

CLIMATE DATABASE PROJECT: A STRATEGY FOR IMPROVING INFORMATION ACCESS ACROSS RESEARCH SITES

Donald L. Henshaw

U.S. Forest Service Pacific Northwest Research Station, 3200 SW Jefferson,
Corvallis, OR 97331

Maryan Stubbs and Barbara J. Benson

Center for Limnology, University of Wisconsin-Madison, Madison, WI 53706

Karen Baker

Scripps Institution of Oceanography, University of California-San Diego, La Jolla, CA 92093

Darrell Blodgett

Forest Soils Laboratory, University of Alaska, Fairbanks, AK 99775

John H. Porter

Department of Environmental Sciences, University of Virginia, Charlottesville, VA 22903

Abstract. To facilitate intersite research among the network of Long-Term Ecological Research sites, information managers are exploring strategies for linking individual site information systems. A prototype to provide climatic summaries dynamically has been developed and serves as one model for improving access to data across sites. Individual sites maintain local climate data in local information systems while a centralized site continually updates and provides access to all sites' data through a common database. Common distribution report formats have been established to meet specific needs of climate data users.

INTRODUCTION

Information managers associated with the Long-Term Ecological Research (LTER) program have developed a basic foundation for a Network Information System (NIS) with a primary goal of facilitating intersite research (Stafford et al. 1994). To accommodate the needs of various intersite studies and synthesis efforts within the LTER, it is critical to develop dynamic systems for providing comparable data from multiple LTER sites. Improving access and adding query capability to intersite data using network information servers is a major component of current NIS development (Brunt 1996). With each site operating its own information management system, the LTER NIS will employ a variety of strategies in linking these individual systems (Porter et al. 1997).

Climate meteorological data are collected at all LTER sites and are the most frequently requested data. Synthesis groups need ready-access to climatic summaries from multiple sites. A NIS prototype to provide climatic summaries dynamically has been developed and serves as one model for improving access to data across sites. This approach allows individual sites to maintain the local climate data in local information systems while a centralized site continually updates and provides access to all sites' data through a common database.

BACKGROUND

A standards document developed by the LTER Climate Committee (Greenland 1986) established baseline meteorological measurements to characterize each LTER site. Standardized measurements provide a basis for coordinating meteorological measurements at two or more sites and enable intersite comparisons. More recently, a project to conduct climatic analyses of the

LTER sites (CLIMDES; <http://lternet.edu/im/climate/climdes/>) gathered individual site temperature and precipitation data (1960-1990) and created on-line monthly summaries for each site (Greenland et al. 1997). While the CLIMDES project satisfied an immediate need for access to monthly site climate data, the structure provided no method for maintaining and updating these summaries or satisfying frequent requests for daily climate data. Most of the LTER sites had their climate data available on the World Wide Web (WWW), but the data sets were sometimes difficult to find and were formatted and aggregated differently at each site.

The NSF-funded XROOTS (<http://lternet.edu/im/xroots/aclim.htm>) project requires intersite climate data to synthesize belowground productivity using root biomass data from multiple sites. The idea that distribution of data in report formats amenable to users independent of the data storage format was explored in an XROOTS climate workshop (Bledsoe et al. 1996). Two monthly distribution report formats were recommended to accommodate both spreadsheet (V-One, i.e., twelve monthly values for one variable per record) and database (V-Many, i.e., one monthly value for many variables per record) users (See Table 1).

OVERVIEW

As part of the LTER Information Managers' NIS development, the LTER climate database project (ClimDB) has developed a prototype for harvesting daily climate data in a standardized exchange format using the WWW from a subgroup of LTER sites. The harvested data are stored in a centralized relational database. Climate variables include daily minimum, maximum, and mean air temperature and daily precipitation. Applications have been developed initially to generate the two XROOT monthly distribution formats using this centralized database of daily values. Additionally, a webpage (<http://www.limnology.wise.edu/climdb.html>) has been created to provide access to the daily and monthly climate data as well as to permit query by LTER site, weather station, and date.

