that will retrieve the correct data files. If the data URLs in the EML document is incorrect, or point to a form, rather than directly to data files, it will be necessary to download the data manually and edit the input section of the program to reflect the location of the data file on the local PC. This should typically not be necessary for publicly-available data in PASTA, because the PASTA system itself provides a copy of the data. However, for data in the LTER Metacat or in other EML documents outside PASTA, the reliability of the links to data files are more variable, so that some automatically-generated programs will work without modification, but others will require editing and for the data to be manual downloaded.
Each time the program created by PASTAprog is run it downloads a fresh copy of the data from the PASTA server. For small datasets, this is not a problem. However for really large datasets, users may wish to store a copy of the data locally and, as discussed above, modify the program to read from local files. Alternatively, the program can be run once, unmodified, and the resulting data structure saved in the formats specific to individual statistical packages. For example, in SAS a data table can be saved as a permanent dataset, or in SPSS saved as a system file, or in R as a saved data frame or workspace. These structures can be retrieved for further analysis without requiring re-ingestion of the data.

Statistical Package-Specific Notes
MATLAB: MATLAB does not allow function file names to include any periods, other than the trailing “.m”. Therefore PASTAprog allows the substitution of the underscore character “_” in place of periods when specifying the package ID. Thus, knb-lter-vcr_26_14.m is a valid replacement for knb-lter-vcr.26.14.m. Alternatively, if an emlurl attribute is specified the program name can be whatever you wish to specify, as long as it ends in “.m” to indicate that a MATLAB progam is desired. The MATLAB stylesheet (http://gce-lter.marsci.uga.edu/public/xsl/toolbox/EMLdataset2mfile.xsl [30]) was created by Wade Sheldon (wsheldon@lternet.edu [31]) at the Georgia Coastal Ecosystems LTER.
Note that the m-file produced by the MATLAB stylesheet contains help on function syntax and usage. For example, if you request knb-lter-vcr_26_14.m, change to the directory where the file was downloaded within MATLAB and type "help knb-lter-vcr_26_14" to view the embedded help text and "data = knb-lter-vcr_26_14" to run the program. Also, in contrast to the R, SAS and SPSS programs the MATLAB program does not perform any specific analyses. All data columns and attribute information are imported as arrays along with metadata content, which are stored in a structure variable for use with MATLAB analyses and plotting tools. For a more complete analytical solution for EML-described data in PASTA, see the GCE Data Toolbox for MATLAB (https://gce-svn.marsci.uga.edu/trac/GCE_Toolbox [32])
R: Data tables are ingested into R data.frames named dataTable1, dataTable2 etc. If there are irregularities in the formatting of one or more numerical data values in a column of data (e.g., a value has a letter O instead of a zero (0) in a number, or non-numeric missing value indicators other than NA), R may automatically convert a data vector from mode NUMERIC to type FACTOR. PASTAprog-generated programs automatically test of the vector mode in R and change the mode back to numeric for columns designated as numeric in the metadata. This may result in the generation of missing value NA codes for data that is really non-numeric.
The current version of the R program does not conduct range checks or do anything with missing values codes other than NA.
SAS: In addition to statistical summaries the SAS programs created by PASTAprog include variable and value-labeling statements, range checks for numerical data and recognize missing values, as specified in the corresponding metadata document.
SPSS: As of this writing, the SPSS programs written by PASTAprog do not automatically retrieve the data from a URL. This feature will be added once I find code that will do the retrieval from inside SPSS. For now, you will need to download the data manually and modify the path to the data file in the program, as specified in the comments. Note that if there are multiple input data tables, you will need to add SAVE statements to save each of the data tables to a system file of your choice. Otherwise, only the last data table read will be available. In addition to statistical summaries the SPSS programs created by PASTAprog include variable and value-labeling statements, range checks for numerical data and recognize missing values, as specified in the corresponding metadata document.

How PASTAprog Works
PASTAprog uses eXtensible Markup Language (XML) stylesheets to transform EML metadata documents (which are XML documents). The system is implemented as a set of PHP and Perl tools that parse the input URL, fetch the EML metadata document and apply the stylesheet transformation. Copies of stylesheets used to create the statistical programs are available in the LTER Network SVN repository (http://svn.lternet.edu [33]). Users are encouraged to improve the stylesheets so that they create better code for subsequent incorporation into the web service.

Troubleshooting
The success of programs created using PASTAprog is ultimately dictated by the quality of the underlying EML metadata and data. If the

metadata incorrectly identifies data columns, the program generated by PASTAprog will as well. Some metadata, while technically correct, incorporate bad practices that may cause specific statistical packages to fail. For example, some EML metadata may include attribute or variable names that include special characters, spaces or mathematical operations as part of the attribute names. This can be highly confusing to statistical packages such names violate the naming conventions of the statistical package, causing programs to fail when run. Some automatic fixes have been incorporated into the PASTAprog program generation, but there may still be times when this automatic correction will fail. When there are inconsistencies between the metadata and the data or problems with attribute naming conventions, it may be necessary to manually edit and correct the PASTAprog-generated program.
Dates are also a common source of problems because they are encoded so many different ways in ecological data, either as combined fields (e.g., “2012-09-07”) or as separate variables. Metadata sometimes lists date data as strings or separate columns of numbers reflecting year, month, day etc. For this reason, you will probably need to do some additional coding within a statistical package to get dates to perform reliably.

Survey on the software used at the USA LTER sites

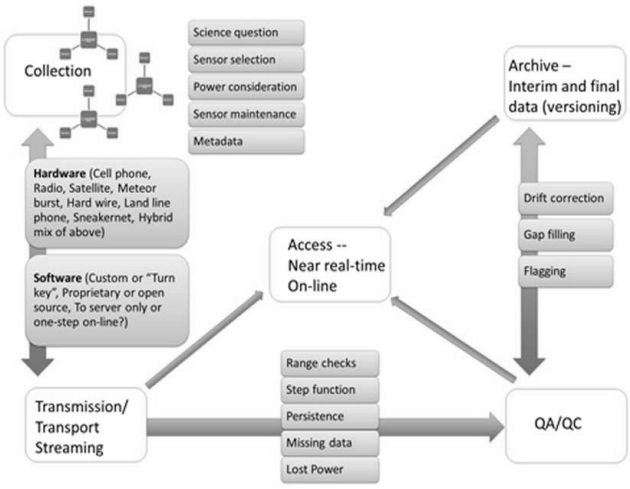
This is the newest version (2005) of a survey designed and performed by Karen Baker where the IMs inform what software they use for specific data management tasks.

Attachment	Size
im_results_05.pdf [34]	48.39 KB

Sensor Data (Material is moving off-site)

Please note: this material is moving to the ESIP EnviroSensing Cluster [35] please don't edit here

In this section we'll discuss all aspects of sensor data and their life cycle from collection to availability and analysis.



Sensor, Site, and Platform selection

Please note: this material is moving to the ESIP EnviroSensing Cluster [35] please don't edit here

Sensor, Site, and Platform Selection
When planning new instrumented research experiments, careful thought should go into the sensor, site, and platform selection. Here, we list several considerations.

- Experimental/project design
  - how do the site and platform fit in with the design?
- Site accessibility
  - physical access *(e.g., by hiking, 4x4, boat)*
  - permission/permits *(e.g., private vs. public vs. federal lands)*
  - is there 24x7x365 access?
  - getting equipment to the site for project construction?
  - enabeling access without impact on environment
- Environmental considerations
  - topography
  - vegetation
  - visibility
  - weather *(e.g., how to deal with extreme temperatures, tropical storms, lightning, snow, uv, rust, wind, lightning, salt water and crystalization,)*
  - wildlife *(e.g., using conduit or fencing, bird specific issues, soil invertebrates)*
  - interference of structure on measurement
- Security
  - protection from vandalism
  - security of the platform (data loggers, buoys)
- Power
  - photovoltaic systems *(e.g., solar panels, batteries)*
  - wind
  - sizing power systems *(e.g., power calculation spreadsheets, sun charts)*
- Communication
  - datalogger-to-sensor communication *(e.g., will wiring need conduit or burial?)*
  - telemetry *(e.g., line-of-sight needed for WiFi connections)*
- Platforms
  - terrestrial sites *(e.g., tripod, tower, pole, tree, tram, kite, uav, below-ground surface, on-ground surface)*
  - freshwater stream sites
  - freshwater lake sites *(e.g., buoy, raft)*

        - Mooring anchors: Iron has the highest density, concrete has a lot less. Concrete filled with scrap iron works well
        - Anchor line considerations: Length depends on water level changes, wind and wave action, other boater and angler traffic on lake. Adding a length of heavy chain adds to flexibility and keeping the buoy in place. Three line moorings should not spin: rotation simply moves the maximum tension from one line to the next. Similarly for two line moorings which are useful in tidal channels where you can anticipate the stress directions. Swivels are necessary on single line moorings or the rotational energy goes into the line with non-optimal results, e.g. shortening of the line, lateral stress on connectors, increased stress on individual strands, etc. Wherever rope meets metal, be it the anchor or an attached chain, use a metal eyelets so that the knotted rope loop goes around the eyelet and you end up with metal eyelet against the metal chain link or anchor.
    - marine sites *(e.g., buoy, tower, rugged containers)*
    - other *(e.g., satellite, airplane)*
- Dataloggers
    - vendors *(e.g., campbell, onset)*
    - software *(e.g., loggernet)*
    - wiring diagrams
- Sensors
    - precision and accuracy
    - power requirements
    - sampling frequency
    - deployment
    - maintenance *(e.g., calibration frequency, removal of bird poop)*
    - reliability
    - robustness
    - communications *(e.g., datalogger compatibility)*
    - dependence on other sensors *(e.g., Vaisala CO2 sensors depend on barometric pressure and temperature)*
    - wiring *(e.g., max distances, extending sensors)*
    - labeling
    - vendor support
-
- Budget
    -

Data Acquisition and Transmission

Please note: this material is moving to the ESIP EnviroSensing Cluster [35] please don't edit here

Data Acquisition and Transmission
Here, we list several considerations for data acquisition and transmission.

- Acquisition Methods
    - manual
    - Wi-Fi
    - cellular
    - satellite
    - serial
- Bandwidth
    - bandwidth requirements *(e.g., is 10Hz data needed?)*
    - where to get broadband POP?
    - ability of equipment to work with required bandwidth
-
- Line of sight (LOS)
    - environment, topography, and vegetation
    - LOS calculators
    - are repeaters required?
- Repeaters
    - *See site and platform selection considerations* [36]
    - distance to repeater
- Power
    - how important is real-time accessibility? *(e.g., what is desired collection frequency?)*
    - power requirements
    - on-site buffer size?
- Other considerations
    - reliability of transmission mode
    - reliability and robustness of equipment
    - configuration and software *(e.g., what level of expertise is needed?, computer requirements)*
    - serial-to-ethernet devices
    - budget

Sensor Management, Tracking, and Documentation

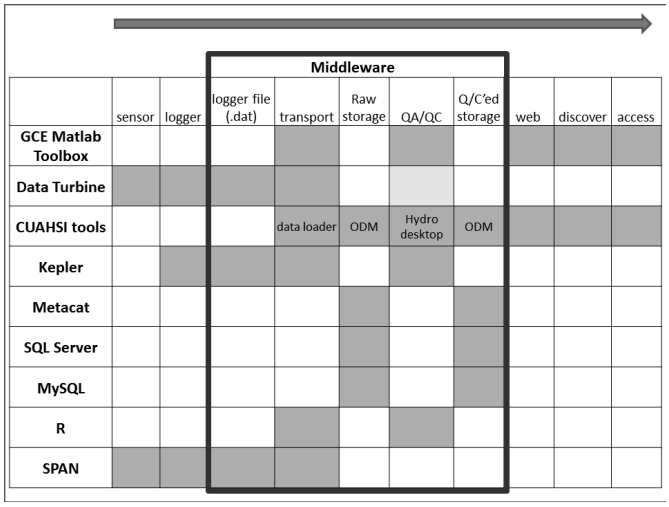
Please note: this material is moving to the ESIP EnviroSensing Cluster [35] please don't edit here

- Sensor Identification & Tracking
  - Location Information (Spatial-Temporal)
  - Service History
  - Tracking Options
    - Bar Codes
    - Geo-Location Tags
    - Microchip Encoded Sensors (NEON 'Grape')
- Maintenance & Tracking of Sensor Events
  - Deployments (Spatial-Temporal)
  - Methodology: Frequency, Calibration, Quality Filters, Sampling Program, Equipment Change
  - Calibrations
  - Other Service Actions
  - Failures
  - Replacement
- Middleware Capabilities
  - Workflow Management
  - Proprietary Software
- Geo-Location Referencing
  - Geo-Referenced Data Stream
  - Photo Documentation
    - Hemispheric photos
    - Time-Series Image Collections
- Data Stream Processing
  - System Configuration (Hardware/Software/History/Legacy)
    - Datalogger Program Archive
    - Sensor Wiring Diagrams
  - Data Workflow
    - Processing Procedures: Current Protocol, Legacy Operations
    - Quality Filters: Application/Log
    - Quality Flags: Application/Log
      - Quality Control Actions
        - Automated
        - Human-Reviewed
        - Log

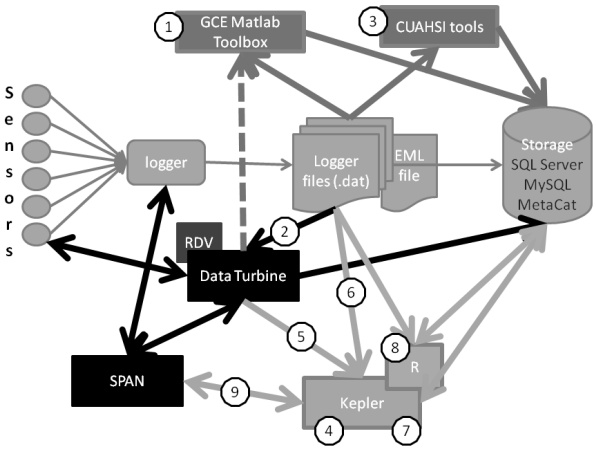
Streaming Data Management Middleware

Please note: this material is moving to the ESIP EnviroSensing Cluster [35] please don't edit here

- Purpose of middleware
  - Data storage / data handling
  - Data aggregation, formatting, filtering
  - Documentation
  - Automated QA/QC on data streams
  - Archiving



	sensor	logger	Middleware: logger file (.dat)	transport	Raw storage	QA/QC	Q/C'ed storage	web	discover	access
GCE Matlab Toolbox										
Data Turbine										
CUAHSI tools				data loader	ODM	Hydro desktop	ODM			
Kepler										
Metacat										
SQL Server										
MySQL										
R										
SPAN										



- Software/middleware

- Proprietary
  - Campbell Scientific LoggerNet (http://www.campbellsci.com/ [37])
  - Hobo (http://www.onsetcomp.com/ [38])
  - Vista Data Vision (http://www.vistadatavision.com/ [39])
  - YSI EcoNet (http://www.ysieconet.com/ [40])
  - Nexsen's Technology (http://nexsens.com [41])
- Proprietary with limited open source (or free package that will fit into a proprietary package)
  - GCE MatLab Toolbox (#1 in graph)
    - web site [42]
    - Presentations [43]
- Open source
  - CUAHSI HIS (#3 in graph)
    - Website [44]
    - Presentations, demos and exercises [45]
  - Open Source DataTurbine Initiative (#2 in graph)
    - Website [46]
    - Presentations [47]
  - Kepler Project
    - Website [48]
    - Presentations, demos and exercises
      - Introduction [47] (#4 in graph)
      - Use with DataTurbine [49] (#5 in graph)
      - Use with databases [50](#6 in graph)
      - Scheduled workflows [51](#7 in graph)
      - Sensor platform [52](#9 in graph)
  - R-project libraries (#8 in graph)
    - Website [53]
    - Presentations [54]
- Custom research applications
  - Open source
    - Python, PHP, MySQL, etc.
  - Proprietary
    - Matlab, Excel, SQLServer, etc.

