



**Network
Information
System**

LNO NIS Status Report

15 May 2006

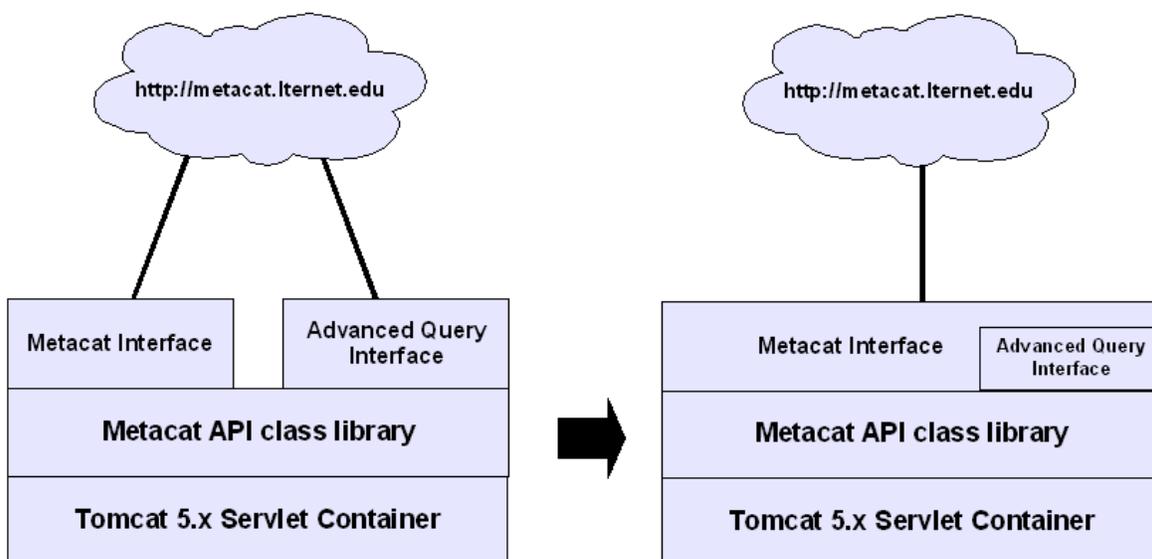
Contact:

*M. Servilla – servilla@lternet.edu
LTER Network Office
Department of Biology, MSC03 2020
1 University of New Mexico
Albuquerque, NM 87131-0001*

1. **Previous Status Report** – Submitted 2 February 2006. The following report contains status on new or ongoing tasks/projects. Please refer to previous reports for completed tasks.
2. **ILTER Data Catalog** – The LTER Data Catalog constitutes tools and services that support the discovery and access to LTER metadata and data, and includes metadata and data from the broader ecological communities that participate in the Metacat replication program.
 - (a) **Metacat** – Metacat operation is status quo ante – no updates or modifications have been implemented at this time. LNO is awaiting the production release of Metacat 1.6 (currently at 1.5); we anticipate significant performance improvements based on early reports from NCEAS/metacat-dev. There are currently 5296 EML documents in the LNO Metacat database.
 - (b) **Harvester** – Support and maintenance of the Metacat Harvester are ongoing. Harvest statistics as of April 2006 can be found in the table below. Total number of harvested EML documents is 4107 (77%).