SPECIFIC EXCHANGE AND DISTRIBUTION FORMATS

Each of the five sites participating in the prototype development process provided climate data files in a standardized daily exchange format at an Internet address (URL). For this model, the site files could be either static or produced by a dynamic script. A comma-delimited format was agreed upon after discussions revealed the diversity of approaches, opinions, and needs among sites. For instance, date can be stored as a single 8-character field, comma separated, or Julian day designated. It is important to note there is not one "right" exchange format. The primary criteria require that individual sites "filter" local site data into the exchange format. The standardized daily exchange format agreed upon is as follows:

Site, station, date, value1, flag1, value2, flag2, value3, flag3, value4, flag4

where,

site	the three-letter LTER site code
station	that site's name for the weather station
date	8-character field, <code>yyyymmdd</code>
value1, flag1	mean air temperature and corresponding flag
value2, flag2	maximum air temperature and corresponding flag
value3, flag3	minimum air temperature and corresponding flag
value4, flag4	precipitation and corresponding flag

All temperature values are reported in degrees Celsius and precipitation in millimeters. Each value has a corresponding data quality flag where flags are coded as follows:

G or blank	value is a good value
E	value is estimated
Q	value is questionable
M	value is missing
T	trace value (for precipitation only)

Here is a brief example of the daily format from the Andrews Forest (AND) site's Primary Meteorological Station (PRIMET) aligned for readability:

```
AND, PRIMET, 19960101, 6.8, , 10.8, Q, 4.5, , 0.0, T
AND, PRIMET, 19960102, 5.3, , 10.6, Q, 0.8, , 4.3,
AND, PRIMET, 19960103, 7.7, , 9.7, , 4.1, , 20.6,
AND, PRIMET, 19960104, 4.2, , 6.7, , 2.4, , 11.4,
AND, PRIMET, 19960105, 4.8, E, 7.4, E, 2.7, E, , M
AND, PRIMET, 19960106, 5.7, E, 9.7, E, 1.3, E, , M
```

Daily climate data from all sites are harvested automatically from the local sites using a simple script calling the WWW line mode browser. An example of the harvest command line for the Andrew's Forest climate data is:

```
www -n -source http://www.fsl.orst.edu/leter/webmast/and_clim.txt >and.dat
```

Data are stored in a relational database at the centralized site. [Note: Currently, the prototype is using the Oracle™ database management system and the centralized site is the North Temperate Lakes LTER Site. Eventually the ClimDB project will move to the LTER Network Office, and the relational database software may change.] Application programs produce two monthly distribution tables (See Table 1).

A webpage allows the user to query for daily data in addition to providing the two monthly tables. Monthly summary values are displayed along with the number of valid daily values included in the summary. Missing and questionable values are excluded from summary values. Listing the number of valid data values used in calculating a monthly value gives the user some assurance about the value's accuracy and represents a valuable addition to any distribution format.

METADATA

Every meteorological station will be described in a central metadata database. An entity-relationship diagram (See Figure 1) shows the proposed schema for the metadata database. The metadata database is currently being developed in Oracle™. LTER-site-level information, individual station descriptions, and specific measurement documentation form the three major entities. Standardized web forms will be used to collect this information from participating sites. Metadata term definitions will be made available on the central webpage. Metadata will be critical for intersite studies in evaluating key differences in site descriptions and methodology.

Table 1. Examples of the two monthly distribution tables (V-One and V-Many) are shown for the Andrews Forest (AND) site's Primary Meteorological Station (PRIMET). The “#” indicates the number of valid daily values (including estimated values) that were used in calculating the monthly summary value.

V-One. V-One displays one variable per table and is primarily intended for use in spreadsheets. These two abbreviated examples show mean monthly air temperature and total precipitation.

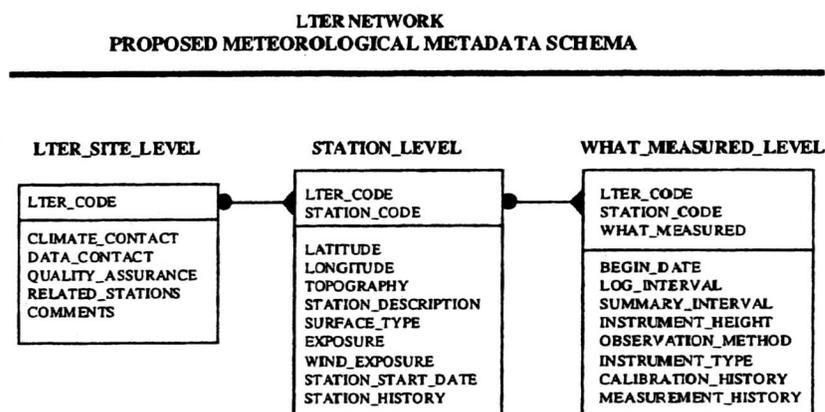
AND PRIMET		Avg_mean_air_temp_c													
Year	Jan	#	Feb	#	Mar	#	Apr	#	May	#	...	Nov	#	Dec	#
1991	0.1	31	5.8	28	4.5	31	6.9	30	10.0	31	...	6.5	30	3.2	31
1992	3.3	29	5.8	29	8.1	30	10.0	30	15.0	31	...	5.0	30	1.0	31
1993	-0.6	31	0.6	28	6.0	31	7.7	30	13.2	31	...	-0.8	30	-0.2	30

