LNO NIS Status Report
10 October 2008

Mark Servilla
Topics

• Data Catalog/Metacat Status

• Active projects
  – EcoTrends Web Portal
  – CREATE Metadata
  – HydroDB/ClimDB Migration
  – Information Management Support
  – Subversion
  – Bugzilla
  – Data Access Server

• Horizon project(s)
  – Web Services for LNO Databases
  – Refactoring LTER Data Catalog
LTER Data Catalog/Metacat Status

- **Metacat News**
  - Upgraded to version 1.8.0 April 2008
    - Performance improvements, including restructured SQL and query caching
    - New skin layout, including LTER
    - New simplified URL syntax for accessing single documents
  - Upgraded to version 1.8.1 June 2008
    - Bug fixes (critical)
  - Added Google Maps to Advanced Search (plus other minor refinements)

- **Metacat Statistics**
  - Total document 26,883 (down from 28,695; NCEAS purged test documents)
  - EML and related 17,720 (14,616 pure EML; 6,378 from LTER)
  - Binary and other 9,163
LTER Data Catalog/Metacat Status

- Harvest Statistics

<table>
<thead>
<tr>
<th>Site</th>
<th>08/05/08</th>
</tr>
</thead>
<tbody>
<tr>
<td>AND</td>
<td>162</td>
</tr>
<tr>
<td>ARC</td>
<td>1671</td>
</tr>
<tr>
<td>BES</td>
<td>189</td>
</tr>
<tr>
<td>BNZ</td>
<td>258</td>
</tr>
<tr>
<td>CAP</td>
<td>182</td>
</tr>
<tr>
<td>CCE</td>
<td>2</td>
</tr>
<tr>
<td>CDR</td>
<td>753</td>
</tr>
<tr>
<td>CWT</td>
<td>234</td>
</tr>
<tr>
<td>FCE</td>
<td>343</td>
</tr>
<tr>
<td>GCE</td>
<td>306</td>
</tr>
<tr>
<td>HBR</td>
<td>123</td>
</tr>
<tr>
<td>HFR</td>
<td>105</td>
</tr>
<tr>
<td>JRN</td>
<td>51</td>
</tr>
<tr>
<td>KBS</td>
<td>42</td>
</tr>
<tr>
<td>KNZ</td>
<td>73</td>
</tr>
<tr>
<td>LNO</td>
<td>475</td>
</tr>
<tr>
<td>LUQ</td>
<td>96</td>
</tr>
<tr>
<td>MCM</td>
<td>181</td>
</tr>
<tr>
<td>MCR</td>
<td>24</td>
</tr>
<tr>
<td>NTL</td>
<td>166</td>
</tr>
<tr>
<td>NWT</td>
<td>145</td>
</tr>
<tr>
<td>PAL</td>
<td>2</td>
</tr>
<tr>
<td>PIE</td>
<td>157</td>
</tr>
<tr>
<td>SBC</td>
<td>24</td>
</tr>
<tr>
<td>SEV</td>
<td>171</td>
</tr>
<tr>
<td>SGS</td>
<td>18</td>
</tr>
<tr>
<td>VCR</td>
<td>118</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>6071</strong></td>
</tr>
</tbody>
</table>

Relative change in harvested documents between 1 May 2007 and 5 August 2008
The EcoTrends Project

The Earth’s environment is changing at local, regional, and global scales. Dramatic changes have occurred over the past century in climate, land cover, and habitat availability with important consequences for plant, animal, microbial, and human populations. Long-term data provide the only means to assess the rate and direction of change, to distinguish directional trends from short-term variability, and to forecast environmental conditions in the future.

The EcoTrends Project is designed to promote and enable the use and synthesis of long-term data to examine these trends in the Earth’s ecosystems.

The EcoTrends project is a collaborative effort among state and federal agencies and institutions, at present primarily in the US, to make long-term ecological data easy to access, analyze, and compare within and across sites. This website is a portal to:

- a large and diverse collection of standardized long-term ecological datasets and their metadata (> 1200 datasets)
- unique data exploration, download, graphing and synthesis tools
- information about participating research sites and their parent agencies

These datasets, tools, and information are available to anyone who would like to: view trends in ecological variables for one or multiple sites or pursue additional statistical analyses of within-site and cross-site comparisons. Please read our data use and citation policies before downloading data.

We are interested in expanding our database to include additional sites from both within the US and in other countries. See Submit New Datasets for details.
EcoTrends Web Portal

• Current Status
  – Transition to production release 1.0 Aug/Sep 2008
  – 23,782 current data sets
  – High-performance search using “Extracat” (non-Metacat) indexing scheme

• Major Components
  – Metacat 1.8.1
  – PostgreSQL 8.3
  – Apache Tomcat 5.5 / Java 1.5 (JSP, Servlets)
  – JFreeChart plots (http://www.jfree.org/jfreechart/)
  – Google Maps
  – PERL-based EML generation and data harvesting
EcoTrends Web Portal

New technology

Existing LTER metadata infrastructure (Metacat and EML)

Source A

Source B

Source C

Data Manager Library

Dataset Registry

EML

Metacat-Harvester

Cache

Workflow Engine

EML.xml

Metadata

Workflow Engine Support for multiple scientific workflow engines (e.g., R script, Kepler, Chimera, D2K)

Derived Data

Metadata and derived data products; metadata as EML

Web API

Exract Metadata

Search

HTML

SOAP

Standard interfaces to support various web portals (e.g., Trends, GEOSS, GEON, NEON, WATERS) and web service APIs

Event driven data loading for synthetic processing (e.g., new data, metadata change)
# EcoTrends Web Portal