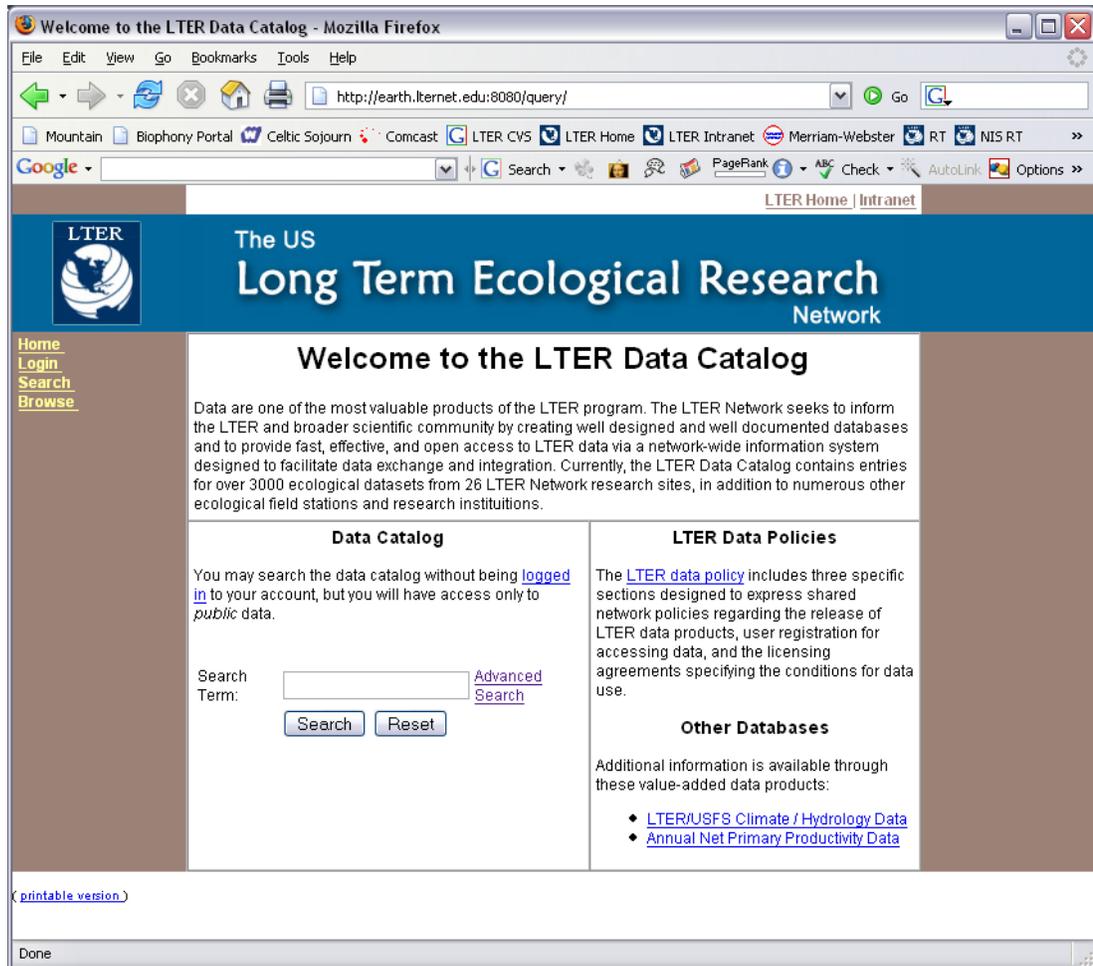
<i>Site</i>	<i>Harvesting Since</i>	<i>Frequency</i>	<i>Datasets</i>
AND	06/24/2005	2 weeks	131
ARC	04/22/2005	1 week	1597
BES	04/19/2005	1 day	11
BNZ	09/22/2005	1 day	139
CAP	08/04/2004	1 week	29
CWT	05/04/2005	1 month	190
FCE	08/24/2005	1 day	269
GCE	04/08/2004	1 week	278
HBR	07/23/2004	1 day	112
HFR	03/13/2006	1 week	96
JRN	09/21/2005	1 week	51
KBS	08/04/2004	1 day	41
KNZ	08/10/2005	1 week	43
LNO	01/18/2005	1 week	360
LUQ	05/06/2005	1 month	96
MCM	02/22/2006	1 day	128
NTL	04/15/2005	1 month	46
NWT	06/21/2005	1 day	139
PIE	07/11/2005	1 month	118
SBC	11/23/2005	1 week	24
SEV	07/25/2005	1 day	91
SGS	06/30/2005	1 month	9
VCR	07/12/2005	1 week	109

(c) Advanced Query Interface – The LTER Data Catalog's development path for the Advanced Query Interface resulted in a split application context between the Advanced Query Interface and Metacat. This split was initiated because the default skin login/session management in Metacat is not functioning correctly. This necessary function was re-written as a Java servlet class (as opposed to Javascript/JSP) and implemented in the Advanced Query Interface as a Java servlet context that runs independent of Metacat (although, it utilizes much of the underlying Metacat class hierarchy). The current model utilizes the Advanced Query Interface context for all web-based presentation with the exception of displaying the EML document details – this utilizes the Metacat interface. Although fully functional, this duality in the LTER Data Catalog has resulted in some minor inconvenience with respect to site specific customization of the LTER Data Catalog presentation (see below). The LNO NIS team is now planning to refactor the Advanced Query Interface so that it resides completely within the Metacat context, thus reducing any unnecessary maintenance overhead. Estimated level of effort is 3 weeks, with an anticipated completion date of 1 June 2006.