Sensor Data Quality Assurance/Quality Control (QA/QC)

Please note: this material is moving to the ESIP EnviroSensing Cluster [35] please don't edit here

- Sensor Quality Assurance (QA) (preventative)
  - Maintenance and calibration schedule
  - Sensor redundancy
  - Continuous monitoring and evaluating
- Near real-time processing and Quality Control (QC) on data streams
  - Timestamp integrity - sequential, fixed intervals
  - Range checks - extreme values based on sensor specifications, or based on historical seasonal ranges
- Other methods for near real-time or post-streaming QC
  - Manual "eyes-on" checks
    - graphics software for displaying near real time data streams
  - Plausibility checks, e.g., TMAX-TMIN>0
  - Variance checks - outlier detection, e.g., Sigma (standard deviation) or Delta/step (difference of subsequent pairs)
  - Sensor drift checks - integration of several data streams (correlations or conditioning with redundant or nearby sensors)
  - Persistence checks - repeating values
- Documentation of methods applied to data
  - Identify streaming QC methods, thresholds, assumptions
  - If no QC, that should be made clear too
- Data qualifiers - data flags
  - Many vocabularies in use - standardize or cross-walk?
  - Rich vocabulary of fine-grained flags for streaming data – intended to guide local review
  - Simpler vocabulary of flags for “final” data for public consumption
  - Define new data columns in lieu of data qualifiers?
    - Method shifts, sensor changes
    - Place key documentation as close to data value as possible

Sensor Data Archiving

Please note: this material is moving to the ESIP EnviroSensing Cluster [35] please don't edit here

- Archival data formats
  - Databases
  - File systems
  - Netcdf, RDF, CUAHSI HIS Observation Data Model, etc.

- Versioning
  - Create a "citable" database: monthly, annual, or periodic snapshots
  - Release of provisional vs. final data
  - Tracking changes to the data, e.g., audit trail
- Identify data level - maintain/provide products at different data levels
  - Level 0 - Raw data, no QC, no data qualifiers applied (data flags)
    - Preservation of original data streams
    - Datalogger conversion of units and formats may be acceptable
  - Level 1 - QC'd, calibrated data, qualifiers added
    - Provisional level (near real-time preparation)
    - Published level (delayed release)
  - Level 2
    - Gap-filled, estimated, or aggregated data
    - Involves interpretation – multiple algorithms possible

GIS Data Management

LTER GIS Data Management Practices

The LTER GIS Working Group has been an informal group of individuals interested in GIS/Remote Sensing Issues/Data related to LTER Information Management. The group first met at the 2000 All Scientist Meeting at SnowBird. The group has met several times, at special meetings through a combination of Network level funding and cooperative efforts with the San Diego Supercomputing Center.

2007 IM Meeting Report Summary:
The working group updated recommendations for GIS at sites. Recommendations included basic spatial information and directions for geo-spatial referencing for study locations. We talked about developing a centralized portal for users to search, display, and access all site spatial data and a strategy for providing basic map service capacity for every site in the network. This group would like to hold an annual GIS workshop/meeting in conjunction with the annual IM meeting, form a GIS Analysis Group to investigate methods for facilitating data analysis across sites and assist researchers with synthesis, and identify necessary funding to accomplish these recommendations.

Beginning Google Maps API Instructions [55]

Related Links

- Best Practices for Documenting Spatial Data (524) [56]
- LTERmapS (529) [57]

Creating shapefiles and kml documents for study site locations

place holder for document on how to create shape files and kml documents for study site locations. These files will be used to store data for linking locations to eml documents.

Servers

Server Practices

Metadata

LTER Metadata Practices

Best Practices and Guides

LTER IMC Guides
N.B. This section is obsolescent. Info about EML and it's use is now maintained by the Environmental Data Initiative. See https://environmentaldatainitiative.org/resources [1]
See "Five phases of data publishing" > "Phase 3"

EML Best Practices, V3
https://environmentaldatainitiative.org/resources/five-phases-of-data-pu... [58]
Previous version (1 and 2, 2004 and 2011 respectively) are below.

Best Practices for Units
[Page below is under construction].

Best Practices for Keywords

[Page under construction]: http://im.lternet.edu/VocabBestPractices [59]

External Guides

ESIP Attribute conventions for data discovery
Uses netcdf format, but terms are mappable to EML
http://wiki.esipfed.org/index.php/Attribute_Convention_for_Data_Discovery_(ACDD)

Related Links

- Best Practices for Documenting Spatial Data (455) [60]

Best Practices for Units

Best Practices for Units
[UNDER CONSTRUCTION]
Another resource for the units model used by EML (thanks, Gastil). It also contains some examples of mathematical and statistical operations that could be allowed on different measurement types. This is also linked on the units dictionary page.
http://en.wikipedia.org/wiki/Level_of_measurement [61]

EML Best Practices for LTER Sites

N.B. This section is obsolescent. Info about EML and it's use is now maintained by the Environmental Data Initiative. See https://environmentaldatainitiative.org/resources [1]
See: "Five phases of data publishing" > "Phase 3"

EML Best Practices, V3
https://environmentaldatainitiative.org/resources/five-phases-of-data-pu... [58]
Previous version (1 and 2, 2004 and 2011 respectively) are below.

Earlier versions of the document are available from the network document archive.

EML Handbook, 2003
EML: Practical Application for Scientists [62] by D. Blankman and J. McGann, LNO. This describes EML use at a basic level. Note some important details such as <objectName> and <attributeLabel> are omitted. However, this is a good place to start an introduction to EML. Note that even back in 2003 it was advised that the data url be at the entity level.

Attachment	Size
2011: EML Best Practices for LTER Sites V 2 (PDF) [63]	880.28 KB
2011: example_full_a.xml [64]	33.94 KB
2011: example_provenance.xml [65]	6.65 KB
2011: example_lter_maps.xml [66]	61.52 KB

I. Introduction

The Ecological Metadata Language (EML) was adopted in 2003 as the exchange format for metadata contributed to the LTER network. As such, EML is one part of an LTER site's information management system. This document contains recommended current views for best practices for EML content for our network's use, and is also intended to augment the EML schema documentation (normative documents) for a less technical audience. Some notes on implementation are included as appropriate. This is one component of several Best Practice documents available to LTER sites, and related documents and resources are listed in the Additional Resources (Section V). The recommendations are directed towards achieving the following specific goals:

a) Provide guidance and clarification in the implementation of EML for datasets

b) Minimize heterogeneity of LTER EML documents to simplify development and re-use of software built for EML datasets

c) Maximize interoperability of LTER EML documents to facilitate data synthesis

Please cite this document as:

Long Term Ecological Research Network. 2011. EML Best Practices for LTER Sites.

I.1. Changes from EML Best Practices Version 1 (2004)

The EML Best Practice recommendations have evolved over time. They are the cumulative result of several working groups comprised of LTER information managers and Network Office representatives (Section VI). Each group aligned its recommendations with the current capability at most of the contributing LTER sites. As of 2011, EML has been widely used for several years with multiple applications written against it, and the community has had the opportunity to observe the consequences of many content patterns. Following are the major changes between this document and Version 1 (2004).
EML Version: EML 2.1 is the current version of the specification, and all examples and recommendations refer to this. However, all recommendations for element content can be applied to EML 2.0.1.

Document Organization and EML "Metadata Levels": Version 1 (2004) of this document discussed specific elements in sections of "levels of EML completeness" which corresponded to functionality tiers identified by NISAC in 2004. It used the completeness levels' increasingly comprehensive descriptions of data resources so that EML could be implemented gradually. In this document (Version 2), the detailed content recommendations (Section II) are organized more simply and more or less in EML-document order. Other organizing concepts were considered, .e.g., distinguishing between 'human-readable' and 'machine-readable' elements. However, at present all information in EML is consumed by humans while only few elements are used by applications, (e.g. access control, geographic coverage, and the structural information in the entity tree). Few applications are able to make semantic use of TextType fields such as <**abstract**> and <**methodStep**>, although these fields can be machine-searched. Some fields with simple string content that is currently ad hoc (e.g., <**attributeName**>, <**customUnit**>) are currently undergoing a network wide standardization effort and will become machine-read in the near future.

Element Descriptions: Several important EML elements can be placed at multiple locations in a dataset, e.g., coverage and methods. To illustrate all possible locations, the allowable XPath locations are listed for every element. XPath is a representation of the document's hierarchical structure, which is similar to fileserver directory paths. For each element, recommendations for content have been grouped together when the element first appears (often at the dataset level). This does not imply that these trees belong only at that level; in fact, locating trees at the most granular level is still advantageous and recommended.

Data Availability: The LTER Network policy on availability is that data products should be generally available. Therefore, there are limited recommendations here for how to describe a data entity which is unavailable to the public. More information on access can be found in the EML normative documents.

Data Types: Recommendations have been added in Section I.2 for describing certain dataset types or for common situations encountered in LTER datasets, specifically, conversions of spatial metadata to EML, and "string-of-pearls" or "attribute – value" data models.

External Applications: A new section (III) has been added to contain recommendations for EML which is to be used in specific external applications. These include Metacat, the LTERMaPs project, and PASTA (specifically, the EML Congruency Checker and its use in validating LTER datasets, and provenance for derived datasets).

I.2. EML Management

The terms "dataset" and "data package" are somewhat interchangeable. The term "data package" is used here to mean the published unit of data and metadata together. "Dataset" has a special meaning within EML documents, because it is the top-level container for data objects (i.e., <**dataset**> as opposed to the other top-level elements, **<citation>, <software>** and **<protocol>**). However, a logical "dataset" may be something slightly different. For example, management needs for an ongoing collection may dictate that a dataset is published in increments, and the entire "Ongoing Climate Dataset" may be composed of multiple contributions, each of which could be an individual EML "dataset". Subsequent data access applications may use the term dataset differently.

I.2.1. Creating Datasets

Several approaches to creating datasets or data packages have emerged in the network and are all valid concepts. In general, this document does not recommend any one pattern. Following are several examples:
(1) Data collected with defined beginning and end dates are published in logical units, with all ancillary data are described together in one EML file (i.e., <**dataset**>). EML accommodates descriptions of multiple data entities in one metadata document. For instance, a data table could be accompanied by a KML file or shapefile to describe the sampling locations.
(2) Data from a large sampling campaign where many somewhat independent parameters are measured could either be combined into one EML dataset with multiple data entities, or broken up into several datasets, all accompanied by some of the same metadata, and each with other unique metadata. Both patterns are valid, and it is left up to the dataset designers to decide which style works best for their particular data and circumstances. It is wise to keep the data user in mind, and whether certain data entities naturally belong together.
(3) For time-series observations (instrumented or not), some designers choose to update existing EML metadata when methods or other ancillary material remains constant. Others may choose to create new datasets for each logical unit (e.g., a summer sampling season). There are advantages and disadvantages to both approaches. The first approach does not allow for strict versioning of the data while the second approach forces the user to download and integrate many separate entities.
(4) The dataset designer should consider some reasonable approaches to deciding which information is encoded directly in EML and which is better left in ancillary files (or tables). For example, codes used in a column of the data table may either be encoded directly in EML or stored in a "look-up table" (described in EML as another entity) and linked via a <**constraint**> (foreignKey constraint) element or described directly in the attribute with the enumeratedDomain/enitityCodeList/ tree. Another example is site locations, which may be encoded in EML under <**methods**> <**sampling**> <**spatialSamplingUnits**> or supplied as a KML or shape file and described as another entity. The decision between these two options probably depends mostly on the number of units to encode in EML. For example, 500 codes are better left in a second data entity, while ten code-definition pairs should be encoded directly in EML.

I.2.2 The Attribute – Value Data Model

The Attribute - Value or “string of pearls” data model is widely used for certain kinds of observational data where the more conventional matrix type model would cause many empty cells. This data model treats each point observation as a single record containing fields for location, time, variable (attribute) name, and value, plus various flags for methods, data quality etc. For ecologists, this data model is especially useful for (1) biodiversity and (2) sensor data that may otherwise require very wide tables in a matrix format, and/or have many empty cells e.g., when species are not regularly observed or sensors are not employed consistently.
The advantages of the Attribute - Value data model are its flexibility and efficiency. However datasets in this format are not easily described in the current version of EML. The basic problem is that the values in the value field do not necessarily share the same attributes (e.g. numerical type, collection methods, unit, and precision). So for now the best practices recommendations are as follows:

Biodiversity data: If all of the values in a given table share the same units (e.g., presence/absence, count, percent abundance), then the table should be represented in this Attribute – Value format and described accurately in EML. If not, then the data should be represented in matrix format, or possibly several matrices if the table is unreasonably wide.

Sensor data: The Attribute - Value format is especially efficient for sensor data and variations of this data model have been developed in communities handling large volumes of sensor data (e.g., ODM or Observation Data Model from CUAHSI). For a single site this data model is useful when sensors are frequently changed or redeployed or when several different sets of sensor deployments exist. For accurate description in EML, it is recommended that each data table contain only the Attribute – Value formatted values for a single sensor. If a single metadata document contains data from multiple sensors, values for each sensor should appear as separate entities (data tables) within one dataset. (Most of these data models describe the concept of a single “data stream,” which may be used to achieve this recommended approach dynamically.)

I.2.3. EML produced from Geographic Information Systems (GIS) systems

Several established standards exist for documenting spatial datasets. The most common formats are Federal Geographic Data Committee (FGDC) geospatial standard, National Biological Information Infrastructure (NBII) biological profile, International Standards Organization (ISO) standard 19115, and ArcGIS metadata format. Tools are available at http://www.fgdc.gov/metadata/geospatial-metadata-tools [67] for preparing FGDC compliant metadata, which can then be converted into EML using XSL transformation style sheets. The LTER network has the following stylesheets available:
esri2eml.xsl: converts XML documents from ArcCatalog metadata editor (up to ARCGIS version 9.3.1)
http://im.lternet.edu/project/Esri2Eml/docs [68]
bdp2eml.xsl: converts FGDC and NBII Biological Profile metadata to EML. http://im.lternet.edu/project/bdp2eml [69]
Valid EML will be produced by esri2eml.xsl if the following guidelines are followed when preparing your FGDC, ISO, or ArcGIS compliant metadata:
a) ***Personnel names***: Last name, first name (i.e., use a comma to separate last name and first name) and for multiple names and contact elements (phone, fax, e-mail), use repeating fields.
b) ***Attribute definition***: Every attribute value needs a corresponding value definition. If the value definition is not available, record the value definition as “unknown”, “none”, or null”.

II. Detailed content recommendations For Elements and Attributes

Following are general best practices for creating EML dataset metadata:

Metadata Distribution: Do not publicly distribute EML documents containing elements with incorrect information as dataset metadata (i.e. as a workaround for problems with metadata content availability or to meet EML validation requirements). EML produced as draft, demonstration or for testing purposes should be clearly identified as such and not contributed to public metadata archives or clearinghouses.

Text Elements: Use EML text formatting tags whenever possible (e.g. <section>, <para>, <orderedlist>, etc.). Use <literalLayout> only when the field must contain HTML for formatting that is not available with Text Type. Note that the TextType elements were taken from docbook, and so do not use the “camelCase” notation that was generally adopted for EML.

Versioning: It is recommended that metadata and data set versioning be handled at the site level. For example, data entities and corresponding metadata for ongoing long term datasets can be archived annually with correct end dates and versions of the EML. Package IDs of the format {scope}.{identifier}.{revision} will be used in the LTER-NIS. Scope is of the format "knb-lter-fls". Data updates or revisions to data or metadata should retain the same identifier while incrementing the revision.

EML “id” Attribute: Many EML elements are allowed to have an “id” attribute. In EML 2.1.0, all ids in a document must be unique. Care should be exercised when using id attributes to reference and re-use EML content. It may be preferable to duplicate content without ids when generating EML dynamically from a relational database system to avoid potential id conflicts. See below for more information on the id attribute.

