

University of New Mexico

UNM Digital Repository

Long Term Ecological Research Network

Long Term Ecological Research (LTER)

2-2013

Executive Board Meeting, February, 2013

Long Term Ecological Research Network

Follow this and additional works at: https://digitalrepository.unm.edu/lter_reports

ILTER Network Information System

Status 28 February 2013

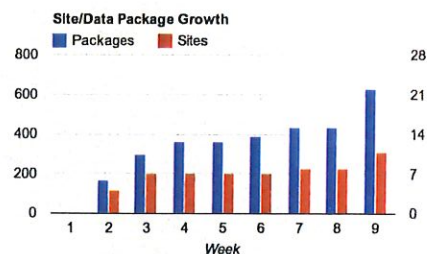
Current Status

The LTER Network Information System (NIS) and PASTA went operationally “live” on 7 January 2013. LTER sites quickly began uploading data packages to PASTA, along with 15,000+ data packages from the EcoTrends synthesis project. Today, there exist over 16,000+ data packages in the LTER NIS with 631 data packages being contributed by 11 sites (including the legacy data packages from the North Inlet LTER site). New site contributions are growing regularly.

All “publicly” accessible data packages are registered with publishing industry standard Digital Object Identifiers (DOIs) that resolve back to the NIS Data Portal (the user website for PASTA) site for the given package. Through access control rules, site information managers regulate what content of the data package, if any, are accessible to the general “public” or to a specific list of registered users.

Quick Facts:

- Operational 7 January 2013
- 631 Site contributed data packages
- 11 Contributing sites (including North Inlet)
- 16034 Total data packages
- 15403 EcoTrends data packages



Site contributed data packages: 631
Total data packages: 16034

Current feature list:

- Data Package Management
 - evaluation and quality metric report
 - upload and revision management
 - data package browse and accessibility
- Digital Object Identifier registration for data packages
- Subscription and notification for data package events
- Provenance-specific metadata for derived/synthetic data
- System “state-of-health” monitoring through Nagios®
- Identity management and authentication
- Data package access control
- Audit logging and reporting
- NIS Data Portal user interface



AGENDA

Finding the Needle in the Haystack:
**A Symposium of the Board on Research Data and Information
on Strategies for Discovering Research Data Online**

**National Academy of Sciences Auditorium
2100 C Street NW, Washington, DC**

**February 26, 2013
3:00 pm - 5:30 pm**

-
- 3:00 pm Framing the issue and introduction of the first panel
- Clifford Lynch, CNI
- 3:10 Panel One: Interdisciplinary Approaches to Discovering Data Online
Session Chair: Clifford Lynch, CNI
- Speakers:
- Data Cite and Science.gov, Lorrie Johnson, OSTI
 - The Data Citation Index, Michael Takats, Thomson Reuters
 - Data Type Registries, Giridhar Manepalli, Corporation for National Research Initiatives
 - General Discussion
- 4:20 Panel Two: Discipline-Specific Examples of Discovering Data Online
Session Chair: Francine Berman, RPI
- Speakers:
- Better Discoverability of Biomedical Data, Michael Huerta, National Library of Medicine, National Institutes of Health (NIH), and Finding Disease Data: The Autism Example, Gregory Farber, National Institute of Mental Health, NIH
 - Strategies for Finding Earth Observation Research Project Data, Suzie Allard, University of Tennessee and the DataOne Project
 - Discovering Social Science Data Online, George Alter, ICPSR
 - General Discussion
- 5:30 End of Symposium – Reception



Finding the Needle in the Haystack:
**A Symposium of the Board on Research Data and Information
on Strategies for Discovering Research Data Online
National Academy of Sciences Auditorium**

One of the problems recognized by experts and casual data users alike has been the inability to find the full array of research databases or factual compilations that are needed to support any given query. As data continue to proliferate and research becomes more data intensive, the discoverability of factual references also grows in importance. For research funders and policymakers, there is a need to better understand data productivity and trends in science, both quantitatively and qualitatively. Yet the deluge of information and the diversity of the datasets makes the task for all users of data and facts that much more difficult.

