Ecosystem services and observing system design  Examples of LTER and a focus on DATA

John R. Vande Castle
LTER Network Office
Department of Biology
University of New Mexico MSC03 2020
Albuquerque, NM, 87131-0001
jvc@unm.edu
Long Term Ecological Research Network

Celebrating 35 Years of Excellence in Long-Term Ecological Research
Ecosystem Services – Watersheds – habitat to houses
Impacts on Ecosystems Services – Mangrove die off
Manipulation Experiments – Sevilleta Wildlife Refuge – streaming sensor data
96 hr, 15 second averaged sampling, 24hr ~ 100kb

Sevilleba Field Station Sensor Web (see map).

Sensor Webs Data Applet, version 1.2

- **LightFlux** (relative)
  - Pod 2
  - Pod 5
  - Pod 9
  - Pod 11

- **AirTemp** (deg C)
  - Pod 2
  - Pod 5
  - Pod 9
  - Pod 11

- **SoilTemp** (deg C)
  - Pod 2
  - Pod 5
  - Pod 9
  - Pod 11

Open Source Data Turbine (OSDT) Buoy – Moorea Coral Reef and North Temperate Lakes LTER

Here – Coral Reef Observing Network (CREON)
Conventional field data collection – Toolik Lake
Satellite Data – Landsat time series to measure changes in kelp density and distribution

Santa Barbra Coastal LTER
Site-regional opportunities: latitudinal differences in relative sea level rise

M. Alber, Ecotrends

- PIE: 2.6 mm y\(^{-1}\)
- VCR: 3.9 mm y\(^{-1}\)
- GCE: 2.9 mm y\(^{-1}\)
- FCE: 2.2 mm y\(^{-1}\)
Palmer Station LTER - 80% drop in dominant penguin species
Characteristics of Ecological Data

- High Data Volume (per dataset)
- Low Data Volume (per dataset)
- High Complexity
- Low Complexity

- Satellite Images
- Streaming Sensor Data
- GIS
- Weather Stations
- Biodiversity Surveys
- Population Data
- Primary Productivity Data
- Gene Sequences
- Soil Cores
Importance of LTER Information Management

• **Site-based**
  - types of data collected differs strongly between sites
  - diversity of physical and computational environments
  - close association of Information Manager’s and active researchers
  - 1 to 2+ FTEs per site dedicated to IM

• **Focus on Long Term**
  - archival: data preserved on decadal time scales
  - legacy datasets
  - Importance of structured metadata (EML) as requirement
Information Management Committee

• Formal membership comprised of 1 individual representing each of the 26 LTER sites plus other site IM personnel (2-4/site)

• Annual meeting to plan and coordinate Network data management activities (60)
• A sub-group (IMExec) meeting 4 times/year
• IMC Working groups - funded and independent – In-person and VTC meetings
• Bi-monthly “water cooler” VTC

• Activity areas that have broad reaching impact include:
  • Metadata standards for ecology
  • Published guidelines for managing research data
  • Development of Network Information System modules
  • National and international training in information management
  • Sponsoring and participating in symposia and other outreach

• Every third year, the IMC co-hosts the larger Environmental Information Management Conference - EIM 2011 Conference (Sept. 28-29, 2011)
This website is maintained by the community of LTER Information Managers. It provides a hub for communication and information/knowledge exchange that we hope will be useful outside of LTER as well. The information provided represents practices and approaches taken at the different LTER sites. Please feel free to contact the authors for help and further information or to contribute to this site.

- **The News Section** has news and updates
- **In the Home Section** reports and meeting notes from IMexec and NISAC and links to the LTER Network office and IM publications may be accessed
- **The Projects Section** contains information on every project the LTER IMC is currently working on, e.g. materials for upcoming video conferences and minutes, workshop materials, documents.
- **Under Meetings** information about annual IMC meetings, virtual updates and training workshops can be found.
- **The Resources Section** is the core of this site with the IM practices which are 'how-to' and the most up to date best practices of general interest to Information Management. The IM guidelines are LTER specific requirements, evaluation criteria etc.
- **LTER Site Profiles** contain updates on LTER site related IM activities and comparisons of how IM is done at the sites.