High-priority Elements: To support locating datasets by time, geographic location, and taxonomically, metadata should provide as much information as possible, in the three <coverage> elements of <temporalCoverage> (when), <taxonomicCoverage> (what), and <geographicCoverage> (where) for the dataset.

Metadata should include detailed descriptions in the <project>, <methods>, <protocols>, and <intellectualRights> elements in order for a potential user to evaluate the relevance of the data package for their research study or synthesis project.

The change history should be recorded in the <maintenance> element.

The root element: <eml:eml>

This element is the root element in all EML documents. The XPath notation is:

/eml:eml

After the XML declaration, all EML documents must have one root element (<**eml:eml**>). There is a benefit to including a schema declaration which points to an online resource. It is recommended that the following schema declaration be included in the root element: xsi:schemaLocation="eml://ecoinformatics.org/eml-2.1.0 http://nis.lternet.edu/schemas/EML/eml-2.1.0/eml.xsd" [70]

An EML dataset is composed of up to three elements under the root element (<**eml:eml**>):

<**access**>

<**dataset**>

<**additionalMetadata**>

@packageId (XML attribute)

This attribute is required in all EML2.1 documents, and is found at this location (XPath):
/eml:eml/@packageId

As outlined in other sections (see versioning above, @**id** and <**alternateIdentifier**>, below) each site should manage unique identifiers and versioning at the local level (see @**system** discussion below). The LTER Network currently uses “Metacat” as its EML repository. Most sites synchronize the **@packageId** with the Metacat “docid” designated in the harvester or at document upload, however synchronization is not required. Metacat and its harvester rely on numerical data set ids and revision numbers for document management and synchronization. This may necessitate a workaround for sites that use non-numeric ids or don’t version data sets. Possible solutions include differentially generating EML optimized for Metacat (for sites capable of dynamic EML generation) and XSLT transformation.

Currently, the **@packageId** attribute is used to identify a site’s EML documents for searches in the LTER Data Catalog. A site’s **@packageId** attribute in EML contributed to the KNB Metacat should be standardized as follows, or another format agreed upon by the site and LNO Data Catalog managers.

> knb-lter-[site].[dataset number].[revision], e.g. knb-lter-fls.187.4

See Section III for other information about EML documents in Metacat.

id, system and scope (XML attribute group)

This attribute group can be used on these EML elements:
<**access**>
<**dataset**>
<**creator**>
<**associatedParty**>
<**contact**>
<**metadataProvider**>
<**publisher**>
<**coverage**>
<**geographicCoverage**>
<**temporalCoverage**,>
<**taxonomicCoverage**>
<**distribution**>
<**software**>
<**citation**>
<**protocol**>
<**project**>
<**dataTable**>
<**otherEntity**>
<**spatialRaster**>
<**spatialReference**>
<**spatialVector**>
<**storedProcedure**>
<**view**>

<**attribute**>
<**constraint**>

These three attributes are found as a group and are usually optional, unless the **@id** attribute is used as a reference. Each **@id** must be unique in one EML document, i.e. a <**creator**> must have a different id than a <**dataTable**>. If the same person appears as dataset creator and protocol or project creator, the same **@id** cannot be repeated, so either the content of the **@id** must be changed or a reference used for repeated instances.

The current restrictions can cause problems when content is drawn from a system with IDs (e.g. a personnel database), and is under consideration by the EML developers. Ideally the three attributes would work together. The **@scope** attribute can have one of two values, “system” or “document”. It is preferred that when the scope is set to “system”, that the **system** attribute defines the ID-system, the **@id** attribute content is (presumably) from that system.

Currently, a reasonable general practice should be to define a **system** on the <**eml:eml**> element and set it to the site (but not set the system attribute at any other level), and to set **scope**=“document” on elements other than <**eml:eml**>.

Example: attributes packageId, id, system, and scope

```
<?xml version="1.0" encoding="UTF-8"?>
<eml:eml xmlns:ds="eml://ecoinformatics.org/dataset-2.1.0"
  xmlns:xs="http://www.w3.org/2001/XMLSchema"
  xmlns:eml="eml://ecoinformatics.org/eml-2.1.0"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns:stmml="http://www.xml-cml.org/schema/stmml-1.1"
  xsi:schemaLocation="eml://ecoinformatics.org/eml-2.1.0 http://nis.lternet.edu/schemas/EML/eml-2.1.0/eml.xsd" [70]
  packageId="knb-lter-fls.21.3"
  system="FLS"
  scope="system">
```

access

The dataset title element is found at this location (XPath):

/eml:eml/access

/eml:eml/[entityType]/physical/distribution/access

<**access**> contains a list of rules defining permissions for this file and its data entity. Values must be applicable by the system where data is stored. Metacat access control format conforms to the LDAP “distinguishedName (dn)” for an individual, as in “uid=FLS,o=LTER,dc=ecoinformatics,dc=org”.

As of EML 2.1.0, <**access**> trees are allowed at two places: as the first child of the <**eml:eml**> root element (a sibling to <**dataset**>) for controlling access to the entire document, and in a **physical/distribution** tree for controlling access to the resource URL. Access elements for documents contributed to the KNB Metacat should be formed according to the Example below. With the exception of certain sensitive information, metadata should be publicly accessible. The <**access**> element is optional, and if omitted, presumably only the dataset submitter will be allowed access.

Example: access

```
 <access authSystem="knb" order="allowFirst" scope="document">
    <allow>
     <principal>uid=FLS,o=lter,dc=ecoinformatics,dc=org</principal>
     <permission>all</permission>
   </allow>
   <allow>
     <principal>public</principal>
     <permission>read</permission>
   </allow>
 </access>
```

datasets

This element is found at these locations (XPath):
/eml:eml/dataset

Under <**dataset**>, the following elements are available. Some are optional, but if they appear, this order is enforced by the schema. Generally, the recommendations are presented here in this order, with the exception of elements related to people and organizations which are grouped together so that the distinctions between the uses of those elements are clear. Elements that can appear at different levels within an EML file are discussed at their first appearance, or highest level (see also section I.1).

<**alternateIdentifer**>
<**shortName**>

<title>
<creator>
<metadataProvider>
<associatedParty>
<pubDate>
<language>
<series>
<abstract>
<keywordSet>
<additionalInfo>
<intellectualRights>
<distribution>
<coverage>
<purpose >
<maintenance>
<contact>
<publisher>
<pubPlace>
<project>

These elements are then followed by one or more elements for the data entity (or entities), designated by choosing:

[**dataTable** | **spatialRaster** | **spatialVector** | **storedProcedure** | **view** | **otherEntity**]

alternateIdentifier

The dataset title element is found at this location (XPath):
/eml:eml/ dataset/alternateIdentifier
/eml:eml/ dataset/[entity]/alternateIdentifier

The site's data set id should be listed as the EML <**alternateIdentifier**>, particularly when it differs from the "**packageId**" attribute in the <**eml:eml**> element. The <**alternateIdentifier**> should also be used to denote that a dataset belongs to more than one LTER site by including the each site's ID in a separate <**alternateIdentifier**> tag. At the entity level, the <**alternateIdentifier**> should contain an alternate name for the data table (or other entity) itself (see additional comments under entities, below.)

title (dataset)

The title element is found at this location (XPath):
/eml:eml/dataset/title
/eml:eml/method/methodStep/protocol/title
/eml:eml/project/title

The dataset <**title**> should be descriptive and mention the data collected, geographic context and research site (what, where). If the data will not be updated, the time frame (when).

Example: dataset, alternateIdentifier, shortName, title

```
<dataset id="FLS-1" system="FLS" scope = "system">
  <alternateIdentifier>FLS-1</alternateIdentifier>
  <shortName>Arthropods</shortName>
  <title>Long-term Ground Arthropod Monitoring Dataset at Ficity, USA from 1998 to 2003</title>
```

People and Organizations (Parties)

People and organizations are all described using a "ResponsibleParty" group of elements, which is found at these locations (XPath):
/eml:eml/dataset/creator
/eml:eml/dataset/contact
/eml:eml/dataset/metadataProvider
/eml:eml/dataset/associatedParty
/eml:eml/dataset/publisher
/eml:eml/dataset/project/creator
/eml:eml/dataset/method/methodStep/protocol/creator

General recommendations: When using <**individualName**> elements anywhere within an EML document, names should be constructed with English alphabetization in mind. Many sites have found that maintaining full contact information for every creator is impractical, however some important contact information should be kept up to date (see below). If a name includes a suffix, it should be included in the <**surName**> element after the last name.

It is recommended to include complete contact information for a permanent role that is independent of the person holding that position. For example, for an information manager or site contact, pay careful attention to phone number and use an e-mail alias that can be

passed on. (see below, under<**contact**>)

associatedParty

This element is found at this location (XPath):
/eml:eml/dataset/associatedParty

List other people who were involved with the data in some way (field technicians, students assistants, etc.) as <**associatedParty**>. All **<associatedParty>** trees require a <**role**> element. The parent university, institution, or agency could also be listed as an <**associatedParty**> using <**role**> of “owner” when appropriate.

Example: associatedParty

```
<associatedParty id="12010" system="FLS" scope="system">
  <individualName>
    <givenName>Ima</givenName>
     <surName>Testuser</surName>
  </individualName>
  <organizationName>FSL LTER</organizationName>
  <address>
     <deliveryPoint>Department for Ecology</deliveryPoint>
     <deliveryPoint>Fictitious State University</deliveryPoint>
     <deliveryPoint>PO Box 111111</deliveryPoint>
     <city>Ficity</city>
     <administrativeArea>FI</administrativeArea>
     <postalCode>11111-1111</postalCode>
  </address>
<phone phonetype="voice">(999) 999-9999</phone>
<electronicMailAddress>itestuser@lternet.edu</electronicMailAddress>
<onlineUrl>http://search.lternet.edu/directory_view.php?personid=12010&amp;query=itestuser</onlineUrl>
<role>Technician</role>
</associatedParty>
```

creator

This element is found at this location (XPath):
/eml:eml/dataset/creator

<creator> The creator is considered to be the author of the dataset, i.e. the person(s) responsible for intellectual input into its creation. <**surName>** and <**givenName>** elements are used to build dataset citations, so these should be completed fully for credit to be understandable. For long term data sets, include the name of the LTER Site (using the <**organizationName**>) or role of Site PI (using <**postitionName**>). It should be kept in mind that different approaches taken by sites have led to confusion over how to best search for long term datasets, and searchers frequently default to searches using PI’s last name. Therefore it is a reasonable practice to include more creators rather than fewer, even if it blurs the credit for long term datasets.

Example: creator

```
<creator id="org-1" system="FLS" scope="system">
  <organizationName>Fictitious LTER Site</organizationName>
  <address>
     <deliveryPoint>Department for Ecology</deliveryPoint>
     <deliveryPoint>Fictitious State University</deliveryPoint>
     <deliveryPoint>PO Box 111111</deliveryPoint>
     <city>Ficity</city>
     <administrativeArea>FI</administrativeArea>
     <postalCode>11111-1111</postalCode>
  </address>
  <phone phonetype="voice">(999) 999-9999</phone>
  <electronicMailAddress>fsu.contact@fi.univ.edu</electronicMailAddress>
  <onlineUrl>http://www.fsu.edu/</onlineUrl>
</creator>
<creator id="pos-1" system="FLS" scope="system">
  <positionName>FLS Lead PI</positionName>
  <address>
     <deliveryPoint>Department for Ecology</deliveryPoint>
     <deliveryPoint>Fictitious State University</deliveryPoint>
     <deliveryPoint>PO Box 111111</deliveryPoint>
     <city>Ficity</city>
     <administrativeArea>FI</administrativeArea>
     <postalCode>11111-1111</postalCode>
  </address>
```

```
  <phone phonetype="voice">(999) 999-9999</phone>
  <electronicMailAddress>fsu.leadPI@fi.univ.edu</electronicMailAddress>
  <onlineUrl>http://www.fsu.edu/</onlineUrl>
</creator>
<creator id="pers-1" system="FLS" scope="system">
  <individualName>
    <salutation>Dr.</salutation>
    <givenName>Joe</givenName>
    <givenName>T.</givenName>
    <surName>Ecologist Jr.</surName>
  </individualName>
  <organizationName>FSL LTER</organizationName>
  <address>
    <deliveryPoint>Department for Ecology</deliveryPoint>
    <deliveryPoint>Fictitious State University</deliveryPoint>
    <deliveryPoint>PO Box 111111</deliveryPoint>
    <city>Ficity</city>
    <administrativeArea>FI</administrativeArea>
    <postalCode>11111-1111</postalCode>
  </address>
  <phone phonetype="voice">(999) 999-9999</phone>
  <electronicMailAddress>jecologist@fi.univ.edu</electronicMailAddress>
  <onlineUrl>http://www.fsu.edu/~jecologist</onlineUrl>
</creator>
```

metadataProvider

This element is found at this location (XPath):
/eml:eml/dataset/metadataProvider

The <**metadataProvider**> element lists the person or organization responsible for producing or providing the metadata content. For primary data sets generated by LTER sites, the LTER site should typically be listed under <**metadataProvider**> using the <**organizationName**> element. For acquired data sets, where the <**creator**> or <**associatedParty**> are not the same people who produced the metadata content, the actual metadata content provider should be listed instead (see Example below). Complete the <**address**>, <**phone**>, <**electronicMailAddress**>, and <**onlineURL**> elements for each <**metadataProvider**> element.

Example: metadataProvider

```
<metadataProvider>
  <organizationName>Fictitious LTER Site</organizationName>
  <address>
    <deliveryPoint>Department of Ecology</deliveryPoint>
    <deliveryPoint>Fictitious State University</deliveryPoint>
    <deliveryPoint>PO Box 111111</deliveryPoint>
    <city>Ficity</city>
    <administrativeArea>FI</administrativeArea>
    <postalCode>11111-1111</postalCode>
  </address>
  <phone phonetype="voice">(999) 999-9999</phone>
  <electronicMailAddress>fsu@fi.univ.edu</electronicMailAddress>
  <onlineUrl>http://www.fsu.edu/</onlineUrl>
</metadataProvider>
```

contact

This element is found at this location (XPath):
/eml:eml/dataset/contact

A <**contact**> element is required in all EML datasets. Full contact information should be included for the position of data manager or other designated site contact, and should be kept current and independent of personnel changes. If several contacts are listed (e.g. both a data and site manager) all should be kept current. Technicians who performed the work belong under <**associatedParty**> rather than <**contact**>. Complete the <**address**>, <**phone**>, <**electronicMailAddress**>, and <**onlineURL**> elements for the <**contact**> element.

Example: contact

```
<contact id="pos-4">
  <positionName> Information Manager</positionName>
  <address>
    <deliveryPoint>Department for Ecology</deliveryPoint>
    <deliveryPoint>Fictitious State University</deliveryPoint>
    <deliveryPoint>PO Box 111111</deliveryPoint>
```

```
    <city>Ficity</city>
    <administrativeArea>FI</administrativeArea>
    <postalCode>11111-1111</postalCode>
  </address>
  <phone phonetype="voice">(999) 999-9999</phone>
  <electronicMailAddress>fsu.data@fi.univ.edu</electronicMailAddress>
  <onlineUrl>http://www.fsu.edu/</onlineUrl>
</contact>
```

publisher

This element is found at this location (XPath):
/eml:eml/dataset/publisher

The LTER site should be listed as the <**publisher**> of the data set. List the LTER site name, fully spelled out, in the <**organizationName**> element. Complete the <**address**>, <**phone**>, <**electronicMailAddress**>, and <**onlineURL**> elements for each publisher element.

