LTER

Network
Information
System

Auditing LTER Data Access

Request for Comments (Version 1.0)

10 May 2007

Mark Servilla
servilla@lternet.edu

LTER Network Office
Department of Biology, MSC03 2020
1 University of New Mexico
Albuquerque, NM 87131-0001

1 Introduction

The LTER Network has invested considerable time, effort, and funding into the collection of scientific data. Access and use of this data is formalized through the end user's acceptance of the LTER Network Data Access Policy, Data Access Requirements, and the General Data Use Agreement (http://www.lternet.edu/data/netpolicy.html), which was approved by the LTER Network Coordinating Committee on 6 April 2005. Motivation behind these policies and agreements is driven by the need to document the flow of data from the LTER Network out to the community to validate broader impacts of the LTER program. As such, the LTER Network has adopted a "standard" for data *access* and *use* that now needs to be implemented into both local and network-wide computing infrastructure. This standard, in simple-terms, requires that the end user registers basic identifying information, including name, affiliation, email address, and full contact information, into a registry of the LTER Network. Further, acknowledgment and acceptance of either the General Public Use Agreement or any Restricted Data Use Agreement applied to a data set, and a statement of the intended use of the LTER data, will be recorded prior to the release of any LTER data.

The following document is a formal Request for Comments (RFC) to the LTER Network from the LTER Network Information System (NIS) development team to solicit input for the design of a network-wide process to address end user compliance to the LTER Data Access Policy. We suggest the use of a centralized Data Access Proxy (DAP) service that brokers access to site data through a managed Data Access Server (DAS), and provides the necessary user registration, identification, and event logging/notification processes necessary to enforce the policy. The design, advantages, and disadvantages will be discussed below. We also propose the development of a portal-based application to allow LTER researchers to monitor and generate reports detailing LTER data access.

2 Background

The LTER Network makes data available through two primary venues: 1) each LTER site supports an independent website of their own management and provides access to static or dynamically streaming data through a URL and 2) direct referencing of LTER data through network-based links (e.g., URL or database connection) that are described in an Ecological Metadata Language (EML) document hosted by one or more Metacat XML database management systems. In either case, access to LTER data is often just a "hyperlink" away. For site-based data access, registration processes and the information collected vary between LTER sites (if at all), and often require the end user to re-register when accessing new or different data (see http://intranet.lternet.edu/modules.php?name=UpDownload&req=getit&lid=423). Notification of a data access event may or may not be furnished to the data owner/provider; and, the end user will likely never be notified of the original data owner/provider contact information for citation purposes. In the case of data access through links in an EML document, the end user is provided site data without any rigorous identification process. If the site performs local event logging when the EML data access takes place, the site is only capable of recording the network address of the computer being used by the end user and, again, may not provide event notifications to the data owner/provider. In some cases, the end user is redirected to a foreign web-page that must be navigated further to reach data. Such efforts to mitigate unrestricted data access often prove fatal to the operation of automated processes. Unfortunately, current site-based or EML/Metacat approaches do not fully meet the requirements of a network-wide LTER Data Access Policy.

Network-wide implementation of the LTER Data Access Policy demands three functional requirements:

1. End user registration for collecting nominal information for entry into a user registry, along with a statement of the intended use for data, and an acceptance acknowledgment of the General Public Use Agreement. End user registration is assumed to be a one-time step that occurs either at the user's convenience prior to any attempt to access data or by the system invoking the registration process when a non-registered user attempts to access data.

2. End user identification to verify user registration and policy acceptance for all data requests. End user credentials must be available to compare against information contained within the user registry. Once verified, a client-side token (e.g., cookie) may be used to automatically identify the end user for future data requests. Strong authentication, such as user verification through a third-party authority, is not assumed for compliance to the LTER Data Access Policy.