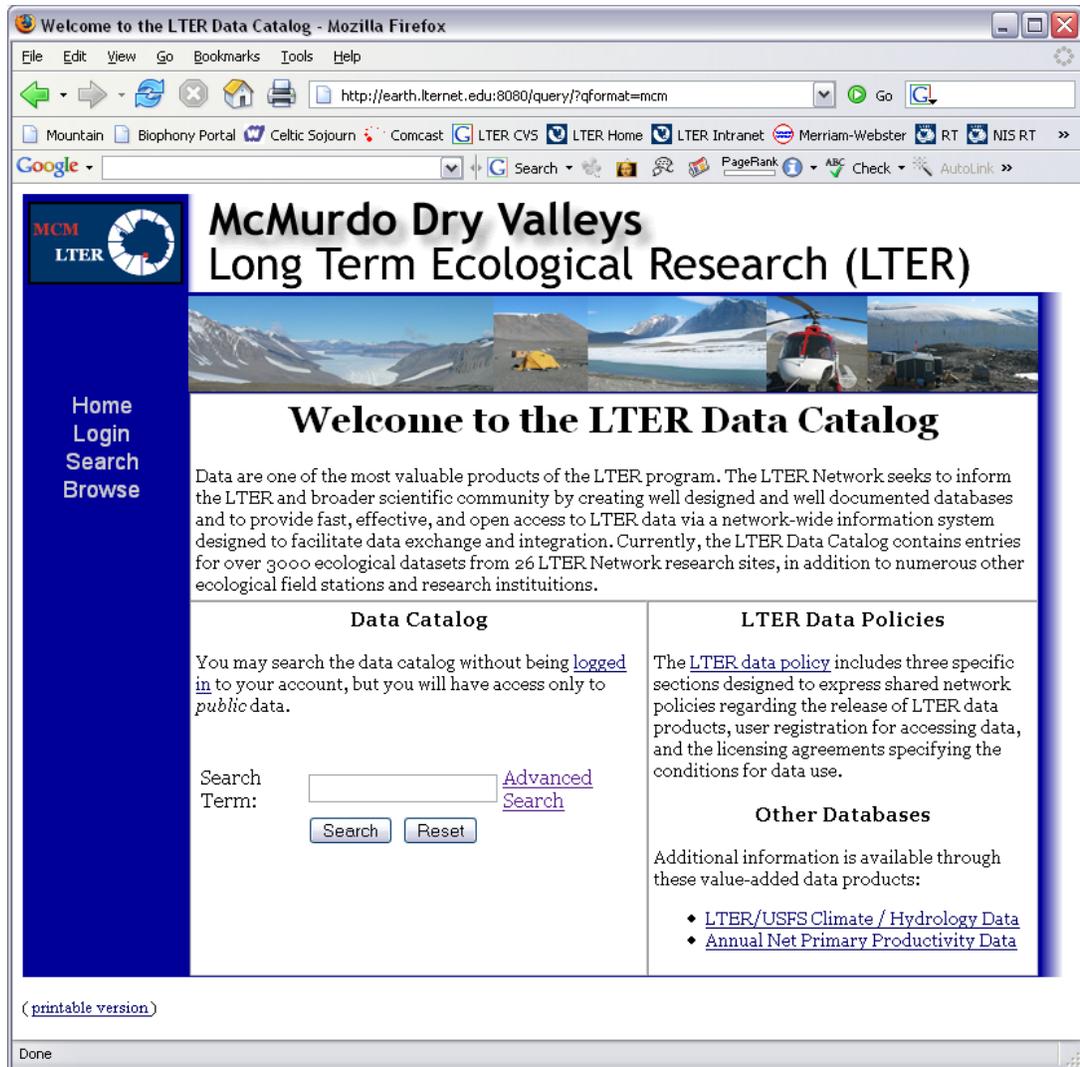


(d) LTER Data Catalog Site Customization – The LNO NIS team (LNT) investigated the customization of the LTER Data Catalog presentation layer to a specific LTER site look and feel. Specifically, LNT worked with the MCM website layout and style sheet to develop a model for “skinning” the LTER Data Catalog interface to the look and feel of another LTER Network site. In this test case, LNT utilized the default layout of the MCM website, along with their CSS, to produce an MCM version of the LTER Data Catalog. Although the process works well, the exercise revealed that maintaining multiple site presentations within the current Advanced Query Interface framework will be

time consuming. For this reason, providing this customization functionality to other sites will be postponed until the Advanced Query Interface is folded into the Metacat application context (see Advanced Query Interface above).