Recommendation for web display of LTER-EML: use <**publisher**> for the organization information

Example: publisher using a reference to a creator from above

```
<publisher>
    <references system="FLS">org-1</references>
</publisher>
```

pubDate

This element is found at this location (XPath):
/eml:eml/dataset/pubDate

The year of public release of data online should be listed as the <**pubDate**> element . The <pubDate> should be updated when data and/or metadata are updated or re released. The format can be either a 4-digit year, or an ISO date (yyyy-mm-dd).

abstract

This element is found at these locations (XPath):
/eml:eml/dataset/abstract
/eml:eml/dataset/project/abstract

In datasets, the abstract element can appear at the resource level or the project level. The <**abstract**> element will be used for full-text searches, and it should be rich with descriptive text. Extensive description should include what, when, and where information, some taxonomic information, as well as whether the dataset is ongoing or completed. Some general methods description is appropriate, and the measured parameters should also be included. For a large number of parameters, use categories instead of listing all parameters (e.g. use the term “nutrients” instead of nitrate, phosphate, calcium, etc.), in combination with the parameters that seem most relevant for searches.

keywordSet and keyword

This element is found at these locations (XPath):
/eml:eml/dataset/keywordSet
/eml:eml/dataset/project/keywordSet

It is recommended that meaningful sets of keywords each be contained within <**keywordSet**> tags. For example, use one <**keywordSet**> for a keyword identifying the LTER site, one for keywords from the LTER controlled vocabulary, one for the LTER core area keywords, etc.. Currently each <**keywordSet**> can include the name of a specific thesaurus (in the optional tag <**keywordThesaurus**>). Keywords that should be included are as many as possible from the LTER controlled vocabulary, at least one LTER core area (if appropriate), the three letter site acronym, some meaningful geographic place names (e.g. state, city, county), network acronym (LTER, ILTER, etc.), organizational affiliation, funding source (i.e. co-funded with other sources, non-LTER funding etc.). In addition to specific keywords, relevant conceptual keywords should also be included.

Example: pubDate, abstract,keywordSet, keyword

```
<pubDate>2000</pubDate>
<abstract>
  <para>Ground arthropods communities are monitored in different habitats in a rapidly changing environment. The arthropods are collected in traps four times a year in ten locations and determined as far as possible to family, genus or species.</para>
</abstract>
<keywordSet>
  <keyword keywordType="place">City</keyword>
```

```
  <keyword keywordType="place">State</keyword>
  <keyword keywordType="place">Region</keyword>
  <keyword keywordType="place">County</keyword>
  <keyword keywordType="theme">FLS</keyword>
  <keyword keywordType="theme">Fictitious LTER Site</keyword>
  <keyword keywordType="theme">LTER</keyword>
  <keyword keywordType="theme">Arthropods</keyword>
  <keyword keywordType="theme">Richness</keyword>
  <keywordThesaurus>FLS site thesaurus</keywordThesaurus>
</keywordSet>
<keywordSet>
  <keyword keywordType="theme">ecology</keyword>
  <keyword keywordType="theme">biodiversity</keyword>
  <keyword keywordType="theme">population dynamics</keyword>
  <keyword keywordType="theme">terrestrial</keyword>
  <keyword keywordType="theme">arthropods</keyword>
  <keyword keywordType="theme">pitfall trap</keyword>
  <keyword keywordType="theme">monitoring</keyword>
  <keyword keywordType="theme">abundance</keyword>
  <keywordThesaurus>LTER controlled vocabulary</keywordThesaurus>
</keywordSet>
<keywordSet>
  <keyword keywordType="theme">populations</keyword>
  <keywordThesaurus>LTER core research areas</keywordThesaurus>
</keywordSet>
```

intellectualRights

This element is found at this location (XPath):
/eml:eml/dataset/intellectualRights

<intellectualRights> should contain site data access policy, plus a description of any deviation from the general access policy specific for this particular dataset (e.g. restricted-access datasets). The timeframe for release should be included as well. For example, LTER Network-wide data should be released on-line within 2-3 years, and if not, the reason needs to be documented in the metadata. (See also LTER Network Data Access Policy: http://www.lternet.edu/data/netpolicy.html [71])

Example: intellectualRights

```
<intellectualRights>
  <section>
    <title>Copyright Notice</title>
    <para> Copyright Board of Regents, Fictitious State University. This dataset is released to the public and may be used for academic, educational, or commercial purposes subject to the following restrictions:</para>
    <para>
      <itemizedlist>
        <listitem>
          <para>While FLS LTER will make every effort possible to control and document the quality of the data it publishes, the data are made available "as is".</para>
        </listitem>
        <listitem>
          <para>FLS LTER cannot assume responsibility for damages resulting from mis-use or mis-interpretation of datasets or from errors or omissions that may exist in the data.</para>
        </listitem>
        <listitem>
          <para>It is considered a matter of professional ethics to acknowledge the work of other scientists that has resulted in data used in subsequent research.</para>
        </listitem>
        <listitem>
          <para>FLS LTER expects that any use of data from this server will be accompanied with the appropriate citations and acknowledgments. </para>
        </listitem>
        <listitem>
          <para>FLS LTER encourages users to contact the original investigator responsible for the data that they are accessing. Where appropriate, researchers whose projects are integrally dependent on FLS LTER data are encouraged to consider collaboration and/or co-authorship with original investigators. </para>
        </listitem>
        <listitem>
          <para>FLS LTER requests that users submit to FLS LTER one copy of any publication resulting from the use of data obtained from this site. </para>
        </listitem>
        <listitem>
          <para>FLS LTER requests that users not redistribute data obtained from this site. However, links or references to this site may be
```

```
freely posted.</para>
        </listitem>
      </itemizedlist>
    </para>
  </section>
</intellectualRights>
```

distribution

This element is found at these locations (XPath):
/eml:eml/dataset/distribution
/eml:eml/dataset/[entity]/physical/distribution

The <**distribution**> element appears at the dataset and entity levels and contains information on how the data described in the EML document can be accessed. The <**distribution**> element has one of three children for describing the location of the resource: <**online**>, <**offline**>, and <**inline**>.

Offline Data: Use the <**offline>** element to describe restricted access data or data that is not available online. The minimum that should be included is the <**mediumName**> tag, if using the <**offline**> element.

Inline Data: The <**inline**> element contains data that is stored directly within the EML document. Data included as text or string will be parsed as XML. If data are not to be parsed, encode them as “CDATA sections,” by surrounding them with “<![CDATA[“ and “]]>” tags.

Online Data: The <**online>** element has two sub elements, <**url**>, and <**onlineDescription**> (optional). <**url**> tags may have an optional attribute named **function**, which may be set to either “download” or “information”. If the URL provides only information about downloading the object but does not directly return the data stream, then the **function** attribute should be set to "information". If accessing the URL directly returns the data stream, then the **function** attribute should be set to "download". If the **function** attribute is omitted, then "download" is implied

An EML dataset should include at least one URL; at a minimum, this should be at the <**dataset**> level (XPath: /eml:eml/dataset/distribution/url), and may point to an application or website. This <**url**> **function** attribute can be set to either “information” or “download”. However, a URL at the entity level (e.g, a dataTable at /eml:eml/dataset/dataTable/physical/distribution/url) should stream data to the requesting application and should include an attribute **function** with the value “download”. In other words, at the entity level, the URL should lead directly to the data and not to a data catalog or intended-use page. For more information about describing a URL connection, see the EML documentation online.

When used at the entity level, an alternative tag is available to <**url**>, <**connection**>. This element is discussed under data entities, below.

As of EML 2.1, there is also an optional <**access**> element in a <**distribution**> tree at the data entity level (/eml:eml/dataset/[entity]/physical/distribution/access). This element is intended specifically for controlling access to the data entity itself. For more information on the <**access**> tree, see above, under the general access discussion.

Data access logging can be implemented by using the LTER Data Access Server (DAS) and its URL-proxy system. For more information about the DAS and how to use it, read the online documentation: [72] http://im.lternet.edu/im_practices/metadata/das [73]

Example: distribution

```
<distribution>
  <online>
    <onlineDescription>f1s-1 Data Web Page</onlineDescription>
    <url function="information">http://www.fsu.edu/lter/data/fls-1.htm</url>
  </online>
</distribution>

<dataTable>
  <physical>
…
    <distribution>
      <online>
        <onlineDescription>f1s-1 Data Web Page</onlineDescription>
        <url function="download">http://www.fsu.edu/lter/data/fls-1.csv</url>
      </online>
    </distribution>
  </physical>
</dataTable>
```

Guidelines for adding keywords for the 5 LTER Core Research Areas

The Executive Board (EB) has mandated that datasets should be searchable according to the 5 LTER Core Research Areas. Sites should include the keywords below to their EML datasets, if applicable. These terms can be added to an existing keywordSet, or may be listed in an independent set.

The 5 terms will be included in the LTER controlled vocabulary (in development), and so putting them in a separate keywordSet would allow them to be linked to that thesaurus in the future.

- The terms to be used to describe LTER core areas are:
  - primary production
  - organic matter
  - disturbance
  - inorganic nutrients
  - populations

Sites may optionally add keywords describing their local areas of research. In this case, site-specific keywords should not be placed in the same keywordSet as the LTER Core Research Area keywords.

Examples:

```
...
<keywordSet>
  <keyword>primary production</keyword>
  <keyword>inorganic nutrients</keyword>
  <keywordThesaurus>LTER Core Research Areas</keywordThesaurus>
</keywordSet>
<keywordSet>
  <keyword>Beach</keyword>
  <keywordThesaurus>Santa Barbara Coastal LTER</keywordThesaurus>
</keywordSet>
...
```

Overview of EML Data Package Completeness (2012) -- DRAFT

Data package completeness determines its potential uses. Below is a summary of EML Metadata content, Data content, and the potential uses of data that are in use at SBC. Asterisks indicate that related issues are being considered and/or tools are being developed at the Network level. The conceptual "LTER Metadata completeness levels (2004)" are included in the last column for comparison. THIS IS A DRAFT (extracted from the SBC IM plan), and has not been reviewed or discussed by the IMC.

Level	EML Metadata content	Data content	Uses
Information	title abstract personnel*, contacts publication date coverage keywords* project description* publisher access and use statements*	Data are optional	1. Information about data only, e.g., LTER Type II data 2. Searches by time, location, taxonomy 3. Data citation Related LTER-2004 levels: “Identification”, “Discovery”
---	---	---	---
Download	*Information* + methods or protocol-link, Data description: includes column names definitions & units* physical description download URL at entity-level*	Data are required, but inconsistencies are allowed	1. Data are available, but the user may need help with interpretation. Related LTER-2004 levels: “Evaluation”, “Access”, “Integration” (with caveats)
---	---	---	---
Integration	*Download* + Metadata congruent with data*	Data are required, and must be clean and congruent with metadata	1. Integration, workflows, further automated processing 2. Query applications 3. Contribute to Network databases Related LTER-2004 level: “Integration” (intended)

Overview of EML Metadata Completeness Levels (2004) -- DEPRECATED

THESE ARE DEPRECATED
Please see this page: http://im.lternet.edu/node/1019 [74] for more information.

The following table summarizes the major levels of EML content “completeness”, or tiers, identified by the two EML working groups. Each level adds more elements from the EML schema to provide a more comprehensive description of the data resources documented by the metadata, and thereby support higher functionality. The table describes Completeness Level, with a Description and Major Elements Added.

1: Identification
Minimum content for adequate data set discovery in a general cataloging system or repository (functionally equivalent to LTER DTOC):
title
creator
contact
publisher
pubDate
keywords
abstract (recommended) dataset/distribution (i.e. url for general dataset information)

2: Discovery
Includes Level 1 content plus coverage information to support targeted searches, adding elements:
geographicCoverage
taxonomicCoverage
temporalCoverage

3: Evaluation
Includes Level 2 content plus data set details to enable end-user evaluation of the methodology and data entities, adding elements:
Intellectual Rights
project
methods
dataTable/entityGroup
dataTable/attributes (see issues outlined in the text)

4: Access
Includes Level 3 content plus data access details to support
automated data retrieval, adding elements:
access
physical

5: Integration
Includes Level 4 content plus complete attribute and quality control details to support computer-assisted data integration and re-sampling, adding elements:
attributeList (full descriptions)
constraint
qualityControl

6: Semantic Use
Includes Level 5 content plus semantic information (currently under development by SEEK, and may require extension to the EML schema)

Documentation : Esri2Eml crosswalk. Migrating Metadata From ESRI or FGDC to EML

This is some documentation on how to use the ESRI to EML crosswalk.

You may find useful some of the tips offered here on how to use the 'esri2eml_for_LTER.xslt' crosswalk. This XSLT stylesheet will allow you to create a EML record out of an XML-record expressed in ESRI - that is, when you use any of the ARC family products (ARC-GIS, ARC-Catalog, ARC-Server, ARCMap etc) you are generating metadata in XML format. This metadata follows the ESRI schema, some version of it, which is based on the Federal Geographic Data Committee schema.

Outline of the documentation:

- Install
    1. Pre-requisites [75]: An XSLT interpreter: Xalan + Xerces
    2. Download the latest version from the LNO CVS, or from the ESRI 2 EML project page [76]or this link [77]
- Dealing with shortcomings of the Esri2Eml crosswalk
    1. Granularity: The crosswalk will store the information that encompasses First Name and Last Name (sometimes multiple authors depends on the orginal document) in one field (XML tag). The crosswalk will throw this information into a creator (or

corresponding party) "last name" (aka surName). You can present this information better by editing the resulting EML -- look for , cut & paste, save.
2. Edit the packageId in the Stylesheet -- you can at least customize the "scope" of the packageID (like knb-lter-sev). You will also have to edit the exact numeric packageId identifier and revision for each resulting EML (knb-lter-sev.32374.1492).
3. Many Esri records at once - You can script some of this, ask the IM community for help, scripts do exist.
4. Adapt the copyright or data usage policies. The style defaults to the LTER Network policies if the originary ESRI document does not specify anything. You may want to see that - look for
5. Check the validity of the document any way you can (XML editor, parser, metacat, etc)
6. Links to the data? check whether it is satisfying
7. It is a Beta tool, chances are it will not always work, and certainly the results will not improve the quality of the original document, furthermore, it is possible that some metadata could be enhanced, this is a helper tool, not an automated process to fix it all.

- ARC-Catalog related tip: You can embed the transformation in Arc-Catalog. In addition to the stylesheet Esri2Eml, you are going to need this additional XML file [78] to embedded in the ArcCatalog
- Support forums, bugs, etc. See this blog [79] about the ESRI10 release and look for related article in databits [80]

Attachment	Size
Converting ArcGIS 10.0 Metadata to EML [81]	30.5 KB
esri2eml.xml [78]	156 bytes

Using the Esri 2 Eml stylesheet

Using the Esri 2 Eml stylesheet with the standalone version of Xalan mentioned is simple:

Just issue this command at the prompt

>xalan -o esri_in_emlFormat_fileName.xml esri_original_record.xml esri2eml_for_lter.xslt

You should see now a new XML file in your folder. Just pick a name more appropriate than "esri_in_emlFormat_fileName".

If you have other XSLT interpreter, a similar procedure will be likely - folow the instructions for your XSLT interpreter, there are way to many ways to do this.

Customizing the Esri 2 EML to your site

You may have a good idea on how to work with the Esri 2 Eml XSLT stylesheet.

You may want also to make a few site-oriented tweaks to it. Meaning, the stylesheet is generic enough, and there is information that applies to all metadata records on your site which will not be entered by the Esri 2 Eml stylesheet as is.

Also, in some circumstances, you may want to enter the general site geographic information (such as description, bounding box coordinates and altitude) or general project-description info, including grant numbers, etc.

Here we explain how to tweak the stylesheet for a few of these site-specific information placeholders (Follow links in item list above or below). Once you have an idea on how to do this, you can change (hard code) any other information placeholders yourself.

Customize Esri2Eml: Adding Data Policies to your EML

The final example on how to customize your Esri 2 Eml stylesheet shows how to encode your site Data Policies.

You basically have to find the <intellectualRights> element in your resulting EML document, and place there your site Data policies, which probably are a superset of the Network Data Policies.

Other option is to adopt the LTER Network Data Policies. In such case, you are in luck. The Esri 2 Eml contains the official LTER Network Data Policies. Those are commented out, meaning, you need to uncomment them if you want to implement them. See the screenshot below, and further comments within the Esri 2 Eml stylesheet.