3. Data access event logging for reporting purposes. All data access events should be recorded in an audit log that includes the identification of the end user, identification of the data accessed, and a date/time stamp of the event. The system portal should provide an interface for the data owner/provider such that events can be queried, viewed, and reports be generated. In addition, the system should provide real-time or near real-time notifications to the data owner/provider at the time of a data access event. Similarly, the system should also provide pertinent contact information of the data owner/provider to the end user for compliance to the General Use Agreement when data is accessed.

3 Proposed Approach

The LTER NIS development team has identified a general model for a Network-wide LTER Data Access Policy implementation strategy called the Data Access Proxy. The DAP model proposes a centralized NIS service that would perform all necessary policy actions, including the pass-through of LTER site data, on behalf of the site. The pass-through process would rely on the replacement of the URL that references site data with a "proxy" URL that points instead to a Data Access Server (DAS) hosted by the LNO. The purpose of the DAS is to validate the end user's credentials, thus confirming their compliance with LTER Data Access Policy, before allowing access to any site data. This approach requires the site to register their data URL with the DAS so that a one-to-one correspondence between the data URL and the proxy URL is declared within the server registry. The proxy URL is used in lieu of the actual data URL within any LTER metadata document (including EML) that is published for public viewing. When an end user wishes to download data by selecting the online distribution URL in the metadata document, they would be directed to the DAS first and have their credentials validated, before a data stream is returned on the site's behalf. If the end user has not registered at this point, they would be directed to the appropriate registration interface. If they have already registered and there exists a token (e.g., cookie) on their workstation, they would be provided the data without restriction. Otherwise, the end user would be directed to a log-in interface prior to receiving any data. Figure 1 presents an overview of the DAP model network-level architecture.

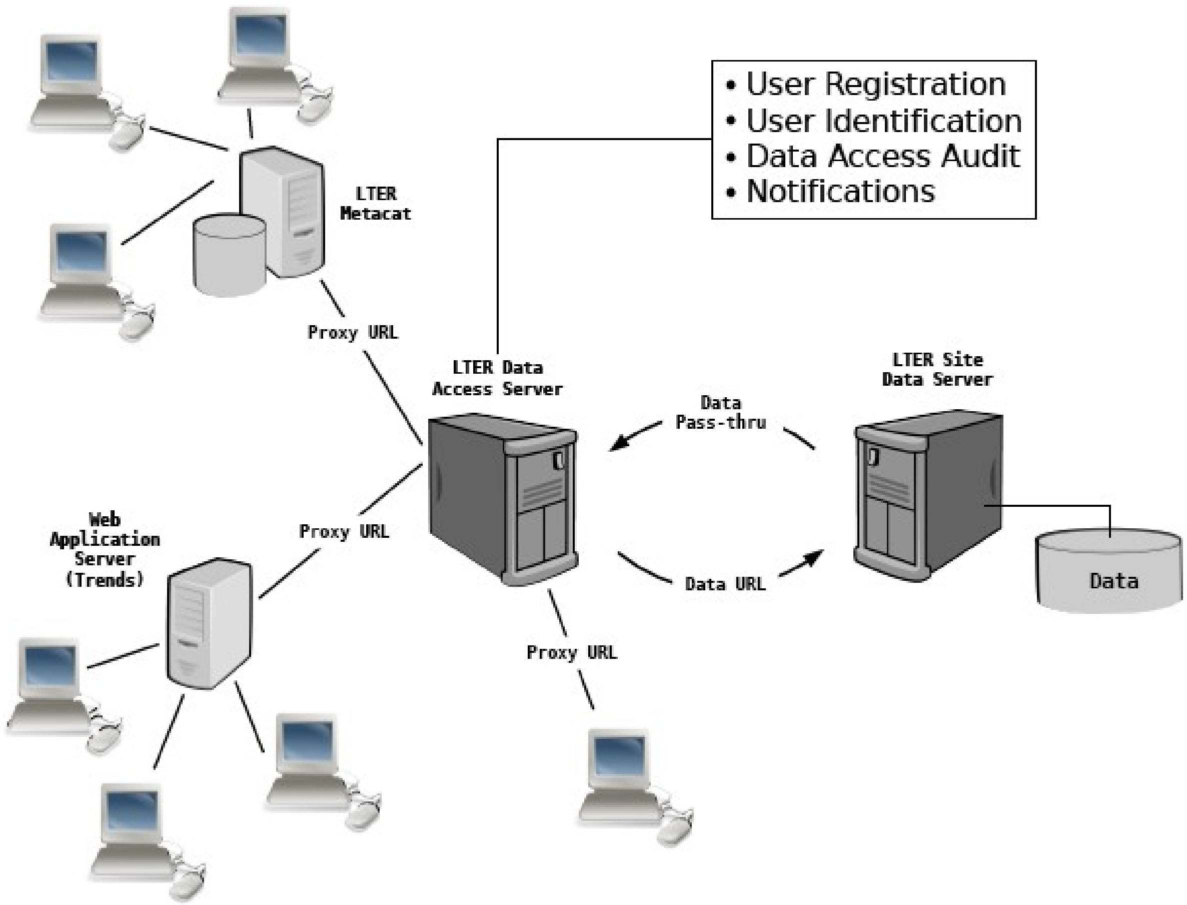
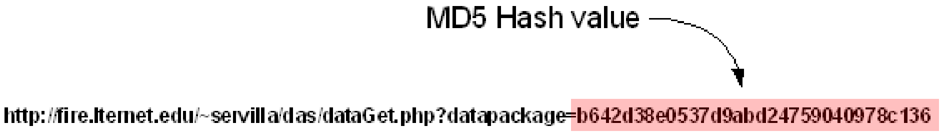