---

**LTER Information Management**

The Long-Term Ecological Research Network was established in 1980 by the National Science Foundation. It includes research sites inside and outside the US representing a wide variety of ecosystems and research emphases. At each site an Information Manager or Information Management Team is concerned with technologies that improve collaborations between researchers and technological solutions to enabling data collection, discovery, access, integration and analysis across disciplinary and scale boundaries.
Ecological Metadata Language (EML)

EML is a metadata specification developed by the ecology discipline and for the ecology discipline. It is based on work done by the Ecological Society of America and associated efforts (Michener et al., 1997, Ecological Applications). EML is implemented as a series of XML document types that can be used in a modular and extensible manner to document ecological data. Each EML module is designed to describe one logical part of the total metadata that should be included with any ecological dataset.

Send any comments, errors, or suggestions to eml-dev@ecoinformatics.org or through the EML Bug Tracking system. The preferred way to submit problems with EML or feature requests is the bug tracking system.

EML Version 2.1.0

You can access the EML specification online by reading it in HTML format, or you can download the entire specification, including both the HTML documentation and the XML Schema files.

- EML 2.1.0 Specification — Read it online
  OR
- Download EML
  The download consists of the EML modules, described in the XML Schema language. In addition, the full documentation on the modules is provided in HTML format.

- Changes to EML in version 2.1.0
- EML Frequently Asked Questions (FAQ)
- Validation service for EML
  In addition to the online service found at the previous link, the EML distribution itself contains the validation software for your use (see "lib/runEMLParser" for details on how to run it).

About the EML Project

The EML project is an open source, community oriented project dedicated to providing a high-quality metadata specification for describing data relevant to the ecological discipline. The project is completely comprised of voluntary project members who donate their time and experience in order to advance information management for ecology. Project decisions are made by consensus according to the voting procedures described in the ecoinformatics.org Charter.

We welcome contributions to this work in any form. Individuals who invest substantial amounts of time and make valuable contributions to the development and maintenance of EML (in the opinion of current project members) will be invited to become EML project members according to the rules set forth in the ecoinformatics.org Charter. Contributions can take many forms, including the development of the EML schemas, writing documentation, and helping with maintenance, among others.

Development Information

Developers may be interested in browsing the source code CVS repository that we use in developing EML. This always contains the most recent development version of EML, and therefore may be in flux, or otherwise broken. It is unlikely that it will contain the same files that are in the current release (2.1.0). Use at your own risk. Write access to this repository is reserved for EML project members. We welcome contributions to this work in any form. Contributions can take many forms, including the development of the EML schemas, writing documentation, and helping with maintenance, among others. Non-project members can contribute by submitting their feedback, revisions, fixes, code, or any other contribution through the eml-dev@ecoinformatics.org mailing list, or...
Welcome to the LTER Data Portal

Data are one of the most valuable products of the LTER program. The LTER Network seeks to inform the LTER and broader scientific community by creating well designed and well documented databases and to provide fast, effective, and open access to LTER data via a network-wide information system designed to facilitate data exchange and integration. Currently, the LTER Data Portal contains entries for over 6000 ecological datasets from 26 LTER Network research sites, and thousands of additional datasets from numerous other ecological field stations and research institutions.

The LTER Data Portal includes content from both LTER and non-LTER data sources including PISCO, KNB, etc. By default, search results display only LTER data sources. You may include non-LTER data sources in your search by selecting the check box below.

NEW! When you begin typing in the search box form below, an auto-completion dialog will suggest ranked terms that originate from key-words and titles within the Data Portal.

Search Term:  

Include non-LTER data

Advanced Search

Search Reset

LTER Data Policies

The LTER data policy includes three specific sections designed to express shared network policies regarding the release of LTER data products, user registration for accessing data, and the licensing agreements specifying the conditions for data use.

