The Importance of Long-Term Ecological Information Management

An LTER Information Manager’s Perspective

Linda Powell
Florida Coastal Everglades (FCE) LTER
Talk Objectives

I. FCE LTER Site Level IMS Perspective
   - FCE LTER Program Quick Facts
   - FCE IMS Development
   - Data Management Policy Elements
   - FCE IMS System Details
   - FCE IMS Program Management
   - FCE IMS Metadata

II. LTER Network Level IMS Perspective
   - Trends
   - Trends Data Store (PASTA)
LTER Site Level IMS Perspective
Florida Coastal Everglades LTER

- FCE LTER founded in May 2000
- FCE research team comprised of 128 members
  - 53 Researchers
  - 8 Postdoctoral Researchers
  - 19 Ph.D. Students
  - 15 Master’s Students
  - 6 Undergrad Students
The majority of the 23 FCE LTER Research Sites are located in freshwater Marsh, Estuarine Mangroves and Seagrass Estuary ecosystems in Everglades National Park.
FCE Information Management System (IMS) Team

FCE IMS Team consists of two full-time managers: Linda Powell, Information Manager and Mike Rugge, Project Manager.

- Linda has a M.S. in geology and has a strong background in GIS, remote sensing and system administration.

- Mike has a M.S. in plant biology and specializes in web development and GIS.
FCE IMS Development

Possible Problems

1) New site comprised of many well established independent laboratories.
   • How to set program procedures that all parties can comply with?
   • How to keep track of program’s research?
2) No real IMS guidance from the LTER network other than research ‘Data’ must be made available via the web.
An IMS was *important* to the FCE LTER for two main reasons:

1) Program Management

2) Data Management

*Our program’s success had a direct correlation to being highly organized.*
One of the FCE LTER Program’s first tasks was to establish an IMS, placing an emphasis on:

1) An IMS *Mission Statement*

2) A set of IMS *Goals*

3) A strong IMS *Data Management Policy*
FCE IMS Development

Mission

• To *facilitate* the site's scientific work

• To *ensure the integrity* of the information and databases resulting from the site's coastal Everglades ecosystem research
FCE IMS Development

Goals

1) *Design and implement an IMS* to handle research contributions from several large FCE research laboratories

2) *Collect and Archive* both FCE and historical Everglades data

3) *Provide Comprehensive metadata* for data interpretation and analysis

4) *Design and implement tools* that facilitate data management, data discovery and data access

5) *Contribute to LTER network* informatic activities
Florida Coastal Everglades LTER

FCE IMS Data Management Policy Elements
FCE IMS *Data Policy Elements*

A Comprehensive *Data Management Policy* greatly facilitates FCE Information Management

• Policy Includes 8 components:
  • FCE IMS Objectives
  • FCE Research Project Management
  • Data Collection Protocols
  • Quality Assurance Procedures
  • Data Organization Description
  • Data Archive and Protection Procedures
  • Data Distribution Information
  • Data User Agreement
FCE IMS Data Policy Elements

Why are each of our Data Policy Elements important?

1) FCE IMS Objectives
   • Clear statement about the IMS goals

2) FCE Research Project Management
   • Can track types and locations of program research being conducted
FCE IMS *Data Policy Elements*

*Why are each of our Data Policy Elements important?*

3) **Data Collection Protocols**
   - **Sets deadlines for when ‘Data’ must be submitted to FCE IMS**
   - **Protects ‘Data’ as information resides in the FCE system AND in individual labs**

4) **Quality Assurance Procedures**
   - **Sets Site-wide Quality Assurance**
Why are each of our Data Policy Elements important?

5) Data Organization Description
   • Sets Data file naming conventions and describes FCE metadata requirements

6) Data Archive and Protection Procedures
   • Sets archive & protection procedures
   • Establishes a disaster recovery plan
FCE IMS Data Policy Elements

Why are each of our Data Policy Elements important?

7) Data Distribution Information
   • Sets Data ‘type’ descriptions and establishes distribution guidelines

8) Data User Agreement
   • Gives Researcher ‘Data’ use information
   • Allows IMS to collect ‘Data’ use statistics
Florida Coastal Everglades LTER

FCE IMS System Details
The FCE IMS is a **HYBRID** system of hierarchical flat files and a robust database.