Figure 1 – Conceptual view of the LTER Data Access Server network architecture.

Any download event invoked by the end user will be logged into an audit record for reporting purposes. At this point, the DAS would send an email notification to the end user with the data owner/provider's contact information, the General Use Agreement, and any special Restricted Data Use Agreement for the specific data set that was downloaded. In addition, the DAS will also send a notification to the data owner/provider of the data download event, along with the end user's contact information and the name of the downloaded data set. The DAP model assumes that the site's data URL provides direct access to the data stream through the HTTP protocol. Further restrictions on data access can be achieved if the site only allows a specified network address or address range to connect to their data source - in this case, the network address for the LNO DAS.

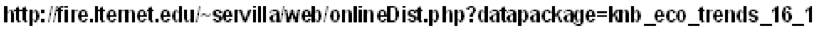
4 Proof of Concept

The NIS development team has deployed a minimal proof-of-concept DAS (http://fire.lternet.edu/~servilla/das) that utilizes data made available for the Trends Project prototype web application (http://fire.lternet.edu/~servilla/web) for demonstrating the use of a proxy URL in place of a data URL. In this case, all URLs displayed within the web application and in the EML metadata documents found within a test Metacat have been replaced with proxy URLs. The proxy URL (Figure 2) points to the DAS and contains a MD5 hash value created from the original data URL (Figure 3) that is used as an index to the site's data URL.



http://fire.lternet.edu/~servilla/das/dataGet.php?datapackage=b642d38e0537d9abd24759040978c136