Customize Esri2Eml: Adding a package ID to your EML

One example of the Esri2Eml site customization is the case of the packageId.

The EML records need a *packageId* to be schema compliant. Consequently, if you want your EML record to be accepted by the central repository, you need to provide a *packageId*. By default, this esri 2 eml stylesheet assigns ESRI's so-called *MetaID* to the EML's *packageId*. The *MetaID* is a unique identifier assigned to your GIS record by the ARC application. However, this is not your best choice, since the *LTER Best Practices* document suggests that the *packageId* should be composed by three parts:

1. a **scope**, which is a string of this sort: "knb-lter-acr", where *acr* is the three letter acronym for your site.
2. a numeric **identifier**, which is a unique number assign to your EML record (perhaps a project number)
3. a numeric **revision** number, which is a natural number(positive integer) that increases as you provide a new version of your EML

How can we do this? You have a number of options. After you converted you ESRI document, you may edit the resulting EML document, and place an appropriate *packageId* in the very first line of the document. You would replace the default value of the XML attribute 'packageId' with the value you need for your document. Other options are editing the original Esri XML record, and add an XML element called packageId directly under the root XML element. The Esri 2 Eml will pick the value contained by that special element and it will place this in the EML document. Finally, you may want to take advantage of some automated process using a Perl script or so, we will discuss this later.

These options are included as comments in the Esri 2 Eml stylesheet - Here is an screeshot 

Customize Esri2Eml: Adding access rules to your EML

Another example on what info in the Esri 2 Eml stylesheet can be customized for your site needs is the EML record access rules.

One EML-specific element absent in Esri is the <access> element, which specifies the access rules (permissions) over the EML record. The access rules specified therein are used by the Metacat repository to give access (read/write/delete) to you or the public at large. Usually, your site will specify full permissions to all your metatada records, and will grant reading permissions to the public.

A generic Esri 2 Eml stylesheet can only provide simple rules like "read permission to everyone", but in order to encode ownership rules to your site for the resulting EML docs, you will need some work.

Here we show some options to customize the access rules.

You can edit by hand the resulting EML. You may want to add similar rules you have for other EML documents you created for your site. This option is a bit time consuming compared to the two next options.

You can encode (hard code) your site rules in the stylesheet. Specifically, you want to ADD a group within the group
The following screen shot shows the section of the Esri 2 Eml that encodes the document access rules in EML.

What you want to do is to add the following group under the group, specifically, above the <allow> element (or tag)

<allow>
<principal>uid=**ACR**,o=lter,dc=ecoinformatics,dc=org</principal>
<permission>all</permission>
</allow>

NOTE: **ACR** is your LTER site acronym.

You may have notice that there is a 'choose' statement. The first 'choice' is connected with yet another customizing choice.

If you were to add the EML-like access rules to the original Esri XML document, this Esri 2 Eml stylesheet will simply translate them. This may be an attractive option when used in conjuction with a Perl parser that prepares the Esri record for multiple issues (granularity + customization). More on the Perl option later.

Differences between the Esri and the Fgdc standards

Here we highlight some of the differences between the ESRI and the FGDC. In a nutshell, ESRI is a superset of FGDC. ESRI is like the FGDC in steriods. Officially, the Esri schema is referred to as the Esri Profile, which is an extension of the FGDC standard. Other significant extension of the FGDC standard is the Biological Data Profile, which extends the geofgraphic-oriented metadata to biological metadata such as taxonomy classifications and methodologies used to collect the data or conduct the experiment.

The elements added by the ESRI Profile are intended to allow a datasets, maps, or other related documents properties to be automatically harvested and recorded in the metadata, to provide information in terms that are more familiar and relevant to ESRI software users, and to allow metadata records to contain graphics and files that describe the dataset. The extended elements also allow improved documentation of items that aren't specifically addressed by the FGDC standard. Overall, this profile helps ESRI to accomplish the goal of making metadata easier to generate, maintain, and use.

The Esri schema can be downloaded here [82].

The FGDC schema files can be also downloaded in the main project page, at the end.

Here are a series of differences.

1. At the top level of the hierarchy, FGDC contains 7 elements corresponding to seven categories:
   Citation Info, Data Quality, Spatial Domain info, Spatial Reference Info, Table information, Distribution Information and Metadata info.
   ESRI provides, in addition to these seven categories, two more categories. These additional informational groups are the Esri

group, and the Binary group. Both are used to tie in the ESRI management systems. Here is a graphical view of the Esri (left) vs FGDC (right) top level differences

2. ESRI - FGDC Differences at each top category :
   - **The Metadata Record Identification Group: <idinfo>**

     Esri has an additional information placeholder to describe Native Data Set enviroment form and sync status.
     Citation/spatial domain
     This ESRI group has more options to describe altitudes and bounding boxes.
     Citation/browse
     This Esri group has one more option to describe a high-level 'browse' form of the data record
     Citation/crossref
     This Esri group has also one more option than the FGDC counterpart.

     Here are these differences graphically: 

   - **The Data Quality group <dataqual>**

     These elements are the same in both the Esri and Fgdc standards

   - **The Spatial Domain Information Group <spdoinfo>**

     Esri has three fundamental expansions for this group.
     Geometrical network information
     Esri-only element to describe relationships between *features*
     Esri Terms
     Properties of objects described using Esri terminology
     Raster Information
     Types and number of raster objects are described in much more details in Esri

     Here are these differences graphically: 

3. **The Spatial Reference <spref>**

   These elements are almost the same in both the Esri and Fgdc standards
   There is one extra element in Esri to allow store a user-defined reference of a coordinate system name. Also, there are a few more named projection systems in Esri.
   

4. **The Data Table information <eainfo> distribution info <distinfo> and the Metadata record information <metainfo>**

   These top-hierarchy groups elements are almost the same in both standards, and the differences are not substantially important for the EML mapping.

Attachment	Size
esriSchema.xsd [83]	95.38 KB

Documentation: Metadata Database

We released a SQL script [84], a mysql workbench schema and a PDF of the corresponding database schema diagram.

Here is a bit of explanation. After several years of working with many LTER sites, we found a common metadata content pattern. The database designed here is intended to provide a platform to contain these common content pattern across LTER sites, and some other ecological data providers.

We used mysql workbench [85] database design tool, which is free for everybody. The tool allows us to graphically design the schema, as well as generate the SQL script that allows us to deploy the database.

The metadata database instance has four main areas. The bottom two areas, "people" and "basic metadata" refer to the metadata content that enables data discovery -the who, when, where and what-. The top two areas, "data file" and "attribute" verse about the details of the actual data containers, whether spreadsheets, tabular web-views of some data-database. Each area contains a number of tables, and it is customizable to the particulars of any ecological metadata provider. For example, the "People" area, where we describe the person, contact information and people-dataset roles, can be used to plug your own personnel directory, or a centralized network directory (or database). There is only one hook from that area to the rest.

A novice user may feel overwhelmed after seeing a total of about 2 dozen tables. In reality, there are a number of tables that appear a number of times in different areas - the reason is to avoid loops (circular relations) in the database schema. Circular relations may jeopardize the integrity of the database. In particular, we repeat the tables concerning "methods" (data acquisition methodologies) and those tables that refer to the geo-temporal context. It is noteworthy, that a dataset may contain many data file (or entities, such as several text files, spreadsheets and the like), and each data file may have a different geo-temporal scope and methodologies, hence the

distinction at the different levels of the hierarchy. Likewise with the more granular area that refers to the individual measurement (i.e; a column in a datafile) -- we included geo-temporal content holders (tables) at that level (database area: "attribute").

How do you populate (migrating the metadata content) the database? we wrote a Perl script that scans rich, quality Ecological Metadata Language documents, and place the content and relationships in this database.

How do you create a web-interface? You can use many programming languages to produce dynamic driven websites about your data. We like to use a content manager to integrate this data with the rest of the website -- this simplifies all aspects of information management.

Documentation: Converting a SPSS Codebook to EML attributeList

Steps for creating metadata for an SPSS file (SPSS v.17, 2009-08-05):

open file in SPSS (or better PASW as it is called now)

setting up the output:

- go to Utilities - OMS control panel
- in the lower panes of the panel click on 'Tables' under 'Output Types'
- on 'Codebook' under 'Command Identifiers'
- on 'Variable Information' under 'Table Subtypes for Selected Commands
- click the radio button for 'File' under 'Output Destination'
- click on the button 'Options' (above the Output Destination)
- in the Options window click the radio button for 'All dimensions in a single row'
  'Output XML' in the pulldown on the top is default
- click 'Continue'
- click 'Browse' next to the File textbox (below the option button again)
- find the folder to save the file and give it a name
- click the 'Add' button above the options button
- click 'ok' and 'ok' again in the confirmation window.

The PASW Statistics Viewer opens with this script:
OMS
/SELECT TABLES
/IF COMMANDS=['Codebook'] SUBTYPES=['Variable Information']
/DESTINATION FORMAT=OXML IMAGES=NO
OUTFILE='C:\path\to\file.xml'
/COLUMNS SEQUENCE=[RALL CALL LALL].

Defining what to output:

- In the PASW Statistics Viewer go to Analyze - Report - Codebook
- in the Codebook output window move all variables from the left to the right side
- click on the output tab and if desiered unclick some of the options (i.e. position and missing value are not necessary in EML)
- click on the statistics tab and unclick all statistics output
- click 'ok'
- now the codebook appears in the PASW Statistics Viewer with all variables on the left side. Make sure they all have an open book icon to be exported.
- In the main window is a nicely formatted codebook with all information for each variable.

Starting the actual output:

- In the PASW Statistics Viewer go back to Utilities - OMS control panel
- higlight the Request and click on the 'End' button.
- click 'ok' and 'ok' again in the confirmation window.

The XML file should now be in the specified location.

Use the attached XSLT stylesheet to convert the data into an attributeList in EML format which can be inserted into the EML file for the dataset. (change extension from .txt to .xsl before use)

Attachment	Size
spss2eml.txt [86]	10.12 KB

EML News

August, 2011
EML Best Practices for LTER Sites, version 2, released.

http://im.lternet.edu/node/910 [87]

July 23, 2010
The LTER Network office maintains a permanent link to the EML schema at:
This may be used in all EML documents in the root element's xsi:schemaLocation attribute, e.g.,
<eml:eml xmlns:eml="eml://ecoinformatics.org/eml-2.1.0"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:ds="eml://ecoinformatics.org/dataset-2.1.0"
xmlns:stmml="http://www.xml-cml.org/schema/stmml-1.1"
xsi:schemaLocation="eml://ecoinformatics.org/eml-2.1.0 https://nis.lternet.edu/eml-2.1.0/eml.xsd" system="knb" scope="system"
packageId="knb-lter-fls.1.1">

June 6, 2010
The EML Best Practices WG has been reconstituted to create the next version of the Best Practices document. The draft should be available by the September meeting at KBS. The WG page:
http://intranet.lternet.edu/im/news/committees/working_groups/emlbestpra... [88]

May 13, 2009
EML 2.1.0 has been released. You can download it at:
http://knb.ecoinformatics.org/software/eml/ [89]
View the README link below for a list of changes between EML 2.0.1 and EML 2.1.0 instance documents. Contact eml-dev@ecoinformatics.org [90] for more information.

EML Versions and Validation

KNB provides an **easy online web form to validate any EML doc**: http://knb.ecoinformatics.org/emlparser/ [91]

- Just click 'Browse', navigate to your EML doc on your local filesystem, and click 'Parse Document'.
- Or you can copy-and-paste the entire EML doc into the text window.
- You will get two results: Validation and Parsing.

oXygen can do validation for you. Parsing does further steps such as testing the uniqueness of id's and connections between id's and references such as with customUnits. **Metacat uses the same parser as that validation tool.**

Recommended Steps:
Since error messages are often terse and not clear at first, and because (just as with xhtml and the W3C checker) often one error fixed just makes the next error show up.

1. Open the EML doc in oXygen
2. be sure the xsd is accessible, ie http://nis.lternet.edu/schemas/EML/eml-2.1.0/eml.xsd [92]
3. notice any red marks
4. starting at the top, hover over the red squiggles and a hint will appear at the bottom of the editor.
5. Or, open a new line and type one < and oXygen will present a menu of tags that can go there.
6. Once you get a green square instead of red marks,...
7. http://knb.ecoinformatics.org/emlparser/ [91]
8. if it passes you're done. If you get an error message, return to oxygen to fix it. (Usually an id or a unit.)

If your site's EML is generated from a system or a script, the next step is to refactor that system so its output EML has the fixes you have identified:

1. Implement the identified fixes in script
2. Re-generate the EML
3. Examine the EML as above

Now that EML doc is ready to be harvested or manually updated or inserted.

1. How to set up a harvest: Metacat Harvester [93]
2. Some examples to help interpret a harvest report error: Categorizing Harvest Error Messages [94]

EML Version History:

EML 2.1.0

Released: August 2008
...
schemaLocation
xsi:schemaLocation="eml://ecoinformatics.org/eml-2.1.0 http://nis.lternet.edu/schemas/EML/eml-2.1.0/eml.xsd"
Major changes from v2.0.1

1. No empty tags
2. valid schema
3. access moves above dataset
4. datetime to dateTime

5. methods plural uniformly

EML 2.0.0/2.0.1

Released 2004
...
schemaLocation
...
Validation notes
...

xsl stylesheet to upgrade EML 2.0.1 to EML 2.1.0

Included with the svn checkout of EML, **eml201to210.xsl** at

- https://code.ecoinformatics.org/code/eml/trunk/style/ [95]

is an xsl stylesheet transform which will automate upgrading*. This was written May 2009 by Jing Tao, one of the Ecoinformatics programmers on the EML development team. Along with it is a **README** file. Both that xsl and README file are attached to this page, (using the most recent version as of Mon, 18 Apr 2011).

Right-click the xsl and README files and Save-As to your local machine.
After downloading, re-name of the README to README.txt so it will open in a text editor.

You can run the transform either on the command line (see README) or with oXygen. oXygen comes with the tool built in. (Document -> Transformation -> Configure Transformation Scenario then Apply Transformation Scenario.)

In addition to the instructions in the README, you will want to **make sure the root element contains an accessible xsd** in the output file. The transform does not substitute that part of the root element. Without a reachable xsd, oXygen will not find errors.

xsi:schemaLocation="eml://ecoinformatics.org/eml-2.1.0 http://nis.lternet.edu/schemas/EML/eml-2.1.0/eml.xsd"

The schemaLocation part on the left it will do for you; but the second half, the http link to the eml.xsd, you will need to replace.
If someone adds that to this xsl please attach the new version to this page.

*The transform cannot fill in missing fields. EML 2.0.1 allows some fields to be empty that EML 2.1.0 does not.
After transforming, fill in any empty fields.
Also there are other upgrades you may wish to do in addition to this transform. It will not move data download urls for you, or remove literalLayout. Somebody want to add those to this stylesheet?? (hint)

EML to ISO19139

GBIF has developed an XSLT to transform our profile of EML 2.1.0 to ISO19139. It may provide a start for a more comprehensive stylesheet. See here:

http://rs.gbif.org/schema/eml-gbif-profile/dev/ [96]

The XSL file is:

http://rs.gbif.org/schema/eml-gbif-profile/dev/eml2iso19139.xsl [97]

Example EML or Harvest Error Messages

Harvest log reports may contain errors in four categories:

- (1) xml not well-formed
- (2) xml not matching eml schema (eml not schema-valid)
- (3) parser errors with references (missing id, or missing customUnit, or duplicate id).
- (4) harvest issues (duplicate revision, ie a docId already submitted, or the url to the eml file is not-found)

The KNB emlParser will find any errors in categories 1, 2 or 3. An xml editor alone will not alert you to category 3 errors but once you have the error message from the parser you can then locate the problem id. The category of the error will determine your approach. So it is useful to recognize which category of error has happened.

