

University of New Mexico

UNM Digital Repository

Long Term Ecological Research Network

Museums and Research Centers

4-2012

Strategies for Improving LTER Network Data Availability

Long Term Ecological Research

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

Recommended Citation

Long Term Ecological Research. "Strategies for Improving LTER Network Data Availability." (2012).
https://digitalrepository.unm.edu/lter_reports/167

This Report is brought to you for free and open access by the Museums and Research Centers at UNM Digital Repository. It has been accepted for inclusion in Long Term Ecological Research Network by an authorized administrator of UNM Digital Repository. For more information, please contact disc@unm.edu.

Strategies for Improving LTER Network Data Availability

IMExec – April 2012

NSF has made increased data availability a high priority, and has emphasized that a) the Network should be able to effectively deliver and provide easily discoverable site data through a single portal, and b) we need to push forward now and not wait for the completion of the PASTA system in 2014 before making more data available. In response to feedback and discussions at its recent meeting¹, IMExec identified four areas for which the Network can take immediate steps to address both the perception that data are not available and the issue of overall availability. The timeline for implementation of these strategies and the development of more concrete plans must be coordinated among the Executive Board, LNO, NISAC and IMExec, and include consultation with members of the Information Management Committee and Science Council.

Four problems (**A-D**) were identified by IMExec and are presented in a tabular format, generally modeled on the structure of the LTER Strategic and Implementation Plan (SIP). Definitions of terms used in this document are presented below this introduction. The problems (**A-D**) are presented next, in order of feasibility and overall importance given recent discussions and feedback from NSF. References to SIP goals are also included. To facilitate implementation, this plan also presents four projects (numbered **1-4**) that could be supported by available ARRA funds, and these are described at the end of the document. All four projects address Issues **A-D** and the linkages between the problems and projects are cross-referenced.

Definitions:

Context: A summary of a current situation or problem creating the perception that LTER data are not available. Some general feedback or concomitant issues may also be included.

Data package: an EML document plus data in the Network catalog

Goal: a possible solution to perceived problem given its context

Where/by whom: the groups who should be responsible

Strategies: how IMExec thinks the Network could address the goal. More specific tasks are still to be developed.

Funds: projects that could address these goals, which are described at end of document.

SIP id: reference to a topic in the Strategic & Implementation Plan, Information Management section.

Current issues affecting network data availability:

A. Context: We have many data packages already submitted in which the path to data is not clear.

EML usage is not yet standardized across the network. Full standardization is our long-term goal, will be required by PASTA, and is best addressed by upgrading site systems. However, some short-term remediation may be helpful as well. Non-standardization also includes the varied use of data registration forms, and standardization may require changes or clarification of access policies that can then be followed by implementation changes (SIP id 2BB).

¹ http://im.lternet.edu/news/committees/im_exec/notes/2012_04_3

Goal: More existing data packages should have visible or easily 'clickable' data URLs

Where	At sites	LNO
By whom	Site IMs, EML mentors or contractors	LNO personnel, contractor
Strategies	<ol style="list-style-type: none"> 1. Bring EML metadata into conformance with a few essential practices, mainly the location of the data URL 2. Develop a practice to identify and display metadata for Type II data. 3. Ensure that all EML includes the 5 essential metadata features (available separately). 	<ol style="list-style-type: none"> 1. Enhance current views of data packages by further adapting Metacat views of search and browse results (current development portal has already been approved by NISAC)
Funds	<ol style="list-style-type: none"> 1. To support EML mentors engaged in educating site personnel, and for site-to-site code assistance (#1, under "Projects for ARRA funds" below) 2. To advance IM systems in general by funding 2 cross-site projects (#2, 3, under "Projects for ARRA funds" below) 	<ol style="list-style-type: none"> 1. To develop XSL templates for LTER-specific EML views (#4, under "Projects for ARRA funds" below)
SIP id	1A, 2BC	

B. Context: There are data at sites that should be available but are not yet described. Nearly all sites have a known backlog of data and an additional unknown quantity of data that should be available, but may not be in the possession of LTER Information Managers.

Goal: More new data packages should be added to the network catalog

Where	At sites, NISAC
By whom	Site personnel, Scientists
Strategies	<ol style="list-style-type: none"> 1. Prioritize backlog of data that is expected, but not yet in site data catalogs. 2. Conduct expeditions for additional unsubmitted data. Use proposals and papers to develop inventories of anticipated data
Funds	<ol style="list-style-type: none"> 1. To support EML mentors engaged in educating site personnel, and for site-to-site code assistance (#1, under "Projects for ARRA funds" below) 2. To advance IM systems in general by funding 2 cross-site projects (#2, 3, under "Projects for ARRA funds" below).
SIP id	1C (all), 1E (all)

C. Context: Data users and NSF expect to be able to find certain kinds of data from all or most sites (e.g, data related to core research areas). Currently, no data are available by category beyond keyword searches, and there is no process for identifying data categorically. Sections C and D are related, in that the same science priorities that define potential data categories here could inform the choices in Section D, (i.e., those data needing workflows).

Goal: Emphasize data that address science priorities

Where	At sites, NISAC	LNO
By whom	NISAC, scientists	LNO personnel, contractor
Strategies	1. Identify high priority derived data 2. Encourage sites to contribute data for high-priority science projects as PASTA-ready EML	Highlight those data in the catalog
Funds	n.a.	To further develop the Network data portal, #4, under "Projects for ARRA funds" below
SIP id	1ABC	

D. Context: NSF has stressed that contributions of data to PASTA could be prioritized, and that not all data may need to be used in workflows, nor described to the level required by PASTA. Data priorities here should be informed by the scientific priorities identified in section C. A related question will be how the NIS has planned to handle and/or display non-PASTA data. Currently there is no process for identifying high-priority data.

Goal: Identify the data that should be in PASTA

Where	Network
By whom	NISAC, scientists
Strategies	Prioritize data for PASTA-readiness. May include: a. data packages needing workflows (see Section C) b. time-series. c. Data useful for cross-site comparisons at [some number of] sites.
Funds	n.a.
SIP id	1AB, 1B (all)

Projects for ARRA Funds:

- EML mentors and technical assistance:** IMExec supports additional funds to sites whose information managers are assisting other sites, either as EML mentors, or with short-term code repair. Site-to-site assistance should focus on producing a) EML in regularized Network formats and b) PASTA-ready data. To ensure consistency, EML mentors should be involved in or advise work on the other projects (addresses Section A, B).
- DEIMS Developer:** the Drupal Environmental Information Management System (DEIMS) is anticipated to include 6-8 sites. This group is developing a common data model and forms for site metadata management. Funds to support a developer could complete the project. Requires participating sites to adopt and support Drupal locally, and for a developer to work closely with the EML mentor(s) (addresses Section A, B).
- Metabase Developer:** the GCE Metabase is currently in use at 4 sites. The database should be enhanced for further cross-site use and/or to export EML using SQL-XML. As with project #2, the EML mentors should advise development (addresses Section A, B).
- Network data catalog improvements:** The current views of data package search results and of individual EML data packages are generic. These could be redesigned to take advantage of regularized EML construction patterns. (Section A). Additional improvements/views could highlight categorized data, or data which have workflows available (addresses Sections C, D)

Priority

1

2.5

2.5

4

\$20K
\$40K
\$80K
\$60K
\$12K