- Data and metadata files from individual research studies are stored in a hierarchical flat file directory system.
- The **Oracle 10g database** currently holds all metadata and project management information and will hold ‘Data’ values in near future.
FCE IMS *System Details*

‘Project’ & ‘Data’ management information is ‘Cross-referenced’ in the Oracle 10g database

• Web-based tools facilitate program and data management
Florida Coastal Everglades LTER

FCE IMS Web-based Tools for Program Management
The FCE Data Management Policy includes research project reporting procedures whereby

- Researchers must submit FCE LTER Project information via a ‘project information’ form upon notification of their project funding and certainly NO later than 6 months after notification.
The collection of FCE project information has allowed for the development of a web-based project management tool.
The Everglades land margin ecosystem in southwest Florida represents a combination of different mangrove ecological types in mainland carbonate environments with gradients in amount of nutrients, hydroperiod and salinity. Resource competition and stress theory will be utilized to understand mangrove community development due to these shifts in nutrient pools and hydroperiod across the coastal gradients of the South Florida mangrove ecosys. These ecological properties are important because potential changes in nutrient and hydrology in the inland watershed may be important impacts on the structure and function of mangroves in the coastal margin of southwest Florida. Initial studies and preliminary modelling efforts of these specific properties of mangroves in the Shark River estuary and those in the Taylor Slough Region will be used to test specific hypotheses in the LTER study. The biogeochemical properties of mangroves are the least understood of ecological processes along the transition from upland to coastal margin ecosystems. Thus the specific nature as to how the distribution of nutrients influences mangrove structure and productivity, and the role of mangroves on the fate of nutrients in sub-tropical estuaries are poorly understood. Continued efforts monitoring these biogeochemical processes along with synoptic studies of productivity, together with further development of ecological models, will provide insights on the response of mangroves to changes in water management of subtropical coastal watersheds.
Project datasets with active links to data summary page & data downloads.
FCE IMS *Project Management*

Sampling Site Locations with active *links* to research site information

**Project Sites**
Search Results give info on a Sampling Attribute's:

- Location
- Project Information
- Datasets

Locations where Porewater Salinity is Collected
Florida Coastal Everglades LTER

FCE IMS Metadata
FCE IMS *Metadata Protocol*

The FCE adheres to the LTER network-wide *Ecological Metadata Language (EML 2.0.1)* standard

• Developed *tool* to collect and convert metadata into EML

FCE Metadata Tool Download:

http://fcelter.fiu.edu/research/information_management/tools/
FCE IMS *Metadata Tool*

**FCE Excel2EML** Metadata Entry and Converter Tool Package

*Two Components:*

- *Excel Spreadsheet*
- *Perl Converter Program*
**FCE IMS EML Metadata Template**

**Step 1:**
Scientists complete Excel Metadata Template

Template Completion = **High level EML**

---

<table>
<thead>
<tr>
<th>Information Requested</th>
<th>Added Information</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>LTER Dataset Information</strong></td>
<td><strong>Template Completion = High level EML</strong></td>
</tr>
<tr>
<td>Dataset Title</td>
<td>Template Completion = High level EML</td>
</tr>
<tr>
<td>Dataset Description</td>
<td>Template Completion = High level EML</td>
</tr>
<tr>
<td>Dataset Keywords</td>
<td>Template Completion = High level EML</td>
</tr>
<tr>
<td>Dataset Contacts</td>
<td>Template Completion = High level EML</td>
</tr>
<tr>
<td>Dataset Abstract</td>
<td>Template Completion = High level EML</td>
</tr>
<tr>
<td>Dataset Notes</td>
<td>Template Completion = High level EML</td>
</tr>
</tbody>
</table>

---

5 Possible Sheets of Information
Step 2: Load Completed Excel Template into Perl Program and Convert to EML
Step 3: Post EML document on the Web & add to LTER Harvest List

FCE EML Conversion Document Produced with Excel2EML Tool
FCE IMS EML XML To HTML

Step 4:
Convert EML XML document into HTML using XSL Style sheet with HTML code & post to the FCE Website.

View Metadata in 3 Formats
Download Data Directly from metadata page
FCE IMS *Metadata Protocol*

EML metadata is extremely important to the FCE IMS

- Provides detailed information on FCE Data, Projects, Attributes and Methods for the Oracle database
- Allows FCE LTER to contribute data information to LTER Network EML database
EML Metadata Protocol

Is compliance with EML cost effective?

• Must retrofit existing metadata into EML or create EML metadata for data with NO metadata- *Lots of man-hours*

• Need to find a tool to generate EML metadata for new and old data- *Every tool seems to have some limitation*
EML Metadata Protocol

Is compliance with EML cost effective?

• In my opinion- YES!
  
  ▪ High level EML enables the LTER network to have a universal ‘standard’ on information about data
  
  ▪ EML facilitates data synthesis across the ecological research community
EML Metadata Protocol

EML has forced the LTER network to look at how well metadata is documented.

Example:

FCE LTER collects surface water salinity but we didn’t describe in the data file’s metadata ‘Methods’ section at what depth in the water column salinity was collected.
EML Metadata Protocol

EML has caused the LTER network Information Managers to address problems when reporting sampling attribute units.

• Many of the LTER site’s local units (i.e. \textit{mg/short shoot/day}) not found in EML, making them generate their own ‘Custom’ units.