Attached is a pdf with a table of example error messages. The error messages are gathered from a metacat harvest report and three diagnostic tools (oXygen, the KNB emlParser, and the pasta qualityChecks in evaluation mode).

example EML doc	docType	Intentional error	oXygen red squiggles	KNB emlparser	Metacat 2.0.0 errorMessage	qualityChecker suggestion

Here are a batch of eml docs which **are all identical except for one intentional mistake in each.** Two are base cases, completely valid and correct (knb-test-fls.1.1 for eml 2.1.0 and knb-test-fls.11.1 for eml 2.0.1.) In some cases I made the same intentional mistake in those two versions of eml to demonstrate how some mistakes were allowed in the older version.

The example eml docs are provided in a zip, along with the corresponding harvest list and metacat harvest report. View them in oXygen. When these eml docs are viewed in the oXygen [98] eml editor, the mistakes are marked with a red squiggle as well as a message at the bottom of the editor pane.

The KNB parser Validation service for EML [91] will validate that your EML file conforms to the rules in the EML spec. It also runs a SAX validation parse to check if your document is schema valid according to the EML schemas. Notice how often the error message from the validation service matches that from the harvest report. Whereas one has to wait for a harvest, the KNB parser tool gives instant feedback.

The pasta qualityChecker can evaluate a whole harvest list in evaluation mode and provides some diagnostic information. (Start at portal.lternet.edu/nis/tools.jsp [99], login, then choose 'Harvest Data Packages' and scroll to bottom of page. Enter harvest list URL and select 'evaluate'.) I only included the 'suggestion' column here; each eml doc gets its own quality report, nicely formatted. It is still a prototype but is already useful. Please be aware that, as a development installation, it can be unavailable at times and does not have the capacity of a production server.

There can be more than one cause for the same error message. A given error message, can be caused more than one way. For example, in knb-test-fls.4.1, that same message could be the result of the entityDescription appearing above the entityName rather than the entityName actually being missing. This is another reason why looking at **validation with an xml editor is going to provide more specific information.** It may be the same message but the red squiggle will happen on a different line. Another way of saying this is that a given error in the eml will always produce the same message but a given message may be caused by more than one error in the eml.

The .zip file would not attach as a .zip so I renamed it .zip_.txt so after download, remove the _.txt extension.

Attachment	Size
Example_EML_Error_Messgaes.pdf [100]	132.09 KB
Example_EML_Error_Messages.zip_.txt (remove _.txt from file name) [101]	110.76 KB

How-To (Videos)

This section contains clips and FAQs that may help you solve specific problems related with metadata, including EML, Harvesting, managing and maintaining issues.

Getting Started With EML - using oXygen to create EML docs

Sometimes we get the question, how do I get started with EML?

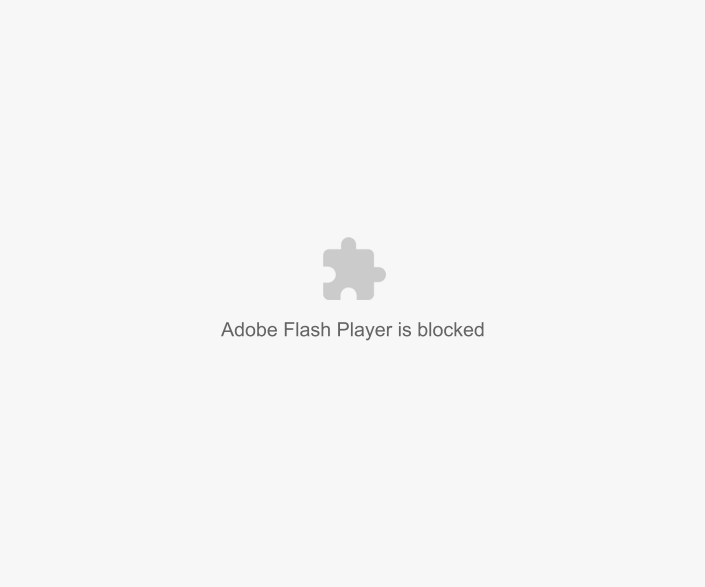
Well, this question is quite generic, and I would urge the inquirer to clarify the question, to narrow it, if possible.

We use EML as a vehicle to share ecological metadata. EML has the unique advantage of being a common specification within the ecological community, a *lingua franca* to share metadata.
EML is mostly for machine consumption, as it is XML based. To read EML documents, it is best to transform the content to make the content (ecological metadata) appealing to the human reader.

The means to produce EML depend on whether you already have metadata in some electronic, structured format or not.

If you where to start from scratch (you do not have any electronic-based structured metadata for your dataset) you can opt to use an XML oriented editor such as oXygen. (yes, you can use notepad, morpho
and others too). Just have your metadata handy when you start!

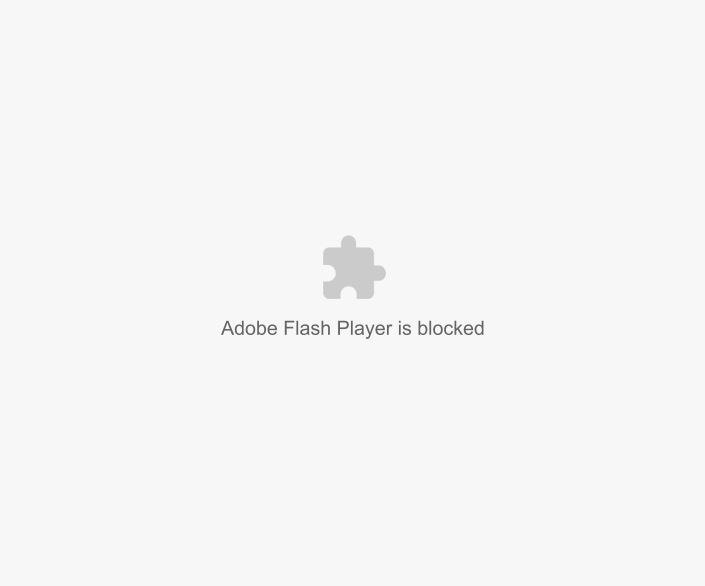
Here is a video that shows how would you use oXygen with the EML schema.



Getting Started With Morpho

We thought you would like to see how would you use the EML oriented editor, Morpho, to create a new EML record. We made a 13 min video that walks you from the where is Morpho, to the basics on how to use it.

Here it is



Importing an EML document into Morpho

Here we show how to import and EML document into Morpho.

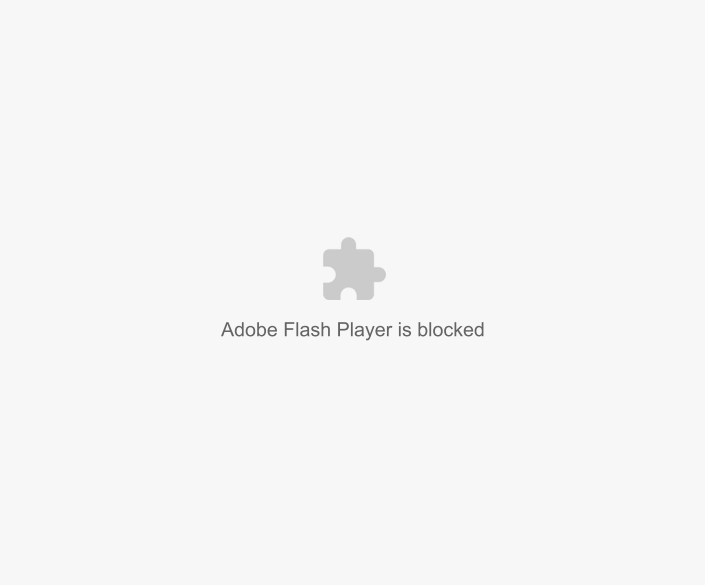
Issue: Sometimes you may be interested on importing an EML document that you created without using Morpho into Morpho.

You just need two things:

1) Name your EML (the file name) using only a X.Y.xml convention, where X is an integer that represents the numeric identifier from the EML packageID. and Y is the version from the packageId.

2) Place the document into the ~/.morpho/profileName/data/scopeFromPackageId/ folder or directory, where the profileName is your Morpho user profile, and the *scopeFromPackageId* is the scope string of your EML, which, for LTER sites is probably "knb-lter-acr" (where acr is your site acronym). If this folder does not exist, just create it.

Now it should be there. See this video to clarify



Converting List of Taxa from ITIS to EML

Issue: You have a list of taxa, and you want this list in EML format, so you can insert it into your existing EML document

The ITIS (www.itis.gov) offers a webform to upload a flat text file with a list of taxa (say Latin Names). The list is compared to its database, and you can choose to get the corresponding taxonomic classification returned as XML complying with the rules of the Biological Data Profile (FGDC) XML schema. In turn, you can change that trivially to an EML-compliant taxonomic classification, since there is a 1-t-1 correspondence. All you have to do, is replace the BDP tags by the corresponding EML tags.

FGDC/BDP EML

You can do this replacement with a "FInd&Replace" function of most editors, or if you wish to do so, you can use the Perl script attached below that has those regexps.

See the video, explains this process visually.

Attachment	Size
itis2eml.pl_.txt [102]	791 bytes

ISO Date and Time formats

Attached below is ISO 8601:2004(E)
This is the international date standard for representing dates and times.

The wikipedia article is als useful:
http://en.wikipedia.org/wiki/ISO_8601 [103]

Attachment	Size
ISO_8601-2004_E.pdf [104]	293.22 KB

ITIS-to-EML XSL Stylesheet

The attached XSL file (I had to rename it to .txt to get it to upload) can be used to convert ITIS web service XML to an EML 'taxonomicCoverage' entity. It is meant to be used with the output from the ITIS 'getFullHierarchyFromTSN' service, which you can find here:

http://www.itis.gov/ws_hierApiDescription.html#getFullHierarchy [105]

It will work with both the SOAP and REST web service outputs. There are a couple things to note:

1) No common names yet. Those are from a different service call, so it would require a separate service layer to pull them both into one EML chunk. It would also be possible to create an EML 'taxonomicSystem' section using this services, but again it would require pulling together multiple ITIS calls.

2) The REST output is not fully valid XML, because it lacks declarations for some of its namespaces. The stylesheet is written to work around this problem, but if you're using the REST output in other ways, be aware. I contacted ITIS about this issue, but according to one of the software engineers, "the REST response is generated by the web service library we're using. Currently, that's Apache Axis2, which gives us the RESTful output 'for free' as part of the SOAP web service. All I do is provide the data, Axis2 takes care of formatting and responses. So, the short answer is that I don't believe there is any way for me to change the REST format at this time without replacing Axis2 with custom code."

Hope this helps you!

Attachment	Size
ITISToEML.txt [106]	1.35 KB

Metacat Harvester

Documentation on how to harvest your site's metadata into the LTER Data Catalog at the LTER Network Office can be found here:

See also Chapter 7.2 (page 84) of harvesterAdministration.html: Configuring a Harvest Site [107]. Two databits articles have been written on this topic: EML Harvesting I [108] and EML Harvesting II [109] written in 2004 and 2005. And for example error messages see Example EML or Harvest Error Messages [94].

Basic Info:

1. Start at the **metacat homepage**: http://metacat.lternet.edu/das/lter/ [110]
2. **Log in as your site*.** Do not log in as yourself.
   This is important. You will see different links and pages.
3. Click on **'Metacat Harvester'** link (at left)

* If you need help logging in as your site, ask an EML Mentor, or the network IM, or tech support. This form is intended only for site IMs to use.

Now you will see a form with these fields:

- **Email Address:** Harvest reports will be emailed to this address
- **Harvest List URL:** The URL to your site's harvest list
- **Harvest Frequency:** (days / weeks / months) Harvests will run once every this number (1-99) of days, weeks, or months

Your site likely already has a method of maintaining the harvest list. The harvest list is an xml document listing which datasets are to be inserted or updated into metacat from your site. See attachment 'example_harvestList.xml'. Here is a screenshot of the first doc in the harvest list:

```
1  <?xml version="1.0" encoding="UTF-8" ?> <hrv:harvestList xmlns:hrv="eml://ecoinformatics.org/harvestList">
2  <document>
3      <docid>
4          <scope>knb-lter-fls</scope>
5          <identifier>101</identifier>
6          <revision>13</revision>
7      </docid>
8      <documentType>eml://ecoinformatics.org/eml-2.0.1</documentType>
9      <documentURL>http://fls.lternet.edu/~janedoe/eml/knb-lter-fls.101.13.xml</documentURL>
10 </document>
11 <document>
```

Some tips:

- **Updates:** A new revision will be harvested to update an existing data package if the docId already in metacat has a lower revision. For example knb-lter-fls.101.12 gets updated with knb-lter-fls.101.13.
- **Inserts:** If there is no docId with knb-lter-fls.101 of any revision already in metacat, then the eml will be Inserted as a new dataset.
- **leading zeros in docIds:** Do not use leading zeroes in the identifier part of the package id.
  - Bad: knb-lter-fls.0052.3
  - Good: knb-lter-fls.52.3
- **docId, packageId, and DAS url should match** It is important that the docId under which you are inserting or updating a data package match the identifier in the packageId (in the root eml element).

```
<?xml version="1.0" encoding="UTF-8"?>
<eml:eml xmlns:eml="eml://ecoinformatics.org/eml-2.1.0"
  xmlns:stmml="http://www.xml-cml.org/schema/stmml-1.1"
    xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
    xsi:schemaLocation="eml://ecoinformatics.org/eml-2.1.0 http://nis.lternet.edu/schemas/EML/eml-2.1.0/eml.xsd"
    packageId="knb-lter-fls.101.13" system="knb" scope="system">
```

If you use the DAS, that the identifier there also match.

```
<distribution>
  <online>
    <url function="download">
      http://metacat.lternet.edu/das/dataAccessServlet?docid=knb-lter-fls.101.13&amp;urlTail=MySitesAltIDForThisPkg
    </url>
  </online>
</distribution>
```

(It is good practice, but not essential, to name the eml documents by either their metacat docId or your site's alternatIdentifier, even though the harvester will accept an eml doc with any filename.)

Attachment	Size
example_harvestList.xml [111]	1.02 KB
eml_root_element_example_knb-lter-fls.101.13.png [112]	29.97 KB
example_DAS_url_knb-lter-fls.101.13.png [113]	18.86 KB
top_of_example_harvestList.png [114]	33.11 KB

Metadata Editors

Description of and help for metadata editors

Morpho

Info for Morpho

Importing an EML document into the Morpho editor

Here we show how to import and EML document into Morpho.

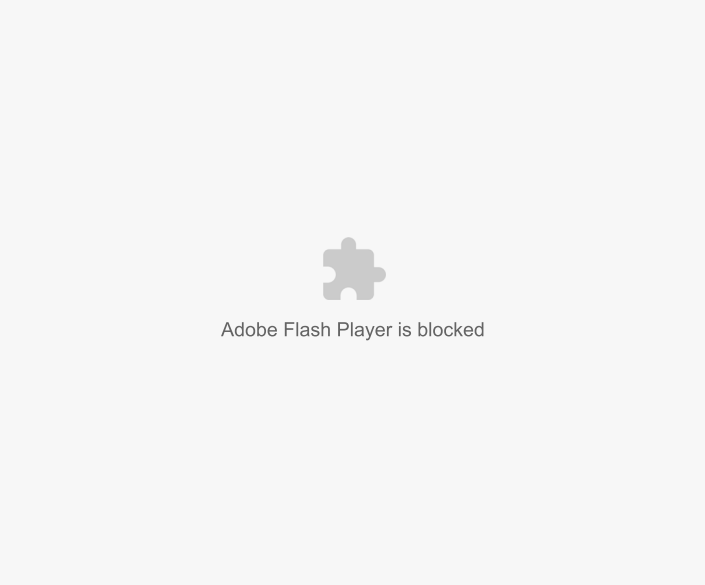
Issue: Sometimes you may be interested on importing an EML document that you created without using Morpho into Morpho.

You just need two things:

1) Name your EML (the file name) using only a X.Y.xml convention, where X is an integer that represents the numeric identifier from the EML packageID. and Y is the version from the packageId.