Other Databases

Additional information is available through these value-added data products:

- Annual Net Primary Productivity Data
- EcoTrends
- LTER/USFS Climate / Hydrology Data
- Remote Sensing Data
## 6 data packages found

<table>
<thead>
<tr>
<th>Total data packages</th>
<th>LTER data packages</th>
<th>MCM</th>
<th>NTL</th>
<th>NWT</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

### Search Results

<table>
<thead>
<tr>
<th>Title</th>
<th>Creators</th>
<th>Organization</th>
<th>Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lake Alboron spillway stream water quality</td>
<td>Caine</td>
<td>NWT</td>
<td>Niwot Ridge, LTER Site, LTER, Colorado hydrology, pH, solute concentration, specific conductance, stream water quality</td>
</tr>
<tr>
<td>ID: knb-lter-nwt.114.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limno Holes, Dive Holes, Lake Level Measurement Locations</td>
<td>Doran</td>
<td>lake</td>
<td>limnology, location data</td>
</tr>
<tr>
<td>ID: knb-lter-mcm.40.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>McMurdo Dry Valley Lake Levels</td>
<td>Doran</td>
<td>lake</td>
<td>lake level</td>
</tr>
<tr>
<td>ID: knb-lter-mcm.68.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>McMurdo Dry Valley Lakes Blue Box data (Continuous stage (lake level), ablation, surface PAR, underwater PAR)</td>
<td>Doran</td>
<td>stage</td>
<td>lake level, ablation, PAR, photosynthetically active radiation, blue box</td>
</tr>
<tr>
<td>ID: knb-lter-mcm.3003.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>North Temperate Lakes LTER: Lake Levels</td>
<td>North Temperate Lakes LTER</td>
<td>North Temperate Lakes LTER</td>
<td>North Temperate Lakes LTER, lake level</td>
</tr>
<tr>
<td>ID: knb-lter-ntl.30.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small survey in Northern Wisconsin Lakes</td>
<td>Solomon Olden</td>
<td>mystery small</td>
<td>chinese mystery</td>
</tr>
<tr>
<td>ID: knb-lter-ntl.40.6</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Data Collection

Data Set Description:
This is a summary of major ion concentrations for stream water samples collected at the spillway from Lake Abibon. At times when the lake level was below the spillway, no samples were taken.

Keywords:
- NWT
- Niwot Ridge LTER Site
- LTER
- Colorado
- hydrology
- pH
- solute concentration
- specific conductance
- stream water quality

Geographic Coverage:
- West: -105.643 degrees
- East: -105.375 degrees
- North: 40.0615 degrees
- South: 39.9932 degrees

Temporal Coverage:
- Begin: 1994-05-21
- End: 2001-11

Maintenance:
The documentation section of this file was developed from the NWT LTER Data Directory (Symphony files DOABSTRA.WR1, DDODOCUME.WR1, DOMASTER.WR1, DDOVARME.WR1; last updated 16 December 1991) by Susan Kindig and Rick Ingersoll in early November 1992. Additional information for the documentation section, as well as the header and data sections were supplied by Neil Caine on 12 February 1993. Data were compiled in January 1993 by Neil Caine from earlier files for individual years [RCL 15 February 1993] This file was put under access control on 30 March 1993. See the appropriate history file for additional information [RCCI 30 March 1993] Data for 1984-1991 were reformatted in order to be consistent with post-1991 data formats on 2 May 1994. The variables, comments, and header sections were modified accordingly [RCCI 12 May 1994] Data for 1993 were appended to the data section on 4 May 1994 [RCCI 4 May 1994] On 11 July 1994, it was discovered that the spring 1994 conversion from mg NO3 to NO3/NH4 had been incorrectly done (mgl, values were multiplied by 61.3 instead of 16.3). The 1984-1991 data were extracted from this file, a script to correct the error was run on the extracted data, and the corrected data were used to replace the erroneous data on 17 July 1994. [RCCI 12 July 1994] Data for 1992 were provided by Chris Seibert on 16 November 1994; these data were processed and inserted into the data section on 17 November 1994 [RCCI 11 November 1994] All data for 1993 were converted to missing because all values were deemed to be too low (presumably because the sample was inadvertently diluted) by Mark Williams and Neil Caine [RCCI 30 January 1995] Data for 1994 were appended to the data section on 12 February 1996. Columns for total and dissolved N and P were also added to the data section on 12 February 1996. [RCCI 12 February 1996] Data for 1995 were appended to the data section on 28 September 1996 [RCCI 28 September 1996] Data for 1996 were appended to the data section on 3 September 1997. Columns for particulate, inorganic, and dissolved organic N and P as well as for PO4— were also added to the data section on that date [RCCI 3 September 1997] The investigator contact information was updated on 9 June 1998 [RCCI 9 June 1998] Added stream water quality to keywords section on 21 July 1998 [RCCI 21 July 1998] Data for 1997 and 1998 were appended to the data section on 31 August 2000 [RCCI 31 August 2000] Data for 1999 were appended to the data section on 11 June 2001 [RCCI 11 June 2001] Data for 2000 were appended to the data section on 05 November 2001 [RCCI 05 November 2001] The values for total P dissolved P, dissolved organic N, particulate N, and inorganic N were determined to be incorrect for 1997, 1998, and 1999. The values were corrected on 10 June 2003 [RCCI 10 June 2003] Data for 2001 were appended to the data section [RCCI 03 November 2003] Data for 2003 were appended to the data section [RCCI 22 March 2003] Four new columns were added to the data set, (33:ODC, 33:OC, 33:ODC, 33:POC). Data for 2004 were appended to the data section [RCCI 29 May 2004] Data for 1991-2004 were reinserted into the data section due to several column errors and the fact that the Niskan Lab had re-run many samples [RCCI 29 May 2007]