• Information Managers developed an \textit{EML Unit Registry} as an aid in developing EML unit descriptions.
Registered Units are searchable and users can get a view of the proper EML XML format.
LTER Network Level IMS Perspective
How an IMS & EML Metadata Facilitates Network Level Science

*Trends* in Long-Term Ecological Data: a multi-agency synthesis project

*(Dr. Debra Peters, Project Leader, Jornada LTER)*
**Trends Project**

The *Trends* project is a collaborative effort among agencies and institutions to make long-term ecological data *easy* to access, analyze, and compare across sites.

- *Easy* ‘Data’ access suggests that the data and metadata *must* be stored in an IMS.
**Trends Project**

**Objectives**

- To *create a platform for synthesis* by producing a compendium of easily accessible long term graphs and data from long-term ecological research sites.

- To *illustrate the utility of this platform* in addressing important within-site and network-level scientific questions.
Trends Products

1) *Folio-sized book* to be published by Oxford Univ. Press

2) *Website* (data, metadata, graphs) for synthesis and analysis
Trends Website

http://trends.sagehost.net/index.php

The Trends Project

The Trends project is a collaborative effort among agencies and institutions to make long-term ecological data easy to access, analyze, and compare across sites. Data from Long Term Ecological Research (LTER) sites funded by the National Science Foundation (NSF), from USDA Forest Service (USDA FS) and Agricultural Research Service (USDA ARS) sites, and from sites operated by universities are being accumulated into a central portal for access.

This website is a portal to:

- a large and diverse collection of derived and current long-term ecological data
- unique data exploration, graphing, and synthesis tools
- information about our large array of collaborating agencies and their individual sites.

These data, tools, and information are freely available to researchers who wish to pursue cross-site synthesis studies or students interested in learning more about resources available to them. Please read our data use and citation policies before downloading data.

In addition, the graphs found in our book, Our Changing World: Insights From Long Term Ecological Research, as well as the data used to generate them, can be easily browsed and viewed here.

Learn more about the background of the Trends project, or browse our site using the buttons to the left.

What’s New?

This website was unveiled to enthusiastic scientists and information managers at the LTER All Scientists’ Meeting (ASM) September 19-24, 2006.

Featured Dataset

Click the image below for a larger view.
**Trends** What they do with the Data

- **Downloading and storing data & documentation**
- **Writing R or SAS scripts to generate:**
  - Datasets containing monthly or annual averages or totals
  - Strict time plots with simple linear regression
  - Tables that record all derived statistics
  - Plots that show change over time among different sites for each variable
  - Anomaly plots of monthly climate data
- **Generating metadata with EML for each derived product. Metadata contains links to original data and associated scripts.**
- **Recording each product (data, metadata, graphs), along with links between products, in a multi-purpose database.**
MULTI-SITE ANALYSES
Nitrate in precipitation

Step 1. Graph similar data through time for sites with those data.
Step 2. Determine trend line by site.
Step 3. Compare slopes of trend lines among sites.

Mean change in total deposition of N in nitrate form in precipitation
How an IMS & EML Metadata Facilitates Network Level Science

Trends Data Store (PASTA)

Mark Servilla
servilla@lternet.edu
The *Trends Data Store* project at LNO automates the *Trends* project data collection and processing steps.

• “The premise of this project is that EML will adequately describe the data set (e.g., entities, attributes, physical characteristics) to allow the capture of distributed data sets into a central SQL database.”
**Trends Data Store**

**Generalized Workflow**

1. Sites collect and document time-series data (e.g., climate, social-economics, …)
2. Sites update EML with a new revision
3. EML is harvested into Metacat
4. EML Loader/Parser loads new/updated dataset into primary database
5. Data integration/transformation converts “raw” data into “derived” data
6. Derived data is stored in secondary database
7. EML is generated for derived data and is stored in Metacat
8. Derived data is made available to *Trends* website
How an IMS & EML Metadata Facilitates Network Level Science

Tools incorporating EML are being developed to automate the data discovery and data synthesis processes:

1) Kepler Project

The Kepler project's overall goal is to produce an open-source scientific workflow system that allows scientists to design scientific workflows and execute them efficiently using emerging Grid-based approaches to distributed computation.

http://www.kepler-project.org/
How an IMS & EML Metadata Facilitates Network Level Science

Tools incorporating EML are being developed to automate the data discovery and data synthesis processes:

1) **EML Statistical Tools** *(J.Porter, VCR LTER)*

   Tools convert EML document to either R, SAS or SPSS programs

   [http://cvs.lternet.edu/cgi-bin/viewcvs.cgi/VCR/eml_statistical_tools/](http://cvs.lternet.edu/cgi-bin/viewcvs.cgi/VCR/eml_statistical_tools/)
Websites of Interest

• http://trends.sagehost.net/index.php
• http://fcelter.fiu.edu
• http://www.kepler-project.org/

Special Thanks to:
Christine Laney, Jornada LTER: Trends Project Information
Mark Servilla, LTER Network Office: Trends Data Store Information