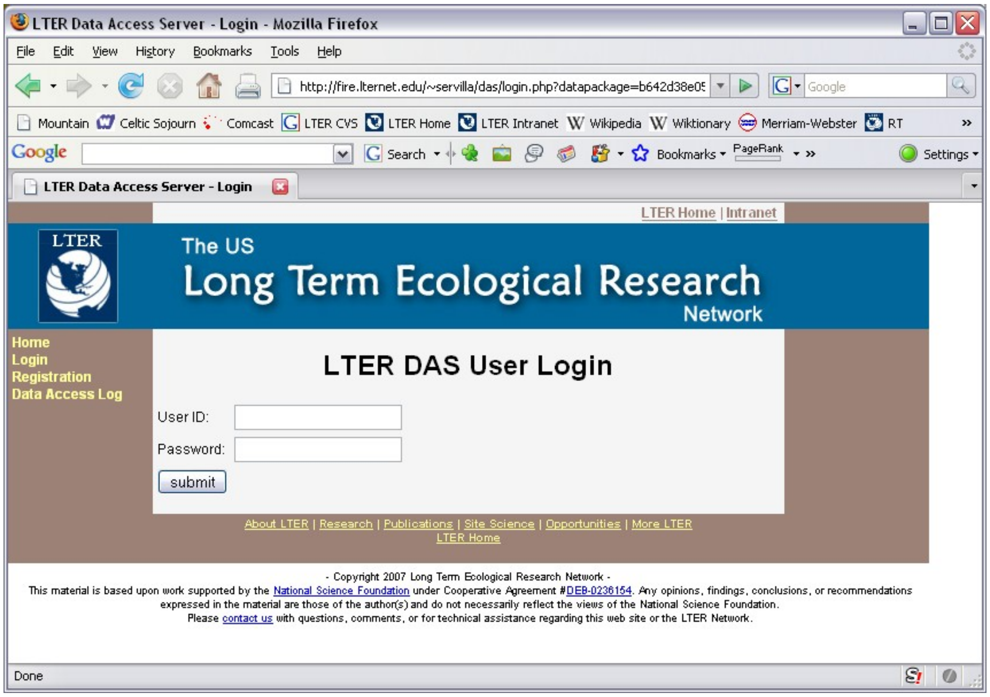
Figure 2 – Example proxy URL.



http://fire.lternet.edu/~servilla/web/onlineDist.php?datapackage=knb_eco_trends_16_1

Figure 3 – Original data URL.

If the end user is correctly identified through either an initial login (Figure 4) or by an existing browser cookie previously loaded by the DAS, the DAS opens a file stream from the site data URL and passes it on to the end user's web browser