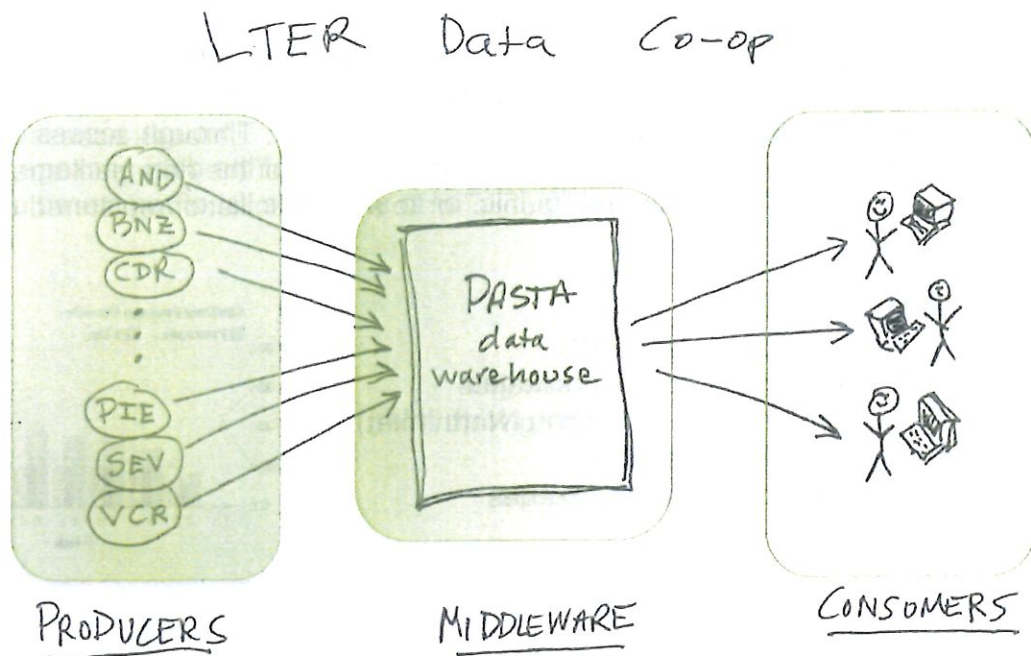
The importance of the problem is not hard to grasp, but the solutions to the better discoverability online of data are not necessarily straightforward or obvious. Libraries have traditionally grouped literature by topic, using tools such as controlled vocabularies and ontologies, and also developed systems of other identifiers, such as call numbers and shelf lists to help the user locate them in the print context. However, physical collocation is not always a necessary or practical strategy for digital datasets. As with literature, the creation of registries, catalogs, and directories may be part of the solution, but we also need to consider online search engines, persistent identifiers and associated metadata, as well as better citation and reference practices in order to enable visibility of and access to digital data. Ontologies created for literature may not always adapt well for use with data sets, particularly if we are to tap the rich potential of unanticipated reuse of data by researchers in other disciplines different from those of the original creators. Links to and between different information sources and objects may be important mechanisms for improving the “findability” of data too.

Despite the proliferation of models and solutions in various disciplines and sectors, there is a recognized need for a pervasive infrastructure, standardization of approaches, and the usual questions of who does what, where, and how? This symposium therefore seeks to highlight some of these different approaches, providing examples that are both broadly interdisciplinary as well as discipline-specific to finding the right data at the right place in the right time. Although we will not offer any common solutions to this set of problems, we do hope to shed some light on the underlying issues and provide an opportunity for experts working in this area to interact, both among each other and with the audience.

The co-chairs of the Board on Research Data and Information, Clifford Lynch of the Coalition on Networked Information, and Francine Berman of the Rensselaer Polytechnic Institute, will lead the symposium discussion, beginning at 3 p.m. on Tuesday, February 26. The event will continue for 2 ½ hours in a mix of short presentations and discussion. The entire proceedings will be recorded and an audio-tape will be archived on the Board’s website. The symposium is free and open to the public, but advance registration is requested. The meeting will be followed by a reception outside the main auditorium.

Upcoming Tasks

The LTER NIS and PASTA went operationally “live” nearly two years before their intended release. Using a “producer/consumer” analogy of a traditional food cooperative, the LTER NIS is cast as the “LTER Data Co-op” to define primary user groups (see below; shading indicates focus). To date, development focus was placed primarily on the loading of data packages from the “producer’s” perspective to ensure viable data packages at the “get-go”. Although data packages are discoverable, and accessible, through PASTA, new focus must be placed on the “consumer’s” experience for the LTER Network Information System to be successful.



The following prioritized list provides a view into tasks still required to improve the “consumer’s” experience of using the LTER Network Information System and PASTA:

- DataONE Generic Member Node
- New metadata engine
- NIS Data Portal enhancements
- Extending Identity Management
- Smart data management
- Performance tuning & system hardening

Although the current project is targeted for “transition to maintenance” in 2014/2015, the LTER NIS and PASTA will continue to evolve as new user requirements and technology improvements are identified.