



LTER Network Data Access Policy Revision: Report and Recommendations

James Brunt, Peter McCartney, Stuart Gage, and Don Henshaw, October 2004, in response to a request from the LTER Network Information System Advisory Committee; Text amended April 13, 2005 to reflect approval of recommendations by the LTER CC.

*Forward comments to:
James Brunt
LTER Network Office
Department of Biology, MSC03 2020
1 University of New Mexico
Albuquerque, NM 87131-0001
jbrunt@LTERnet.edu*

Table of Contents

Introduction.....	3
Purpose.....	3
Analysis of current LTER Site Data Access Policies	3
Recommendations.....	6
Summary Statement	7
Appendices.....	7

Introduction

This document is a report on work carried out to update the LTER Network Data Access Policy. The current LTER Network Data Access Policy, approved by the coordinating committee in 1997, has been in use since 1990. An analysis of the current policies related to the release, access, and use of LTER data has been undertaken by a sub-committee of the LTER Network Information System Advisory Committee with the goal of standardizing policies across the network. Recommendations for revising LTER Network data policies, especially with respect to those defining requirements for access to LTER data and terms of acceptable use of LTER data have been proposed and a revised LTER data policy has been drafted based on these recommendations and is attached as an appendix to this report.

Purpose

Assessments of the current LTER data access policy by the Information Management committee at its 2004 annual meeting found a number of shortcomings in the current policy that were brought to the attention of the Coordinating Committee (CC) at its annual meeting. A motion was approved at that meeting to charge the IM committee to work with the Network Information System Advisory Committee (NISAC) to revise the policy (1) to better align it with the requirements of the developing Network Information System (e.g., automated data distribution), (2) to reconcile LTER policies with those of other NSF programs such as Polar Programs, and (3) to reaffirm commitment to the LTER Data Access Policy by all 26 member sites.

Reexamination of the policy serves to assure data providers of appropriate use of data sets, while promoting a culture within the ecological community of making data available to the general scientific community in a timely fashion. The analysis led to the following proposals: 1) an expansion of the current LTER Data Access Policy to cover the release (timing of availability of data from a site) of data and metadata, 2) requirements for accessing (retrieving data from a site's repository) LTER data, and 3) the provision for data use agreements (specifying the conditions "some rights reserved" for the use of LTER data by others).

Analysis of current LTER Site Data Access Policies

A draft network LTER policy was created and adopted in 1997 to define the minimum standards for timely data release. The draft was posted online at <http://lternet.edu> and is repeated below. It was intended to be used as a template from which sites would craft their own data release policies that would be compatible with both the Network expectations and the sites individual requirements.

The network recommendations dealt only with the policy for the release of LTER datasets to access by the public and scientific communities. The policy provided no language for protecting the rights of data providers over Type 1 datasets other than allowing that some datasets may be

DATA ACCESS POLICY FOR THE LTER NETWORK

- 1) There are two types of data: Type I (data that is freely available within 2-3 years) with minimum restrictions regarding its use by others and, Type II (Exceptional data sets that are available only with written permission from the PI/investigator(s)). Implied in this timetable, is the assumption that some data sets require more effort to get on-line and that no "blanket policy" is going to cover all data sets at all sites. However, each site would pursue getting all of their data on-line in the most expedient fashion possible.
- 2) The number of data sets that are assigned TYPE II status should be rare in occurrence and that the justification for exceptions must be well documented and approved by the lead PI and site data manager. Some examples of Type II data may include: locations of rare or endangered species, data that are covered by copyright laws (e.g. TM and/or SPOT satellite data) or some types of census data involving human subjects.

available only on a restricted basis (Type II data). It is likely that the lack of such a provision has led to a greater number of datasets classified as Type II than had been anticipated. As sites incorporated this template into their own site data access policy, many attempted to define conditions for acceptable use; this resulted in a considerable range of diversity across the network.

The policy also failed to address criteria that have been used either explicitly or implicitly to except certain datasets from data access obligations, or to rank priorities for release. Such criteria have included exemptions for student collected data or data not collected by LTER, but used in LTER research. This was found to be inconsistent with the review guidelines for Information Management under preparation by the IM committee.

An analysis of the web data access work flows for each of the 24 LTER sites was performed and general components of the policy related to data release, access, and use were noted. A second quantitative pass was performed to determine the number of sites exhibiting each component (Table 1). The analysis indicates that a variety of information types are co-mingled. For example, at several sites there is a general "data management policy" that contains information about "data use" requirements (terms a data user agrees to abide by when downloading data), "data access" requirements (process of registering or providing tracking information in obtaining data), and "data release" policy (policy describing which data sets will be made available and when – data provider requirements) as well as information about the local data management practices. At other sites these components are presented in a variety of approaches. We did find the greatest consistency among the sites in terms of the "data use" policies. For sites that had requirements regarding access to data, there was consistency among the types of information required for data access. Table 1 shows the number of LTER sites that identify specific components of data access and use as site policy.