Default LTER Data Catalog presentation.



LTER Data Catalog using an MCM presentation.

- (e) Controlled Vocabulary Metadata Index** – The Controlled Vocabulary Metadata Index is a sub-project of the LTER Controlled Vocabulary Working Group. The basic concept is to “crawl” the LTER Metacat on a nightly basis looking for all active EML documents that contain one or more of the vocabulary terms and generate a list that contains direct links to each document – the index. This indexed list will then be linked to the “Catalog Browse” page of the LTER Data Catalog, which will allow users to access all documents associated with a term in the vocabulary. The advantage of this strategy is that accessing EML documents from a direct link is significantly faster than performing a search through the Metacat API. The vocabulary list of terms can be modified to add or delete terms as necessary. The frequency of “crawling” can be adjusted to meet the update frequency of the Metacat database. Currently, the list of vocabulary terms include the set of

terms found on the Catalog Browse page and all LTER site three letter acronyms – the latter being important to quickly find all documents from a specific LTER site. This project is in progress and with an approximate completion date July/August 2006.

3. EML Metadata Site Support

(a) Site visits – Recent site support includes:

- i. CDR: Anticipated harvest of all data/metadata at or near Level 5 EML.
- ii. LUQ: Test-bed for Metadata Management Model (see below).
- iii. MCM: Level 5 EML harvested for 85% of data/metadata.
- iv. SGS: Transformation tool for converting MS Access DB to EML at or near Level 5.

(b) EML Quality Assessment – A follow-up study was performed to assess the availability/usability of site data by noting the presence of the “online” and “dataTable” elements in the EML documents found in the LTER Metacat. The EML “online” element usually contains a URL that should point to the data described in the EML document. The EML “dataTable” section describes in great detail the structure of the data entity itself, that is, column names, units, etc. These EML elements are indicators of richer metadata and an important attribute of the metadata that may support automated data access and use. Of about 4000 documents analyzed, 96% contained an “online” element that pointed (via an URL) to either an intermediate web-page or directly to the data. Five LTER sites have a direct, unobstructed URL to the data, which is the target goal. About 25% of the documents had content in the “dataTable” element, which must be present to make useful any streamlined use of data (e.g., a programmatic use of the data). Overall, we have verified the progress of metadata enrichment and suggest that a network wide effort can be conducted to keep improving the data access as well as other aspects of the quality of the standardized metadata.

4. NBII Strategic Relationship

(a) EML to BDP Crosswalk – The EML2BDP crosswalk XSL style sheet has been extensively modified to add additional features of EML. Updates include minor bug fixes to the “geospatial” section of the transformation. One potentially major issue is that BDP does not support the notion of multiple dataset entities – that is, multiple tables, raster or vector objects, or other entities are not supported in a single BDP document. Solutions are being investigated by Inigo San Gil (LNO) and Chris Lindsley (ORNL).

(b) BDP to EML Crosswalk – The BDP2EML crosswalk XSL style sheet is near completion, and will allow reverse harvesting from the NBII clearinghouse into the LTER Metacat for greater metadata discovery – status quo ante.

(c) ESRI to EML Crosswalk – An improved version of the ESRI to EML XSL style sheet for the conversion of geospatial content is in development – status quo ante.

(d) Crosswalk Web Interface – A web interface to the crosswalk tools is in development. The website will allow a user to upload one or more documents in a particular standard and have it transformed into one of the previously mentioned standards (EML, BDP, or ESRI). It is the intention to provide this service as a formal Web Service – status quo ante.

5. Custom Unit Registry – The LTER Custom Unit Registry is a database registry of scientific units defined by the Scientific, Technical, and Medical Markup Language (STMML) and required by the EML 2.0.1 specification for describing such units. The registry is currently available through a website interface that has been status quo since Fall 2005. Part of the code base for the registry was utilized by PAL and others for use in related unit and attribute registries/dictionaries (see article in LTER Databits 2006). There is renewed interest in extending the functionality of the registry; specifically, by Karen Baker at PAL/CCE. Additional requirements may include:

- (a)** Authentication and authorization functionality to allow approved users access to upload, modify, and perhaps delete registry entries.
- (b)** Utilize the Life Science Identifier mechanism for uniquely marking each unit and unit-type.
- (c)** Provide a web service wrapper around the registry as an alternative access API.