AND PRIMET		Totl_precip_mm													
Year	Jan	#	Feb	#	Mar	#	Apr	#	May	#	...	Nov	#	Dec	#
1991	232	31	208	28	221	31	242	30	195	31	...	451	30	214	31
1992	160	31	201	29	40	31	290	30	20	31	...	337	30	419	31
1993	242	31	95	28	354	31	394	30	237	31	...	103	30	278	31

V-Many. V-Many displays many variables per table and is primarily intended for use in relational databases. This example includes all four prototype variables of monthly mean, maximum, and minimum air temperature and total monthly precipitation.

AND PRIMET									
Year	Month	Mean	#	Max	#	Min	#	Ppn	#
1991	Jan	0.1	31	5.3	31	-3.0	31	232	31
1991	Feb	5.8	28	12.4	28	2.1	28	208	28
1991	Mar	4.5	31	11.2	31	0.3	31	221	31
1991	Apr	6.9	30	13.3	30	2.6	30	242	30

Figure 1. Proposed schema for the metadata database



CONCLUSIONS

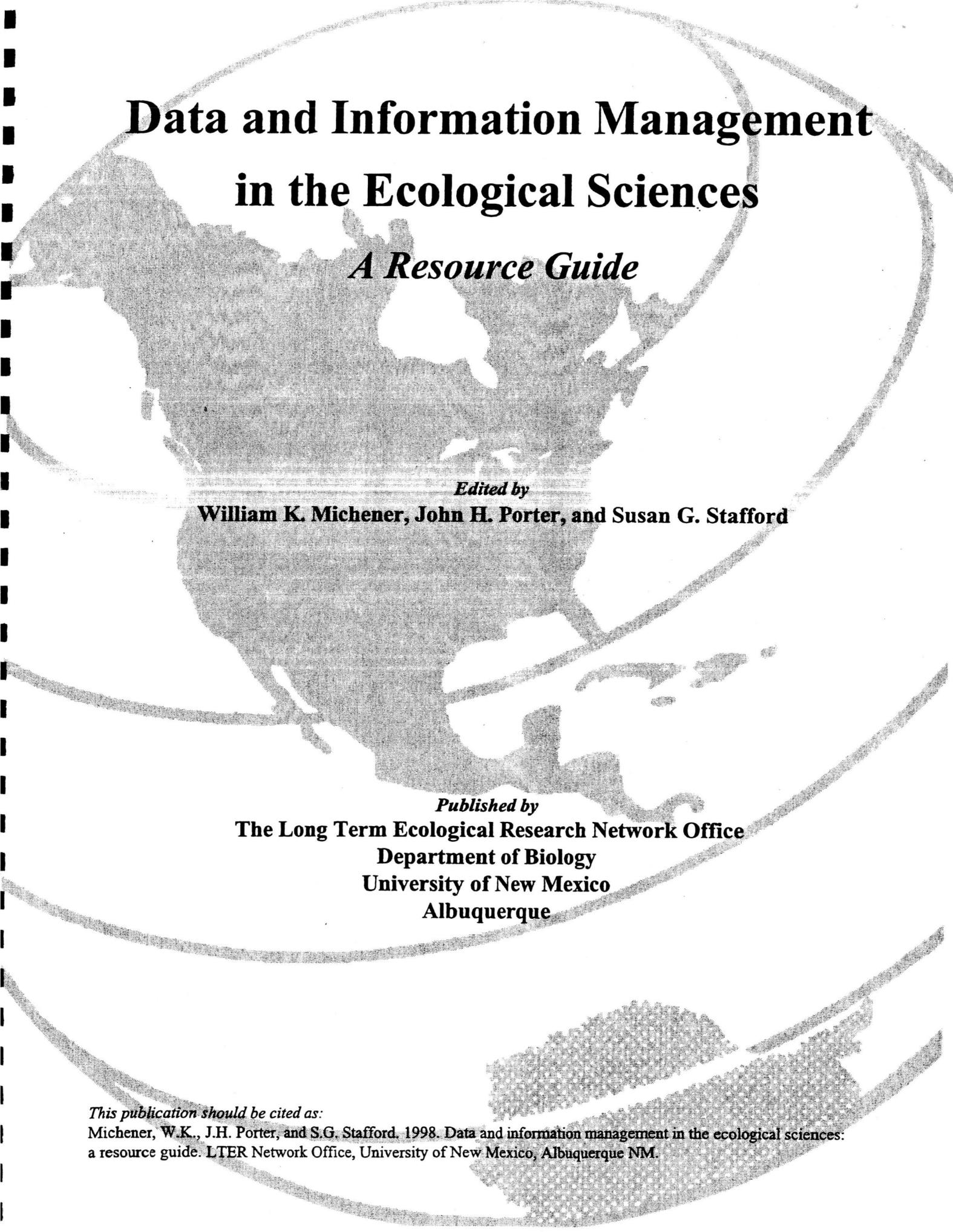
With an increasing focus on intersite activities within the LTER program, the LTER Information Managers are developing a Network Information System to facilitate intersite research. This LTER NIS prototype for climate data will serve as a model for other intersite data set integration efforts. The approach allows for the diversity in information management systems across the LTER network. Data sets are distributed across multiple sites, but are accessible in common distribution formats from a central site. Specially formatted distribution reports have been established to meet specific needs of climate data users, but the design is extensible in that it permits update with additional formats as the need arises.