Table 1. Summary of site data policies.

Data Access and Use Policy Component	# of LTER sites
General Statement of Use	20
Statement of use restricted to legitimate scientific interests	6
Statement of intellectual property rights	2
Statement of code of ethics	2
Statement of Encouraged Collaboration, Consultation, and Co-Authorship	7
Statement of precedence of site activities	3
Statement of no redistribution	10
Acknowledgment of LTER and NSF in derivative works	14
Suggested acknowledgment	13
Requirement for citation in derivative works	14
Suggested citation	1
Requirement for communication of reprints of published work	15
Requirement for preprint/manuscript of derivative works	5
Requirement for notification of PI/contact upon accessing data	6
Requirement for communication of improvements	1
Statement of action to be taken in violation of use terms	2
Disclaimer	9
Statement of QA/QC	1
Statement of Human Subjects Requirements	2

As LTER sites began to provide interactive web applications to disseminate their online data, many took advantage of the medium to request information from users for the purpose of tracking their own use accounting and to protect intellectual rights of the data providers. Again, in the absence of network standards, considerable diversity in the kind and extent of requirements for accessing data were observed (Table 2).

Table 2. Classification of data access requirements by 24 LTER site data systems:

Data Access Requirement	# of LTER sites
No Requirements	12
Voluntary Registration	2
Required Registration	10
Name	11
Email Address	11
Full Contact Information (may include affiliation, mailing address, phone, fax)	10
Agreement to Data Use Policy	9
Requirement of Statement of Use	6
Authentication (reusable login)	4

Finally, data release policies associated with following NSF areas were reviewed; 1) Division of Ocean Sciences, 2) Office of Polar Programs, 3) Behavioral and Cognitive Sciences, and 4) Division of Earth Sciences. These policies ranged from the very general encouragement of sharing to the very specific. Although there are a number of domain specific requirements regarding sharing and submission to particular public archives, the only general provision that should be added to the LTER Network policy to keep it inline with the general guidelines of these policies is the requirement for submission of metadata on datasets within 60 days of data collection (for continuous observations metadata should be submitted periodically if there is significant change in sampling).

Recommendations

Based on this analysis of LTER site policies and requirements in use today, we have generalized across the broadest cross-section of these policies to recommend new draft policies for data release, data access, and data use. The adoption and use of these revised policies by the LTER Network is motivated by the following needs:

- to ensure that the data release, data access, and data use language is consistent and clear across the network,
- to standardize the approach to data access requirements for automated data distribution,
- to standardize the requirements for data provision and use, and
- to ensure that the LTER Network data policies are consistent and inter-operable with those of other NSF-funded programs.

We recommend that the LTER Network Data Access Policy be revised to include three main components. A proposed draft of this document is provided in Appendix I. First, an **LTER Network Data Release Policy** based on the language of the existing policy should include stronger guidelines for timely release and should specify obligations for the release of metadata consistent with the NSF OCE, OPP policies. This will partition the issue of requirements for “accessing” LTER data from the timing of the availability of LTER data. Secondly, although fully half of the sites imposed no requirements for the access of data, we recommend the use of an **LTER Network Data Access Requirements**, that standardizes access requirements and can be implemented via software agents to provide sites with the confidence and tracking mechanisms they need for accounting and reporting data use, while not causing excessive obstacles to accessing data. Thirdly, we recommend the LTER Data Access policy stipulate that all datasets released by LTER be accompanied by metadata that specifies terms of use in the form of a user agreement. A **General Public Data Use Agreement** for use with all Type I data is drafted and included as Appendix II. It is recommended that sites use this document as a template for crafting a **Restricted Data Use Agreement** for Type II data, specifying the unique restrictions and their reasons. This agreement may require additional supporting documents depending on nature of the restrictions. Note: The NISAC working group is currently investigating the work of “Creative Commons” for applying the concept of “some rights reserved” to publicly accessible data and may modify the draft data use agreement to align it more with this concept (<http://science.creativecommons.org>) which is geared specifically towards the products of scientific innovation. Science Commons works within current copyright and patent law to promote legal and technical mechanisms that remove barriers to sharing.