2) Place the document into the ~/.morpho/profileName/data/scopeFromPackageId/ folder or directory, where the profileName is your Morpho user profile, and the *scopeFromPackageId* is the scope string of your EML, which, for LTER sites is probably knb-lter-acr (acr is your site acronym). If this folder does not exist, just create it.

Now it should be there. See this video to clarify



oXygen EML Editor

oXygen is far more efficient and effective than a plain text editor for examining EML at any stage of development.

1. Get the program:
Download oXygen EDITOR from this page: http://www.oxygenxml.com/download.html [115]
As of 2012, you click on the logo with a blue box with a red 'X'

2. Get the license:
The oXygen XML Editor is free to LTER sites as part of their "Support Life Program": http://www.oxygenxml.com/support_life.html [98]

In the 'Describe your activity' box, give the URL to your site and say that you use oXygen for ecological research, and with open source software. Individual sites have put their site's research gpal in that space as well. Scroll down the page to see some examples. Within a couple days, they will email you a license.

3. Validation:
When an eml doc has an accessible .xsd specified in its root element, oXygen can validate it.
Ask for help if you need it.

4. Other links:
(Add links to other pages on this site such as the intro video.)

Metadata Library

This page is a collection of recent papers relating to metadata management.

Ontology Papers:
Joshua Madin, Shawn Bowers, Mark Schildhauer, Sergeui Krivov, Deanna Pennington, Ferdinando Villa. 2007. An ontology for describing and synthesizing ecological observation data. Ecol. Informatics, doi:10.1016/j.ecoinf.2007.05.004.
A Data Bits Good Read: http://intranet.lternet.edu/archives/documents/Newsletters/DataBits/07fa... [116]

Joshua S. Madin, Shawn Bowers, Mark P. Schildhauer and Matthew B. Jones. 2008. Advancing ecological research with ontologies. Trends in Ecology and Evolution. doi: 10.1016/j.tree.2007.11.007

The Luquillo Legacy Metadata Standardization Project Documentation

The starting point of this project is the creation metadata inventory for the Luquillo LTER

1. A careful and manual process for metadata control checks after metadata was handed by the principal investigator to the data manager. This process is conducted by John, Gabriella, and Eda.
2. A web-friendly version of the metadata was placed for public examination, with all details related to each particular data and metadata set.
3. High level metadata content was stored in a relational database (Paradox instance). The unnormalized database schema had no constraints, and it was used

to produce some custom views and reports to help the Luquillo investigators.

Based on this, the Luquillo Information manager and the LNO assistant decided that the first short term goal was to contribute to the LTER data catalog with *Discovery Level Metadata*.

A longer term goal involved transforming the existing, richer metadata, into the EML specification. This involved quite a bit of quality control, manual, time-consuming work (hence the denomitation as longer term). The details for this step extend the details of the previous implementation, and we can organized them as here:

- Workflow and scripts involved
- How to analyse and review the results, log files.
- Fixing problems, examples
- Dealing with EML versions and the harvest list

Luquillo EML implementation QA/QC Examples

Here we walk through a couple of examples for diagnosing and fixing EML resulting from the Standardization workflow.

First Example: we will understand why a *discovery* level EML document (aka level 3 EML) was produced in lieu of a content-rich EML document (EML Level 5, *attribute level* EML).

We ran the perl script. Apparently, the EML document looks fine - it is also EML-schema compliant- that is *valid* EML. However, we are surprised to see a relatively lean document - none of the attribute level EML is reflected in the produced document. What happened? We know we have information about the data table, variables, etc.

When we have content-rich metadata at the source level, but EML level 3 is produced, the most likely cause is a mismatch in the actual data filenames. Our original source of information comes from the comma delimited file *filesTab* for the particular data set. For example, the contents of *lterdb127-filesTab.csv* for the Luquillo dataset number 127, a **meteorological** dataset, look like this:

```
Data File No., Data File Identifier, On-Line Filename, Starting Date, Periodicity of sample, End Period
1, Hourly data 2001 - 2004, EVFSTowerMetHourly2001-2004.txt, 2001, hourly, 2004
2, Hourly data 2005 - 2008, EVFSTowerMetHourly2005-2008.txt, 2005, hourly, 2008
3, Daily (24 h) data, EVFSTowerMetDaily.txt, 2000, daily, 2008
```

Here we see that this dataset contains three data files named "EVFSTowerMetDaily.txt" and " EVFSTowerMetDaily2001-2004.txt", etc.

Independently, we created a file with a list of data file names and header lines. This is a relevant excerpt of this file:

```
127, EVFSTowerMetHourly20012004.txt, ARRAY, DATE_MMDDYY, YEAR, DAY, TIME, PPFD_MICROMOLES/M2, PPFD_MILLIMOLES/M2/HOUR, PPDFMAX,
PPDFMAXTIME, PPDFMAXSEC, SOLARRAD_WATTS/M2, SOLARRAD_KJOULES/M2/DAY, SOLRADMAX,SOLRADTIME, SOLRADSEC, TEMP_DEGREES_CELSIUS,
RELHUM_percent, WINDMAX, WINDMAXTIME, WINDSPEEDAVER_M/S, WINDDIR_DEGREES, SDWINDDIR, RAIN_MM,WINDROSE1-45_M/S, WINDROSE46-90_M/S,
WINDROSE91-135_M/S, WINDROSE136-180_M/S, WINDROSE181-225_M/S, WINDROSE226-270_M/S, WINDROSE271-315_M/S, WINDROSE316-360_M/S
127,EVFSTowerMetHourly20052008.txt,ARRAY,DATE_MMDDYY,YEAR,DAY,TIME,PPFD_MICROMO[.....]
```

At first sight, all may look OK, but after close inspection of the file names, we see that a hyphen is missing. We have EVFSTowerMetDaily20012004.txt instead of EVFSTowerMetDaily2001-2004.txt. This small mismatch produced a scaled back version of the EML metadata, in an attempt to avoid inconsistencies. We solved this by introducing a hyphen on the list, then a level 5- EML document [117] was produced.

Similar mismatches in the variable labels (also known as the "attributeLabel" in EML; the data file column labels) result into problems in the EML outcome. This is perhaps the most common problem encountered. This added check provides a good consistency check for the content in the data file, and the descriptive content in the source (original) metadata. Of course, problems of other nature may pass undetected by these mechanisms.

Second Example: Mismatch on the attribute problem. This is, as mentioned earlier, the most common problem that occurs when we generate EML from the metadata sources. Sometimes the mismatch is a typo, but other times the mismatch points to a deeper problem, such as a missing column description. In this example, we show a column label mismatch caused by a difference in the granularity of the information.

Luquillo's 137th dataset *Community Composition of Litter Invertebrate* uses a single data table (spreadsheet) to store the data. The data file columns are described by the following comma delimited descriptor:

```
Abbreviation (Col. Label), Taxon, Classification, State, Mm, Year_code/forest_code/palmtype_code
Name of Variable, Species name, Classification, State, Length, "99,00,01 = year; T, C, D = tabonuco, palo Colorado, dwarf forest; P, NP
= palm, non-palm (columns 5-76)"
[...]
```

Reviewing the resulting EMLs

Here is where we describe the manual quality control and assurance process for the Luquillo metadata standardization project.

The workflow described before outlined the one-directional flow of the standardization process. However, there are many points in the process that may introduce problems. Here we describe how we detect and solve some of these problems, and how we create a feedback mechanism for our described workflow.

Some of the XML files created in the standardization process may not be EML-schema compliant. Usually, these bad EML documents are a result of one of these two reasons:

1. Problems with the code or XSLT (bugs)
2. Source of metadata non-conforming with the template expectations

It is our turn to manually debug the problems for these files, and also, to verify the content of the EML-schema compliant generated documents. We used oXygen and XMLspy to manually inspect the files.

Manual inspection of the files requires quite a bit of time, as well as good organization. We need:

1. A sequential and systematic inspection of all files. Evaluate the content level per quantity. Evaluate the quality -- do the information placeholders contain relevant metadata? are there empty XML elements? Do URL conduct to actual data files, broken links or generic catalogs?
2. A process to document all problems and solutions provided in the reviewing process
3. Access to source metadata files to trace problems on the original metadata
4. Access to the intermidiate XML-generated files, to help with the debugging of both the Perl and the XSLT code.
5. Open communication with the dataset contact people and data information manager to clarify or complement metadata.
6. Access to the actual data files, to verify the content and some obvious aspects of the metadata

See some practical examples, a walk through the QA/QC process.

Data Access Server

The Data Access Server (DAS) is a model for abstracting access to LTER generated data behind a proxy server that provides key services in support of the LTER Data Policy, including authentication, auditing, and notification.

- Detailed documentation [118]
- Instructions for registering with the DAS [119]

Update: The plans described below were implemented and the DAS is available for use by sites. Instructions on how to use the DAS for your site are described in the attached powerpoint (Feb 2009) and pdf (Oct 2008).

Introduction

The LTER Network has invested considerable time, effort, and funding into the collection of scientific data. Access and use of this data is formalized through the end user's acceptance of the LTER Network Data Access Policy, Data Access Requirements, and the General Data Use Agreement (http://www.lternet.edu/data/netpolicy.html [71]), which were approved by the LTER Network Coordinating Committee on 6 April 2005. Motivation behind these policies and agreements is driven by the need to document the flow of data from the LTER Network out to the community to validate broader impacts of the LTER program. As such, the LTER Network has adopted a "standard" for data access and use that needs to be implemented into both local and network-wide computing infrastructure. This standard requires that the end user provide basic identifying information, including name, affiliation, email address, and contact information, in an electronic format that can be provided to the data owner. Further, acknowledgment and acceptance of either the General Public Use Agreement and or the Restricted Data Use Agreement applied to a data set, and a statement of the intended use of the LTER data, will be recorded prior to the release of any LTER data.

Approach

To support sites in their compliance with LTER data policies, the DAS project will provide sites with the ability to replace “direct link” data URLs with proxy URLs, which will route all data requests through an authentication, auditing, and notification service called the Data Access Server (DAS). The DAS will support five primary objectives that have been defined through analysis of use case scenarios:

1. end user registration and acceptance of the LTER data agreements, including authentication;
2. URL management;
3. notification of data access to the Data Set Contact;
4. auditing of all data access; and
5. reporting.

The basic notion of the DAS is that sites may register a data URL expression into the DAS URL management interface and replace the data URL with a similar representation using a proxy URL. The goal of the DAS is to route end users (including non-interactive applications) through the DAS so that access to LTER data can be logged and the appropriate contact be notified about the access event. The DAS strives to use as much existing LTER cyberinfrastructure as reasonable, including:

1. the LTER Data Catalog servlet infrastructure to support the integration of DAS components;
2. Ecological Metadata Language documents in the LTER Data Catalog to identify appropriate contact information for notification and to ascertain data table entity names; and

3. the LTER LDAP for end user contact information and authentication (authentication will also use LDAP databases of our affiliate networks, such as NCEAS and PISCO).

A graphical representation of the DAS network topology is shown in Figure 1.

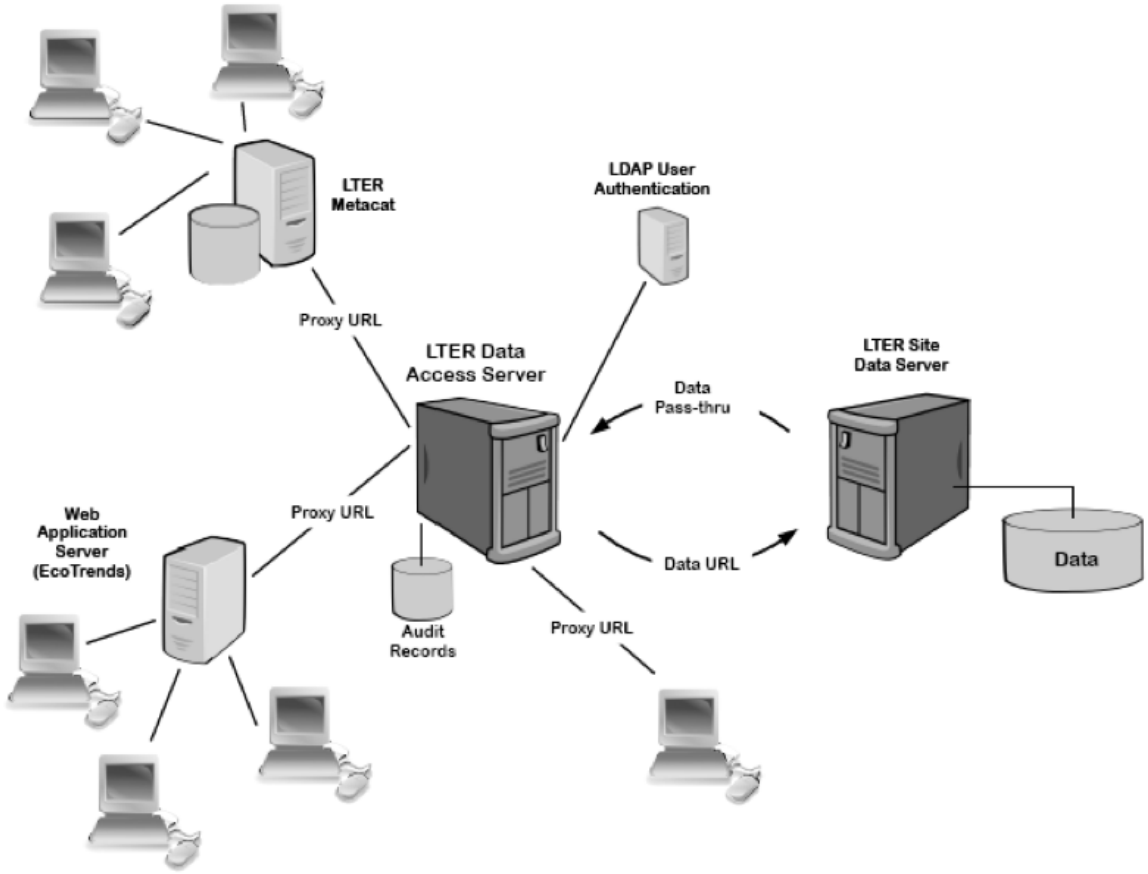


Figure 1: Hypothetical network topology of the Data Access Server.

The LTER NIS development team has identified a general model for a Network-wide LTER Data Access Policy implementation strategy called the Data Access Server (DAS). The DAS model proposes a centralized NIS service that would be integrated into the LTER Data Catalog infrastructure (see Figure 2). This approach takes advantage of common tasks already implemented within the Data Catalog, such as authentication and audit logging. The DAS service components would add five areas of functionality to the LTER Data Catalog as noted above.

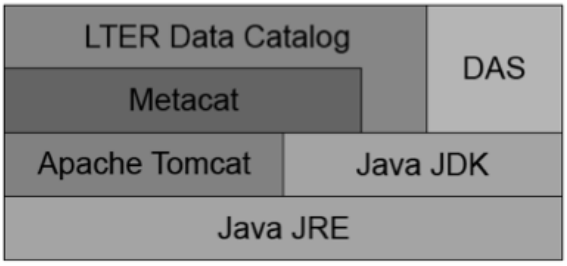


Figure 2: Relationship of the Data Access Server software with other components.

End user registration and authentication will take place through the standard LTER Data Catalog user interface. Registration will include forms to accept the LTER Data Access Policy and an area to enter the intended use of the LTER data. Once registered and authenticated, users may accept a web-browser cookie that will allow seamless access to data for future requests.

The pass through process will rely on the replacement of the URL that references site data with a "proxy" URL that points instead to the DAS. This approach requires the site to register their data URL with the DAS so that a one-to-one correspondence between the data URL and the proxy URL can be determined by the DAS service. The proxy URL is used in lieu of the actual data URL within any LTER metadata document (including EML) that is published for public viewing. When an end user wishes to download data by selecting the online distribution URL in the metadata document, they would be directed to the DAS first and have their credentials validated, before a data stream is returned on the site's behalf. If the end user has not registered at this point, they would be directed to the appropriate registration interface. If they have already registered and there exists a token (e.g., cookie) on their workstation, they would be provided the data without restriction. Otherwise, the end user would be directed to a log-in interface prior to receiving any data.