ACKNOWLEDGMENTS

The authors would like to acknowledge contributions from the North Temperate Lakes LTER site (DEB 9632853) for participating in the development of this prototype and for supporting the centralized database and web pages. Contributions from the H. J. Andrews Experimental Forest (NSF grant DEB 9632921), the Bonanza Creek Experimental Forest (DEB 9211769), Palmer Station (OPP 9632763), and the Virginia Coast Reserve LTER (DEB 9411974) sites are also recognized for participating in the development of this prototype. LTER sites are funded all or in part by the National Science Foundation. We also wish to acknowledge the efforts of Caroline Bledsoe for her strong support and continued interest in this project.

LITERATURE CITED

- Bledsoe, C., J. Hastings, and R. Nottrott. 1996. Xclimate workshop. Davis, CA.
<http://lternet.edu/im/xroots/aclim.htm>
- Brunt, J. W. 1996. Developing an LTER network information system for the 21st century.
<http://lternet.edu/is/is18Jan96.htm>
- Greenland, D., T. Kittel, B.P. Hayden, and D.S. Schimel. 1997. A climatic analysis of Long-Term Ecological Research sites. *<http://lternet.edu/im/climate/climdes/>*
- Greenland, D. 1986. Standardized meteorological measurements for Long-Term Ecological Research sites. *Bulletin of the Ecological Society of America* 67:275-277.
lternet.edu/im/climate/standard86.htm
- Porter, J., D.L. Henshaw, and S.G. Stafford. 1997. Research metadata in Long-Term Ecological Research (LTER). Proceedings of the Second IEEE Metadata Conference. Silver Spring, MD.
http://www.computer.org/conferen/proceed/meta97/list_papers.html
- Stafford, S.G., J.W. Brunt, and W.K. Michener. 1994. Integration of scientific information management and environmental research. Pages 3-19 in W. K. Michener, J. W. Brunt and S. G. Stafford editors. *Environmental information management and analysis: ecosystem to global scales*. Taylor & Francis, Bristol, PA.



Data and Information Management in the Ecological Sciences

A Resource Guide

Edited by

William K. Michener, John H. Porter, and Susan G. Stafford

Published by

**The Long Term Ecological Research Network Office
Department of Biology
University of New Mexico
Albuquerque**

This publication should be cited as:

Michener, W.K., J.H. Porter, and S.G. Stafford. 1998. Data and information management in the ecological sciences: a resource guide. LTER Network Office, University of New Mexico, Albuquerque NM.