Summary Statement

In summary, this document is developed to encourage access and use of LTER Data Sets. Due to the complexity of the Sites, the flow of scientific expertise through the LTER infrastructure and the scientific value of long term information collected by site personnel it is of significant value to have a standard policy for data access and use. LTER promotes the adoption of the recommendations of this document by the broader ecological research community.

Amendment (April 13, 2005)

The appended data access policy and general data use agreement have been voted on and unanimously approved for adoption by the LTER Network Coordinating Committee in the course of their meeting April 6-7, 2005 in Key Largo Florida.

Appendices

- I. Draft Revised LTER Data Access Policy Version 3
- II. Draft General Data Use Agreement

Appendix I: Long Term Ecological Research Network Data Access Policy - Version 3

The LTER data policy includes three specific sections designed to express shared network policies regarding the release of LTER data products, user registration for accessing data, and the licensing agreements specifying the conditions for data use.

LTER Network Data Release Policy

Data and information derived from publicly funded research in the U.S. LTER Network, totally or partially from LTER funds from NSF, Institutional Cost-Share, or Partner Agency or Institution where a formal memorandum of understanding with LTER has been established, are made available online with as few restrictions as possible, on a nondiscriminatory basis. LTER Network scientists should make every effort to release data in a timely fashion and with attention to accurate and complete metadata.

Data

There are two data types:

Type I – data are to be released to the general public according to the terms of the general data use agreement (*see Section 3 below*) within 2 years from collection and no later than the publication of the main findings from the dataset and,

Type II - data are to be released to restricted audiences according to terms specified by the owners of the data. Type II data are considered to be exceptional and should be rare in occurrence. The justification for exceptions must be well documented and approved by the lead PI and Site Data Manager. Some examples of Type II data restrictions may include: locations of rare or endangered species, data that are covered under prior licensing or copyright (e.g., SPOT satellite data), or covered by the Human Subjects Act. Researchers that make use of Type II Data may be subject to additional restrictions to protect any applicable commercial or confidentiality interests.

While the spirit of this document is to promote maximum availability for ecological data in either Type I or II status, there are criteria by which priority for data release may be determined. Primary observations collected for core research activities directly supported by LTER research must receive the highest priority for data release. Data collected by other sources to which LTER supported research has added value is also a high priority. Other types of data including non-LTER data that was acquired for LTER research, student thesis data, schoolyard LTER data, or legacy data that already suffer from inadequate documentation or format obsolescence may be ranked a lower priority by a site with justifications provided in their data management policy. Finally, some data may be determined of lowest priority for archiving on the grounds that they are interim data that led to final products that carry the scientific value. These might include data files created during stages within an analytic workflow, raw or replicate data values that were subsequently aggregated or processed for release, or individual outputs from stochastic models.

Metadata

1. Metadata documenting archived/online data sets of all types listed above will be made available when, or before, the dataset itself is released according to the terms above.
2. All metadata will be publicly available regardless of any restrictions on access to the data.
3. All metadata will follow LTER recommended standards and will minimally contain adequate information on proper citation, access, contact information, and discovery. Complete information including methods, structure, semantics, and quality control/assurance is expected for most datasets and is strongly encouraged.

1. LTER Network Data Access Requirements

The access to all LTER data is subject to requirements set forth by this policy document to enable data providers to track usage, evaluate its impact in the community, and confirm users' acceptance of the terms of acceptable use. These requirements are standardized across the LTER Network to provide contractual exchange of data between Site Data Providers, Network Data Providers, and Data Users that can be encoded into electronic form and exchanged between computers. This will allow direct access to data via a common portal once these requirements have been fulfilled. The following information may be required directly or by proxy prior to the transference of any data object:

- Registration
 1. Name
 2. Affiliation
 3. Email Address
 4. Full Contact Information
- Acceptance of the General Public Use Agreement or Restricted Data Use Agreement, as applicable.
- A Statement of Intended Use that is compliant with the above agreements. Such statements may be made submitted explicitly or made implicitly via the data access portal interface.

Data providers wishing to impose further requirements beyond these are encouraged to include them in their Restricted Data Use Agreements accompanying the datasets.

- **Data Use Agreements**

Datasets released by LTER sites or the network will be accompanied with a use agreement that specifies the conditions for data use. For Type I data, this shall be the General Data Use Agreement (see appendix II). This document specifies general roles and the obligations and rights enjoyed by each regarding the use of most dataset released for general public use. For Type II datasets, a Restricted Data Use Agreement must be provided with the dataset that identifies the specific restrictions on the use of the data and their justification. Because these are expected to be unique to the dataset, no template is provided although in most cases the General Data Use Agreement can be modified to serve. Grounds for restricting data may include the need to restrict access to species, habitats or cultural resources protected by legislation; rights of privacy granted by human subjects legislation; or protection of intellectual, financial or legal rights over the data held by a third party.