as if the original site data URL were accessed. New users may register through a typical "forms" web page (Figure 5) that saves their contact information and data use intent statement in a local database. A production system, however, would likely utilize a modified version of



LTER DAS User Login

User ID:

Password:

submit

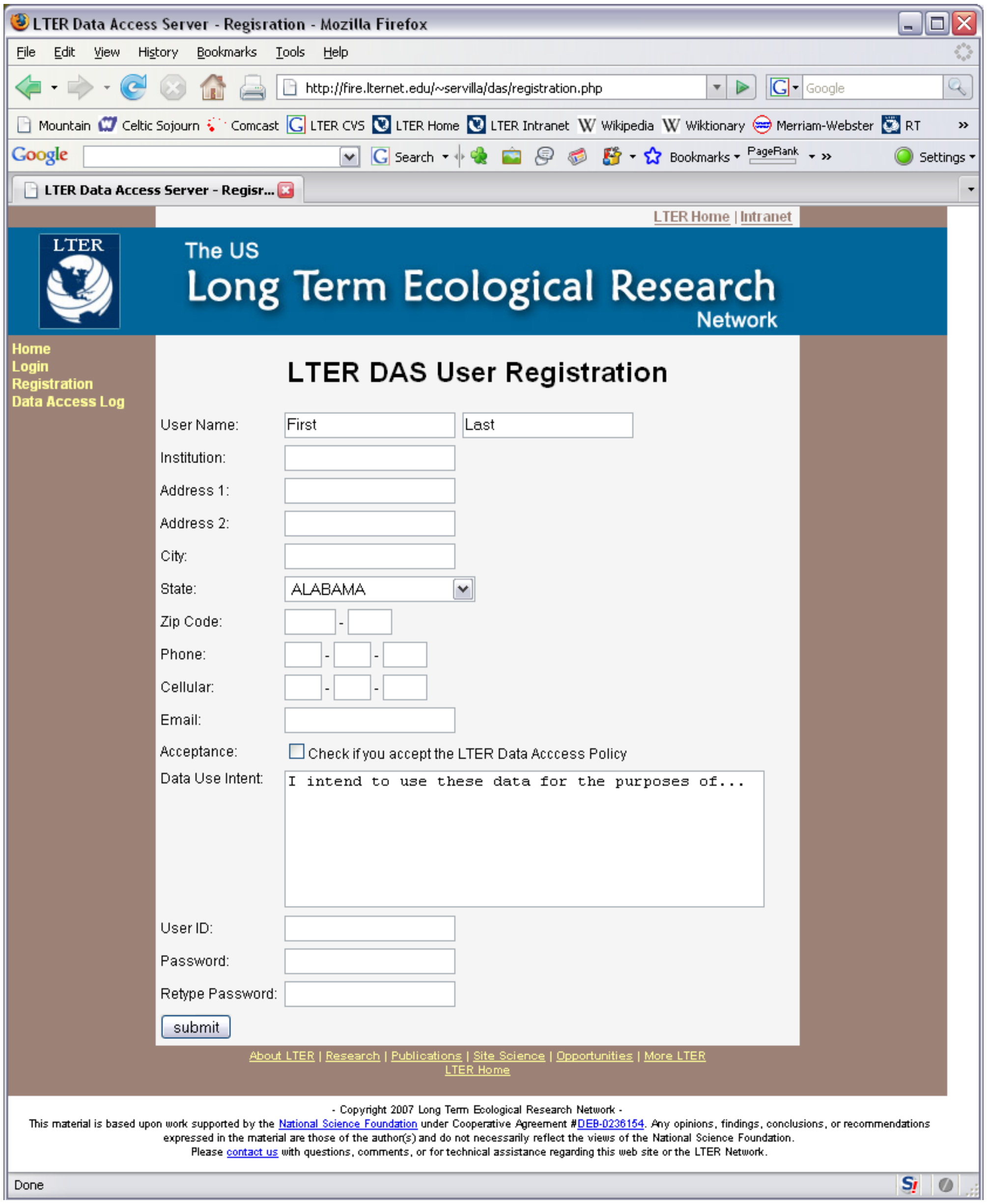
- Copyright 2007 Long Term Ecological Research Network -