Any download event from through the DAS invoked by the end user will send an email notification to the data owner/provider of the data download event, along with the end user's contact information and the name of the downloaded data set. In addition, the DAS will also send a notification to the end user with the data owner/provider's contact information, the General Use Agreement, and any special Restricted Data Use Agreement for the specific data set that was downloaded.

The DAS will also log all event information into its audit log table; this process will be an addendum to the the logging that already is part of the routine of the Data Catalog.

The DAS will provide periodic reports of all data access events to LTER sites, in addition to supporting an interface that will allow site information managers to perform interactive queries for specific events.

Advantages

1. The DAS model does not require sites to participate or change their current practice of providing direct access to their data. It is a model that may be utilized at the site's convenience, perhaps addressing sensitive or high-profile data first.
2. Since the DAS would run as a centralized service (potentially distributed) at the LTER Network Office, tools and enhancements based on the DAS model would be available to all participating sites, including data access reports that can be perused directly by NSF officials. This can be an effective method for standing groups like the Information Manager Executive committee or the LTER Executive Board to analyze LTER data access through a single interface.
3. The DAS model fits nicely within the current LTER LDAP user registry used by the LTER Metacat for user identification. Other Metacat sites (and their users) would not have to conform to the LTER Data Access Policy, but their users would have to register with the DAS before being allowed access to LTER data.

Disadvantages

1. The current DAS proof-of-concept relies on the use of HTTP cookies for identifying registered users. Cookies, when enabled by the web browser, are sent automatically with each client request to the web server (in this case, the DAS). Any other application (e.g., Kepler or MatLab) that could not send a cookie would automatically fail the user identification process and not be allowed access to LTER data. A more robust method for providing generic identification would have to be identified.
2. The DAS model requires sites to change their data access URLs within their EML documents and/or any data references that would be bound by the LTER Data Access Policy.
3. A new registration interface would be required to collect the necessary Data Access Policy information. This would require users to submit new information into the DAS, even those who are already registered in the LTER LDAP.

Conclusion

The DAS model is one method for sites to easily conform to the LTER Data Access Policy. A fully functioning DAS implementation is expected in early 2009. The LTER NIS Development Team welcomes all comments and suggestions for improving this model, and anticipates working closely with beta-sites to evaluate and test the DAS model. For more detailed information, please refer to the DAS Project Plan (http://urban.lternet.edu:80/viewvc/trunk/projects/DAS/docs/DAS-ProjectPlan.pdf?view=co&root=NIS).

For additional information – Mark Servilla, LTER Network Office (servilla@lternet.edu)

Attachment	Size
dataAccessServer.pdf [118]	382.96 KB
DAS-WaterCooler.ppt [120]	2.4 MB
DAS_Fig_1.png [121]	118.75 KB
DAS_fig_2.png [122]	31.33 KB

Web Sites

Please see new LTER web design guidelines here [123]

Data Warehouses

Data warehousing practices

ClimDB/HydroDB

The Climate and Hydrologic Database Project is a component of the LTER Network Information System (NIS), the suite of LTER intersite database modules being created to promote synthetic ecological research. The ClimDB and associated HydroDB were developed in response to research scientist need for current and comparable climate and hydrologic data summaries for LTER sites and U.S. Forest Service Experimental Watersheds widely used in intersite comparisons, modeling studies, and land management-related studies (Henshaw et al. 1998). ClimDB/HydroDB is composed of a harvester system that continually captures and updates data from multiple

sites into a central database, as well as a web interface that allows graphical display and download of data in common formats (Baker et al. 2000).

Participant Page - http://climhy.lternet.edu/harvest/harvest.htm
Public Data Access Page - http://climhy.lternet.edu/

Sites can trigger a harvest as a scheduled event or through a program by using the following URL:
<code>http://climhy.lternet.edu/harvest.pl?module=<#>&site=<ltersite></code>

where,
<#> = URL option number (1 or 2)
<ltersite> = Three letter LTER Site code (i.e., AND for Andrews)

Two harvest URL's are maintained in the database for each site (option 1 and option 2), either URL can be used for any data set or scripting program (option 3 is reserved for USGS data harvest only). These URLs can be edited from the participant page under "Update Metadata" and then "Research Area Information". Sites employing real-time USGS Gauging Station measurements can have this data regularly harvested (LTER Network News, Fall 2003). Please contact the ClimHy Data Manager.

Sample Handling, Voucher Specimens, Collections

Sample Handling, Voucher Specimens, Collections practices

Communication Technology

Communication Technology Practices

Scheduling VTC protocol

Video Teleconferencing (VTC) is increasingly being used to hold meetings, share information, and basically improve communication among sites and scientists within the LTER. In 2007, the LTER Network Office (LNO) sent two computer-mounted PolyCom units to each LTER site with the suggestion that one unit should be assigned to the information manager at the site because of the frequent need for interaction within this group (see email from Bob Waide from 7/3/2007 attached below). The sites used PVX Licenses provided by LNO to activate the service at each site. The information management committee (IMC) members use this technology frequently including monthly virtual water coolers (VWC - informal information sharing), monthly meetings of the IM-Exec group, and opportunistic working group meetings. However, the PolyCom is also used widely outside of the IMC and scheduling conflicts are inevitable.

Since initiation in 2007, James Williams has been the LNO's PolyCom contact person. While James Brunt serves as a backup. When using the LNO PolyCom hub for a meeting, the call is initiated from the LNO. These initiations begin about 10 minutes before the call begins. This allows time for participants to get connected and work out any issues without taking up meeting time.

Following are some basic protocols that should be used when scheduling a meeting to use the PolyCom system.

(1) The LNO must confirm that the scheduled date and time for a PolyCom meeting are available.
For VWC, it's recommended to give the James' a list of dates and times for the coming year. This is easy since there is a regular pattern for these meetings. For other meetings, typically a Doodle Calendar is set up and the LNO PolyCom contacts should always be cc'd. This keeps them in the loop and enables them to see who needs to be connected for that particular meeting.

(2) Send LNO an anticipated list of meeting participants.
This can be via a Doodle Calendar (as described above) or just an email list of names. This helps the LNO set up the PolyCom calls.

(3) Send out a meeting reminder.
It's always a good idea to send a meeting reminder to the participants and to the LNO PolyCom contact a day or two before the meeting.

Attachment	Size
LTER-im PolyCom units sne to sites.txt [124]	1.24 KB

This material is based upon work supported by the National Science Foundation under Cooperative Agreement #DEB-0236154. Any opinions, findings, conclusions, or recommendations expressed in the material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

[8] http://im.lternet.edu/links/goto/475/89/links_related
[9] http://im.lternet.edu/links/goto/475/88/links_related
[10] http://im.lternet.edu/links/goto/475/87/links_related
[11] http://im.lternet.edu/links/goto/475/84/links_related
[12] http://im.lternet.edu/links/goto/475/83/links_related
[13] http://im.lternet.edu/links/goto/475/90/links_related
[14] http://im.lternet.edu/sites/im.lternet.edu/files/bibliography_guidelines.pdf
[15] http://im.lternet.edu/sites/im.lternet.edu/files/emlCitation2endnoteText_xsl.txt
[16] http://www.research.gov/common/robohelp/public/WebHelp/Project_Reports.htm
[17] https://verbosus.com/bibtex-style-examples.html
[18] http://demo.research.gov
[19] http://im.lternet.edu/links/goto/8/170/links_related
[20] http://im.lternet.edu/links/goto/8/169/links_related
[21] http://intranet2.lternet.edu/content/develop-lter-nis-best-practices-designing-and-writing-workflow-scripts-pasta-framework
[22] http://mtsms.unm.edu/Mediasite/Catalog/Full/cf020feb9c6a4222a57d8daa334f7e3521/b671822aa4c04e6caa5d2157a55da6e514/cf020feb9c6a4222a57d8daa334f7e3521/?state=GHrbcctw5uPTuUXj0ogE
[23] http://im.lternet.edu/im_practices/data_management/nis_workflows/PASTAprog
[24] http://im.lternet.edu/sites/im.lternet.edu/files/NISdataworkflowsbestpractices0.2.pdf
[25] http://im.lternet.edu/sites/im.lternet.edu/files/NISdataworkflowsbestpractices0.1.pdf
[26] http://www.vcrlter.virginia.edu/webservice/PASTAprog/
[27] http://www.vcrlter.virginia.edu/webservice/PASTAprog/knb-lter-vcr.26.14.r
[28] http://www.vcrlter.virginia.edu/webservice/PASTAprog/myprog.sas?emlurl=http://myserver.edu/myEMLdoc.xml
[29] http://myserver.edu/myEMLdoc.xml
[30] http://gce-lter.marsci.uga.edu/public/xsl/toolbox/EMLdataset2mfile.xsl
[31] mailto:wsheldon@lternet.edu
[32] https://gce-svn.marsci.uga.edu/trac/GCE_Toolbox
[33] http://svn.lternet.edu
[34] http://im.lternet.edu/sites/im.lternet.edu/files/im_results_05.pdf
[35] http://wiki.esipfed.org/index.php/EnviroSensing_Cluster
[36] http://im.lternet.edu/959
[37] http://www.campbellsci.com/
[38] http://www.onsetcomp.com/
[39] http://www.vistadatavision.com/
[40] http://www.ysieconet.com/
[41] http://nexsens.com
[42] http://gce-lter.marsci.uga.edu/public/im/tools/data_toolbox.htm
[43] http://im.lternet.edu/node/1007
[44] http://his.cuahsi.org/
[45] http://im.lternet.edu/node/1003
[46] http://www.dataturbine.org
[47] http://im.lternet.edu/node/1010
[48] https://kepler-project.org/
[49] http://im.lternet.edu/node/1014
[50] http://im.lternet.edu/node/1005
[51] http://im.lternet.edu/node/1015
[52] http://im.lternet.edu/node/1016
[53] http://www.r-project.org/
[54] http://im.lternet.edu/node/1013
[55] http://intranet.lternet.edu/im/files/im/Beginning_Google_Maps_API.pdf
[56] http://im.lternet.edu/links/goto/9/161/links_related
[57] http://im.lternet.edu/links/goto/9/111/links_related
[58] https://environmentaldatainitiative.org/resources/five-phases-of-data-publishing/phase-3/metadata-best-practices/
[59] http://im.lternet.edu/VocabBestPractices
[60] http://im.lternet.edu/links/goto/80/161/links_related
[61] http://en.wikipedia.org/wiki/Level_of_measurement
[62] http://im.lternet.edu/sites/im.lternet.edu/files/emlHandbook.pdf
[63] http://im.lternet.edu/sites/im.lternet.edu/files/emlbestpractices-2.0-FINAL-20110801_0.pdf
[64] http://im.lternet.edu/sites/im.lternet.edu/files/example_full_a.xml
[65] http://im.lternet.edu/sites/im.lternet.edu/files/example_provenance.xml
[66] http://im.lternet.edu/sites/im.lternet.edu/files/example_lter_maps.xml
[67] http://www.fgdc.gov/metadata/geospatial-metadata-tools
[68] http://intranet.lternet.edu/im/project/Esri2Eml/docs
[69] http://intranet.lternet.edu/im/project/bdp2eml
[70] http://nis.lternet.edu/schemas/EML/eml-2.1.0/eml.xsd"
[71] http://www.lternet.edu/data/netpolicy.html
[72] http://www.google.com/url?q=http%3A%2F%2Fintranet.lternet.edu%2Fim%2Fim_practices%2Fmetadata%2Fdas&sa=D&sntz=1&usg=AFQjCNF5qpXdlir91CpUJAK7phLSymWTrQ
[73] http://im.lternet.edu/im_practices/metadata/das
[74] http://im.lternet.edu/node/1019
[75] http://intranet.lternet.edu/im/project/Esri2Eml/docs/preRequisites
[76] http://intranet.lternet.edu/im/project/Esri2Eml
[77] http://intranet.lternet.edu/im/project/Esri2Eml/currentRelease
[78] http://im.lternet.edu/sites/im.lternet.edu/files/esri2eml.xml
[79] http://lno.lternet.edu/node/325
[80] http://databits.lternet.edu
[81] http://im.lternet.edu/sites/im.lternet.edu/files/esri2eml_how.doc
[82] http://intranet.lternet.edu/im/files/im/esriSchema.xsd
[83] http://im.lternet.edu/sites/im.lternet.edu/files/esriSchema.xsd
[84] http://intranet.lternet.edu/im/node/429
[85] http://dev.mysql.com/downloads/workbench
[86] http://im.lternet.edu/sites/im.lternet.edu/files/spss2eml.txt
[87] http://im.lternet.edu/node/910
[88] http://intranet.lternet.edu/im/news/committees/working_groups/emlbestpractices
[89] http://knb.ecoinformatics.org/software/eml/
[90] mailto:eml-dev@ecoinformatics.org
[91] http://knb.ecoinformatics.org/emlparser/
[92] http://nis.lternet.edu/schemas/EML/eml-2.1.0/eml.xsd

[93] http://im.lternet.edu/node/418
[94] http://im.lternet.edu/node/1024
[95] https://code.ecoinformatics.org/code/eml/trunk/style/
[96] http://rs.gbif.org/schema/eml-gbif-profile/dev/
[97] http://rs.gbif.org/schema/eml-gbif-profile/dev/eml2iso19139.xsl
[98] http://www.oxygenxml.com/support_life.html
[99] https://portal.lternet.edu/nis/tools.jsp
[100] http://im.lternet.edu/sites/im.lternet.edu/files/Example_EML_Error_Messgaes.pdf
[101] http://im.lternet.edu/sites/im.lternet.edu/files/Example_EML_Error_Messages.zip__0.txt
[102] http://im.lternet.edu/sites/im.lternet.edu/files/itis2eml.pl_.txt
[103] http://en.wikipedia.org/wiki/ISO_8601
[104] http://im.lternet.edu/sites/im.lternet.edu/files/ISO_8601-2004_E.pdf
[105] http://www.itis.gov/ws_hierApiDescription.html#getFullHierarchy
[106] http://im.lternet.edu/sites/im.lternet.edu/files/ITISToEML.txt
[107] http://knb.ecoinformatics.org/software/metacat/MetacatAdministratorGuide.pdf
[108] http://databits.lternet.edu/fall-2004/eml-harvesting-i-metacat-harvester-overview-and-management
[109] http://databits.lternet.edu/spring-2005/eml-harvesting-ii-preparing-site-metadata-and-harvest-lists
[110] http://metacat.lternet.edu/das/lter/
[111] http://im.lternet.edu/sites/im.lternet.edu/files/example_harvestList.xml
[112] http://im.lternet.edu/sites/im.lternet.edu/files/eml_root_element_example_knb-lter-fls.101.13.png
[113] http://im.lternet.edu/sites/im.lternet.edu/files/example_DAS_url_knb-lter-fls.101.13.png
[114] http://im.lternet.edu/sites/im.lternet.edu/files/top_of_example_harvestList.png
[115] http://www.oxygenxml.com/download.html
[116] http://intranet.lternet.edu/archives/documents/Newsletters/DataBits/07fall/#gr3
[117] http://intranet.lternet.edu/im/files/im/LUQMetadata127.xml_.txt
[118] http://im.lternet.edu/sites/im.lternet.edu/files/dataAccessServer.pdf
[119] https://nis.lternet.edu/NIS/?q=node/60
[120] http://im.lternet.edu/sites/im.lternet.edu/files/DAS-WaterCooler_0.ppt
[121] http://im.lternet.edu/sites/im.lternet.edu/files/DAS_Fig_1.png
[122] http://im.lternet.edu/sites/im.lternet.edu/files/DAS_fig_2.png
[123] http://intranet.lternet.edu/im/im_requirements/webdesign_guidelines
[124] http://im.lternet.edu/sites/im.lternet.edu/files/LTER-im PolyCom units sne to sites.txt