A more detailed summary of potential work will be elaborated in a NIS Request for Comments document in Spring 2006.

6. Web Services Model – The LTER Web Services Model is a project that is evaluating current web service technologies that may be applied to data integration and database synchronization across the LTER Network. The canonical problem that is currently being studied is the synchronization of LTER Network-wide databases, such as PersonnelDB, SiteDB, and potentially the new TrendsDB. The LNO NIS team has developed a prototype web service that performs remote searches, inserts, updates, and deletes to a test personnel database, in addition to obtaining table attributes and datatypes. The current prototype has provided a solid foundation for testing different models for integration and synchronization. A more detailed summary of potential uses of web services within the LTER Network will be elaborated in a NIS Request for Comments document in Spring 2006.

7. Trends Data Module – The Trends Data Module is based on the integration of long term core datasets of LTER sites into derived products. This work originated with Dr. Debra Peters of the Jornada LTER. The LTER Coordinating Committee, in September 2005, accepted the Trends project as a new NIS Data Module. The LNO NIS team has met with Christine Laney (JRN LTER), project manager, in November 2005 and at the Trends Editorial Committee meeting at the Sevilleta LTER 14-16 February 2006. In addition to meeting with Christine Laney at the February meeting, the LNO NIS team met with members of NISAC

(Wade Sheldon), IMExec (Ken Ramsey), and NCEAS (Mark Schildhauer) to develop a strategy for the Trends Data Module. As an outcome of the February 2006 meeting, the goal for a single website that complements the Trends Book was to be developed in two phases: First, develop a simple and static Trends website that contains near-identical content as the Trends Book and in a downloadable format by the September 2006 All Scientists Meeting (ASM) – this task is the responsibility of Christine Laney, with guidance provided by the LNO Web Development team. Second, (and the responsibility of the LNO NIS team) develop a prototype application that dynamically uploads site data, performs a limited integration and synthesis process, including generating the appropriate EML for all derived products, and making the derived data accessible for graphing and download via the Trends website interface, with an anticipated demonstration by the September 2006 ASM – this is in accordance with the discussions held by IMExec on 7-8 February 2006 in Santa Barbara, California (see email from Don Henshaw to NISAC, 10 February 2006). The LNO NIS team has initiated a working document that outlines the vision, goals, strategies, and implementation plan for developing the Trends Data Module for the LTER NIS (please refer to <http://cvs.lternet.edu/cgi-bin/viewcvs.cgi/NIS/projects/trends/docs/>). A project task has been added to the LNO NIS team “dotproject” management application (please refer to <http://fire.lternet.edu/dotproject/> and login with guest/guest).

- 8. Grid Proposal** – In collaboration with NCSA, KBS LTER, and SBC LTER, the LNO submitted on 25 January 2006 a full proposal to NSF in response to the RFP “Cyberinfrastructure for Environmental Observatories: Prototype Systems to Address Cross-Cutting Needs”. No update at this time on the status of this proposal.
- 9. Metadata Management Model** – A preliminary approach to developing a site-based Metadata Management Model is being prototyped through a collaborative development between the LNO Web Development team, the LNO NIS team, and the LUQ LTER. The design is attempting to integrate existing database content from the Network Personnel and Site databases into a single view that supports site-specific metadata. A summary report will be available Spring 2006.
- 10. Network Database Status** – LNO is in the process of requesting input on various aspects of the network databases – comments have been received regarding the use of unique ID's in the LTER Bibliography and the responses are being incorporated into a web-services RFC that will be circulated in the near future. Changes to the personnel database to accommodate multiple responsible sites are underway and are being closely coordinated with the web services developments. Of the 23 sites contributing to the LTER Bibliography, 17 of those sites have records for 2005 or 2006.
- 11. CI Strategic Plan** – Mark Servilla, John Vande Castle, Bob Waide, and James Brunt have been contributing significant time to providing input and editing on the Cyberinfrastructure (CI) Strategic Plan being developed for the Network as part of the LTER science planning process.