This material is based upon work supported by the National Science Foundation under Cooperative Agreement #DEB-0236154. Any opinions, findings, conclusions, or recommendations expressed in the material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Please contact us with questions, comments, or for technical assistance regarding this web site or the LTER Network.

Figure 4 – DAS login page.

the current LDAP user registry. A simple list of data access events (Figure 6) is available through the DAS web site. More detailed user information (Figure 7) is obtained by selecting the user's name in the list. A notification process has not been implemented in this proof-of-concept.


LTER Data Access Server - Regisration - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://fire.lternet.edu/~servilla/das/registration.php
Google
Mountain Celtic Sojourn Comcast LTER CVS LTER Home LTER Intranet Wikipedia Wiktionary Merriam-Webster RT
Google Search Bookmarks PageRank Settings
LTER Data Access Server - Regisr...
LTER Home | Intranet
LTER
The US
Long Term Ecological Research
Network
Home
Login
Registration
Data Access Log
Done

LTER DAS User Registration

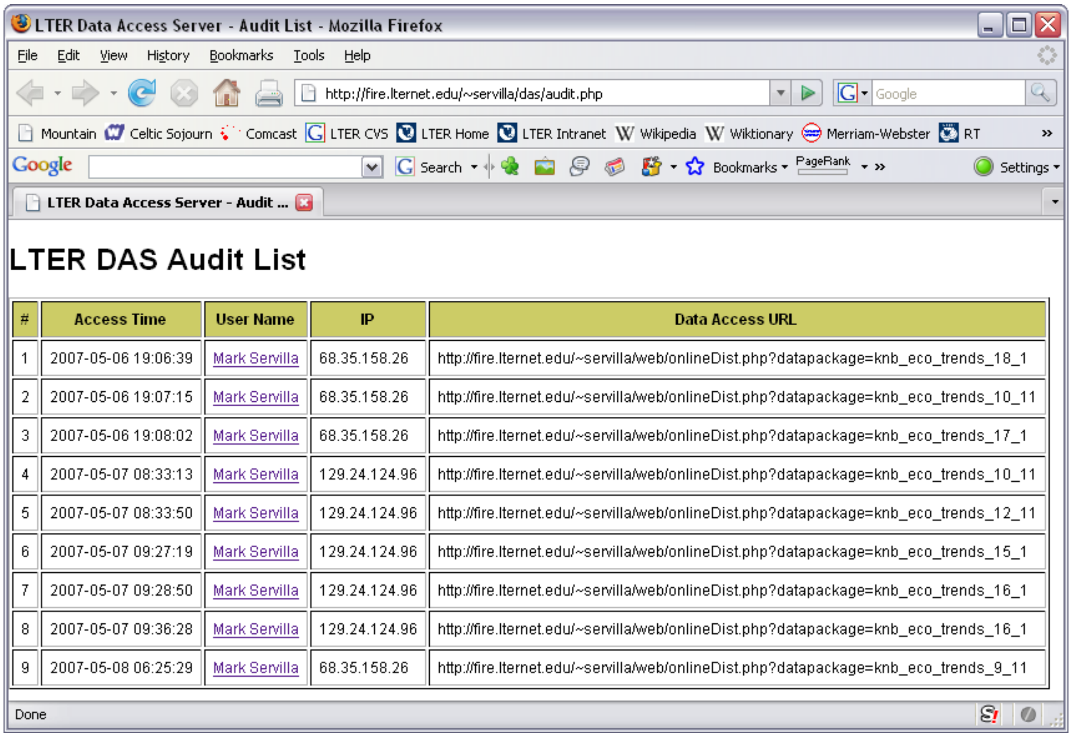
User Name:	First · Last
Institution:	
Address 1:	
Address 2:	
City:	
State:	ALABAMA
Zip Code:	-
Phone:	- -
Cellular:	- -
Email:	
Acceptance:	□ Check if you accept the LTER Data Acccess Policy
Data Use Intent:	I intend to use these data for the purposes of...
User ID:	
Password:	
Retype Password:	

submit

About LTER | Research | Publications | Site Science | Opportunities | More LTER
LTER Home

- Copyright 2007 Long Term Ecological Research Network -
This material is based upon work supported by the National Science Foundation under Cooperative Agreement #DEB-0236154. Any opinions, findings, conclusions, or recommendations expressed in the material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.
Please contact us with questions, comments, or for technical assistance regarding this web site or the LTER Network.

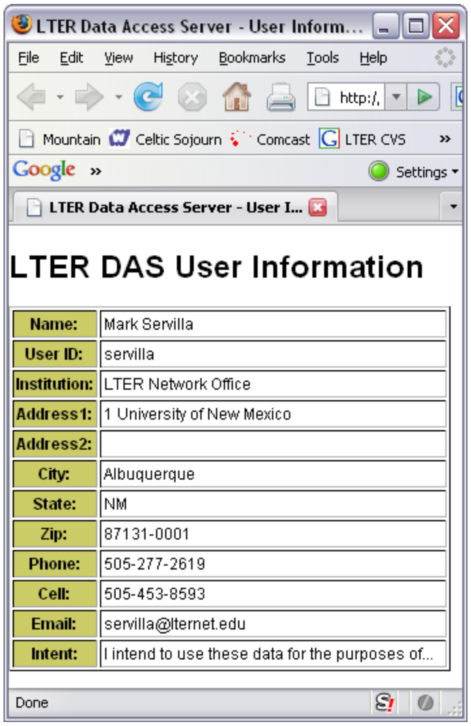
Figure 5 – DAS registration form.