Frequency:
- Contact:
  - Organization: Niwot Ridge LTER/University of Colorado
  - Position: Information Manager
  - Address: 1960 30th Street, CB 450,
    Boulder, CO 80309 USA
  - Email Address: lternw@colorado.edu
- Publisher:
  - Organization: Niwot Ridge LTER/University of Colorado
  - Address: 1960 30th Street, CB 450,
    Boulder, CO 80309 USA
  - Email Address: lternw@colorado.edu

Additional Metadata
LTER NIS – beyond the Data Portal:

• Based on LTER Strategic Research Initiative and Cyberinfrastructure Planning
• Initiated by $1.5m NSF Strategic Planning Grant to LTER

• NIS plan integrated into 2009 LTER Network Office Grant
• External review of LTER-LNO cooperative agreement
• Follow-on NSF review team focused on NIS component

• Information System Interoperability (NBII, ORNL, GEON, CUAHSI, DataONE) is critical
  • “NSF won’t fund 26 different LTER information systems”
  • Standardized interfaces to permit transparent links to and from external systems
LTER Network Information System (NIS)

- **Goal:** Support basic ecological research and science, at both the site and network levels.

- **Objective:** Design and develop a distributed, LTER-wide information system using a modular approach, while maintaining and building on present functionality.

- **Purpose:** To promote advances in collaborative and synthetic ecological science at multiple temporal and spatial scales by providing the information management and technology infrastructure to increase:
  - availability and quality of data from LTER sites – by the use and support of standardized approaches to metadata management and access to data;
  - timeliness and number of LTER derived data products – by creating a suite of middleware programs and workflows that make it easy to create and maintain integrated data sets derived from LTER data; and
  - knowledge generated from the synthesis of LTER data – by creating standardized access and easy to use applications to discover, access, and use LTER data.
LTER NIS – the PASTA Framework Overview:

• Utilize a Provenance Aware Synthesis Tracking Architecture (PASTA)
  • Metadata-driven data-flow for automatically loading data from LTER research sites
  • Data availability through a standard and well defined software interface

• All data products have associated metadata including relevant provenance information

• Site data automatically harvested by the NIS

• Initial focus on well documented tabular data,
  • Followed by more complex data (GIS/remote sensing data, and/or video data)

• Site and Site Information Manager participation critical to NIS success
  • Development of NIS tools and applications to support site information management needs
  • Development of PASTA components to include Site knowledge and experience
  • Creating well documented and quality data products for harvesting into the NIS
LTER Data Plan:

• NSF Mandate of 2 year data release – linked to review/funding
  • Some site funding more strict (immediate release)
  • Provision for protection of specific data (thesis, sensitive)

• Attention to LTER Information Management Committee
• Adoption of Ecological Metadata Language (EML)
• Focus on LTER NIS for data interoperability and migration