## Features
- Browse by category
- Browse by site

### Browse by Topic

<table>
<thead>
<tr>
<th>Biogeochemistry</th>
<th>Atmospheric chemistry</th>
<th>Dry deposition chemistry</th>
<th>Litter and decomposition</th>
<th>Metals</th>
<th>Precipitation chemistry (636)</th>
<th>Soil chemistry</th>
<th>Surface water chemistry (136)</th>
<th>Vegetation chemistry</th>
<th>Water quality (3)</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Biotic structure</th>
<th>Biomass</th>
<th>Cover and density of organisms</th>
<th>Land cover</th>
<th>Phenology</th>
<th>Production</th>
<th>Species richness and species diversity</th>
</tr>
</thead>
</table>


### Browse by Site


---

LNO NIS
EcoTrends Web Portal

- Features
  - Multiple levels of metadata
EcoTrends Web Portal

- **Features**
  - Multiple levels of metadata
EcoTrends Web Portal

- **Features**
  - Multiple levels of metadata
EcoTrends Web Portal

- **Features**
  - My Data Store

<table>
<thead>
<tr>
<th>Site</th>
<th>Stations</th>
<th>Topic</th>
<th>Variable (Unit)</th>
<th>Timestep</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baltimore Ecosystem Study LTER</td>
<td>NOAA Station 8574680, Baltimore, MD</td>
<td>Climate and physical variability</td>
<td>mean sea level (meter)</td>
<td>monthly</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>California Current Ecosystem</td>
<td>NOAA Station 9410170, San Diego, CA</td>
<td>Climate and physical variability</td>
<td>mean sea level (meter)</td>
<td>monthly</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>California Current Ecosystem</td>
<td>NOAA Station 9410170, San Diego, CA</td>
<td>Climate and physical variability</td>
<td>mean sea level (meter)</td>
<td>yearly</td>
</tr>
</tbody>
</table>

Any use of data or figures from the EcoTrends project must include the following statement (replace [Original Data Source] and [EcoTrends Dataset Identifier] with values from the website as appropriate): "Data and figures were obtained from the EcoTrends Project (http://www.ecotrends.info) funded by the National Science Foundation and USDA Agricultural Research Service. These data are from [Original Data Source]; [EcoTrends Dataset Identifier]."
EcoTrends Web Portal

- Features
  - Data Set
  - Plotting
EcoTrends Web Portal

- Features
  - Data Set
  - Plotting

Legend:
- Baltimore Ecosystem Study LTER, NOAA Station 8574680, Baltimore, MD, mean sea level, meter, monthly
- California Current Ecosystem, NOAA Station 9410170, San Diego, CA, mean sea level, meter, monthly
- Florida Coastal Everglades, NOAA Station 8724880, Key West, FL, mean sea level, meter, monthly
- Georgia Coastal Ecosystems, NOAA Station 8670870, Pt. Pulaski, GA, mean sea level, meter, monthly

Original Data Source: National Oceanic and Atmospheric Administration (http://www.noaa.gov/)

LNO NIS
EcoTrends Web Portal

- Features
  - Data Set
  - Plotting
EcoTrends Web Portal

• Features
  – Data Set Plotting
EcoTrends Web Portal

- Features
  - Administration
  - Audit Trails
EcoTrends Web Portal

- Features
  - Administration
  - Audit Trails
EcoTrends Web Portal

• Features
  – Administration
  – Audit Trails
CREATE Metadata

- MODIS data for LTER sites
- Collections 3-4 per day
- Products include:
  - NDIV/NDVI cloud mask (Normalized Difference Vegetation Index)
  - EVI/EVI cloud mask (Enhanced Vegetation Index)
  - FIRE (Active fire detection)
  - Water Vapor Infrared
  - Snow Cover
- EML – one document per product per site (26 x 5 = 130)
- Daily updates to EML
- Daily harvests into Metacat (prototype only)
ClimbDB/HydroDB Migration

- ClimbDB/HydroDB migration from AND to LNO
- Planning Jul/Aug 2008
- Replicate identical hardware/software Sep/Oct 2008
- Move into PASTA framework 2009
  - Enables planning and testing of auto-loading of site data
  - Add CUAHSI web service wrapper
Information Management Support

- Controlled Vocabulary
  - Auto-completion query form
- Unit Dictionary
  - Import site custom units
  - SVN development support for next generation tools
Subversion

• **Migrate current LTER CVS to Subversion**
  – A modern Source Code repository

• **Benefits**
  – Being used by majority of community
  – Not time-stamp dependent; uses hash values
  – Versions both files and directories
  – Provides for deletes, moves, copies
  – Simplified access control (svn://urban.lternet.edu/project)

• **Schedule**
  – Planning Aug 2008 ✓
  – Deploy SVN Sep 2008 ✓
  – Replicate CVS into SVN Sep 2008 ✓
  – Deprecate CVS Nov 2008
Bugzilla

• Bug/feature tracking application for software projects
• Flexible interface with lots of bells and whistles
• Open source community standard
• Open to community for input at http://fire.lternet.edu/bugzilla
• Ready for business!
Data Access Server

• Managed access to LTER data products
  – User registration and acceptance of LTER Data Policy
  – Data access notification to data owners
  – Audit logging of data access
  – Data URL/Proxy URL management interface

• References
  – http://lno.lternet.edu/projects/das
  – http://intranet.lternet.edu/archives/documents/Newletters/DataBits/07fall/#fa3

• Prototyped and RFC May 2007
• Full planning and design Oct 2008
Horizon Projects

• Web Services for LNO databases
  – Prototyped April 2006
  – RFC for LTER implementation in progress
  – Full planning and design Jan 2009

• Refactoring the LTER Data Catalog
  – Remove from Metacat skins framework
  – Increase and improve flexibility
  – Based EcoTrends model
  – Limits fail-over through NCEAS :-(
  – On the table for discussions