LTER Data Access Server - Audit List - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://fire.lternet.edu/~servilla/das/audit.php
Google
Mountain
Celtic Sojourn
Comcast
LTER CVS
LTER Home
LTER Intranet
Wikipedia
Wiktionary
Merriam-Webster
RT
Search
Bookmarks
PageRank
Settings
LTER Data Access Server - Audit ...
Done

LTER DAS Audit List

#	Access Time	User Name	IP	Data Access URL
1	2007-05-06 19:06:39	Mark Servilla	68.35.158.26	http://fire.lternet.edu/~servilla/web/onlineDist.php?datapackage=knb_eco_trends_18_1
2	2007-05-06 19:07:15	Mark Servilla	68.35.158.26	http://fire.lternet.edu/~servilla/web/onlineDist.php?datapackage=knb_eco_trends_10_11
3	2007-05-06 19:08:02	Mark Servilla	68.35.158.26	http://fire.lternet.edu/~servilla/web/onlineDist.php?datapackage=knb_eco_trends_17_1
4	2007-05-07 08:33:13	Mark Servilla	129.24.124.96	http://fire.lternet.edu/~servilla/web/onlineDist.php?datapackage=knb_eco_trends_10_11
5	2007-05-07 08:33:50	Mark Servilla	129.24.124.96	http://fire.lternet.edu/~servilla/web/onlineDist.php?datapackage=knb_eco_trends_12_11
6	2007-05-07 09:27:19	Mark Servilla	129.24.124.96	http://fire.lternet.edu/~servilla/web/onlineDist.php?datapackage=knb_eco_trends_15_1
7	2007-05-07 09:28:50	Mark Servilla	129.24.124.96	http://fire.lternet.edu/~servilla/web/onlineDist.php?datapackage=knb_eco_trends_16_1
8	2007-05-07 09:36:28	Mark Servilla	129.24.124.96	http://fire.lternet.edu/~servilla/web/onlineDist.php?datapackage=knb_eco_trends_16_1
9	2007-05-08 06:25:29	Mark Servilla	68.35.158.26	http://fire.lternet.edu/~servilla/web/onlineDist.php?datapackage=knb_eco_trends_9_11

Figure 6 – DAS audit list.


LTER Data Access Server - User Inform...
File Edit View History Bookmarks Tools Help
http:/,
Mountain
Celtic Sojourn
Comcast
LTER CVS
Google
Settings
LTER Data Access Server - User I...
Done

LTER DAS User Information

Name:	Mark Servilla
User ID:	servilla
Institution:	LTER Network Office
Address1:	1 University of New Mexico
Address2:	
City:	Albuquerque
State:	NM
Zip:	87131-0001
Phone:	505-277-2619
Cell:	505-453-8593
Email:	servilla@lternet.edu
Intent:	I intend to use these data for the purposes of...

Figure 7 – DAS user information.

5 Advantages

1. The DAP model does not require sites to participate or change their current practice of providing direct access to their data. It is a model that may be utilized at the site's convenience, perhaps addressing sensitive or high-profile data first.
2. The DAP model is not tightly coupled to EML, the Metacat, or any other sub-system, and therefore, it can be used at both the site and network-level. Figure 1 shows that the proxy URL can be used through links embedded in the EML metadata document residing in Metacat, as part of a data link reference in a separate web application, such as the Trends Project, or simply as a data link provided in an email message.
3. Since the DAS would run as a centralized service at the LTER Network Office, tools and enhancements based on the DAP model would be available to all participating sites, including data access reports that can be perused directly by NSF officials. This can be an effective method for standing groups like the Information Manager Executive committee or the LTER Executive Board to analyze LTER data access through a single interface.
4. The DAP model fits nicely within the current LTER LDAP user registry used with Metacat for user identification. Other Metacat sites (and their users) would not have to conform to the LTER Data Access Policy, but their users would not be allowed access to LTER data until they register with the LTER LDAP.

6 Disadvantages

1. The DAS model requires sites to change their data access URLs within their EML documents and/or any data references provided to the general community. Although this burden would be the responsibility of the site, the NIS development team could provide conversion scripts that would replace the site data URL with the correct proxy URL on a case-by-case basis. We also envision both a web interface and an application programmable interface (API) that would allow sites to manage their own URL conversions.
2. A new registration interface would be required to collect the necessary Data Access Policy information and store it in the LTER LDAP directory.