4. This policy becomes effective when approved by the LTER Network Coordinating Committee. It may be revised by, or at the request of, the same body.

Appendix II: General Data Use Agreement

Definitions

“Data Set” – Digital data and its metadata derived from any research activity such as field observations, collections, laboratory analysis, experiments, or the post-processing of existing data and identified by a unique identifier issued by a recognized cataloging authority such as a site, university, agency, or other organization.

“Data User” - individual to whom access has been granted to this Data Set, including his or her immediate collaboration sphere, defined here as the institutions, partners, students and staff with whom the Data User collaborates, and with whom access must be granted, in order to fulfill the Data User's intended use of the Data Set

“Data Set Creator” - individual or institution that produced the Data Set

“Data Set Owner” – individual or institution that holds intellectual property rights to the dataset. Note that this may or may not be defined as a legal copyright. If no other party is designated in the metadata as Data Set Owner, it may be presumed that these rights are held by the Data Set Creator.

“Data Set Distributor” - individual or institution providing access to the Data Sets.

“Data Set Contact” - party designated in the accompanying metadata of the Data Set as the primary contact for the Data Set.

Conditions of Use

The re-use of scientific data has the potential to greatly increase communication, collaboration and synthesis within and among disciplines, and thus is fostered, supported and encouraged. Permission to use this dataset is granted to the Data User free of charge subject to the following terms:

- 2) *Acceptable use.* Use of the dataset will be restricted to academic, research, educational, government, recreational, or other not-for-profit professional purposes. The Data User is permitted to produce and distribute derived works from this dataset provided that they are released under the same license terms as those accompanying this Data Set. Any other uses for the Data Set or its derived products will require explicit permission from the dataset owner.
- 3) *Redistribution.* The data are provided for use by the Data User. The metadata and this license must accompany all copies made and be available to all users of this Data Set. The Data User will not redistribute the original Data Set beyond this collaboration sphere.
- 4) *Citation.* It is considered a matter of professional ethics to acknowledge the work of other scientists. Thus, the Data User will properly cite the Data Set in any publications or in the metadata of any derived data products that were produced using the Data Set. Citation should take the following general form: *Creator, Year of Data Publication, Title of Dataset, Publisher, Dataset identifier. For example:*

McKee, W. 2001. Vascular plant list on the Andrews Experimental Forest and nearby Research Natural Areas: Long-Term Ecological Research. Corvallis, OR: Forest Science Data Bank: SA002. [Database]. <http://www.fsl.orst.edu/lter/data/abstract.cfm?dbcode=SA002>. (21 October 2004)

- 5) *Acknowledgement*. The Data User should acknowledge any institutional support or specific funding awards referenced in the metadata accompanying this dataset in any publications where the Data Set contributed significantly to its content. Acknowledgements should identify the supporting party, the party that received the support, and any identifying information such as grant numbers. For example:

Data sets were provided by the Forest Science Data Bank, a partnership between the Department of Forest Science, Oregon State University, and the U.S. Forest Service Pacific Northwest Research Station, Corvallis, Oregon. Significant funding for collection of these data was provided by the National Science Foundation Long-Term Ecological Research program (NSF Grant numbers BSR-90-11663 and DEB-96-32921).

- 6) *Notification*. The Data User will notify the Data Set Contact when any derivative work or publication based on or derived from the Data Set is distributed. The Data User will provide the data contact with two reprints of any publications resulting from use of the Data Set and will provide copies, or on-line access to, any derived digital products. Notification will include an explanation of how the Data Set was used to produce the derived work.
- 7) *Collaboration*. The Data Set has been released in the spirit of open scientific collaboration. Data Users are thus strongly encouraged to consider consultation, collaboration and/or co-authorship with the Data Set Creator.

By accepting this Data Set, the Data User agrees to abide by the terms of this agreement. The Data Owner shall have the right to terminate this agreement immediately by written notice upon the Data User's breach of, or non-compliance with, any of its terms. The Data User may be held responsible for any misuse that is caused or encouraged by the Data User's failure to abide by the terms of this agreement.

Disclaimer

While substantial efforts are made to ensure the accuracy of data and documentation contained in this Data Set, complete accuracy of data and metadata cannot be guaranteed. All data and metadata are made available "as is". The Data User holds all parties involved in the production or distribution of the Data Set harmless for damages resulting from its use or interpretation.