

# INFORMATION ACCESS AND DATABASE INTEGRITY AT THE NORTH TEMPERATE LAKES LONG-TERM ECOLOGICAL RESEARCH PROJECT

Barbara J. Benson and Maryan Stubbs

Center for Limnology, University of Wisconsin-Madison,  
680 N. Park Street, Madison, Wisconsin 53706

*Abstract.* The North Temperate Lakes LTER data and information system is designed to facilitate ecological research. The primary information management goals focus on information access and database integrity. The use of relational database software is discussed in the context of these two goals.

## INTRODUCTION

The North Temperate Lakes Long-Term Ecological Research (NTL-LTER) program (Magnuson et al. 1991) was established in 1981 by the National Science Foundation as part of the LTER network of research sites (Callahan 1984, Swanson and Franklin 1988). The NTL-LTER project has two study areas in Wisconsin where patterns, processes, and interactions of lakes and their surroundings are examined at a nested set of spatial and temporal scales.

Data management is an integral component of the NTL-LTER project (Benson 1996). The design of an information system must be based upon the research agenda. NTL-LTER has research goals (Table 1) which require diverse data sets, linkages among data sets, and multiple spatial scales. Expansion of our research agenda has included regional-scale investigations and the study of human interactions with lake ecosystems.

Table 1. NTL-LTER research goals.

- |   |
|---|
| <ol style="list-style-type: none"><li>1) Perceive long-term changes in the physical, chemical, and biological properties of lake ecosystems.</li><li>2) Understand interactions among physical, chemical, and biological processes within lakes and their influences on lake characteristics and long-term dynamics.</li><li>3) Develop a regional understanding of lake ecosystems through an analysis of the patterns and processes organizing lake districts.</li><li>4) Develop a regional understanding of lake ecosystems through integration of atmospheric, hydrologic, and biotic processes.</li><li>5) Understand the way human, hydrologic, and biogeochemical processes interact within the terrestrial landscape to affect lakes and the way lakes, in turn, influence these interactions.</li></ol> |
|---|

The NTL-LTER data and information system has been designed to facilitate interdisciplinary research. Our primary goals have been 1) to create a powerful and accessible environment for the retrieval of information that facilitates linkages among diverse data sets and 2) to maintain database integrity.

## INFORMATION ACCESS

To provide an optimal environment for information access, we required the following criteria be met: 1) the data structures facilitate queries, 2) the client interfaces are easy to use, and 3) adequate metadata are available to permit data interpretation. We implemented our information system using Oracle™ database software on a Sun Ultrasparc™ 2. Other LTER sites have built successful information systems without using relational databases; however, there were some strong reasons to use a relational database for our data. The relational structure is well-suited for queries and facilitates linking data sets. Relational databases can handle large, complex data sets. Database structures can be easily expanded or changed. Normalizing data tables can eliminate data redundancy. Built-in security, recovery and export capabilities create a more secure environment for access, updates, and backup. Through the use of SQL (Celco 1995, Date 1997, Ladanyi 1997) scripts, procedures can be saved, therefore, documented and reused. Our system, Oracle™ is multi-user and multi-tasking.

Industrial-strength relational databases such as Oracle™ can be rather expensive to purchase unless your organization has special terms with the vendor. These products are very large and complex, with numerous configuration options, and there tends to be a significant learning curve to utilize database features fully. However, benefits include full concurrency control, recoverability, high performance, and stable vendor presence in the field.

Researchers at NTL-LTER can use an end-user query tool (Oracle Discoverer 2000™, formerly called Oracle Data Browser™) to retrieve data from the database. This point-and-click interface is being used routinely by researchers to obtain exactly the data sets of interest. Joining of tables, aggregation, and sorting can be performed with this tool.

An alternative access method is through the World Wide Web (WWW). On-line data sets can be accessed through the data catalog (<http://limnosun.limnology.wisc.edu/datacat.html>). These data sets are text files retrieved from the Oracle database with metadata at the top of each file. However, maintaining the data then requires maintaining both the database and the retrieved text files. We are now developing dynamic query capability from the WWW to the Oracle™ database (Stubbs and Benson 1996). In addition to avoiding maintaining text files, these dynamic queries also permit more powerful information retrieval for the user.

Currently, we have implemented dynamic queries for meteorological data (<http://limnosun.limnology.wisc.edu/climate.html>). The user can select parameters to retrieve and specify the time period of interest. It is also possible to generate summaries over a specified time period (e.g., total precipitation or mean air temperature) or to graph a parameter over time.

## DATABASE INTEGRITY

Maintaining the integrity of a database requires controlling the access for writing to the database. In addition, the database needs to be protected by an adequate backup system and be recoverable. The data in the database must have been subjected to quality control/quality assurance protocols. File format and storage media need to be addressed to guarantee useable long-term archiving.

The Oracle™ database software provides considerable functionality for database integrity issues. Setting up passwords, privileges, and roles controls read and write access. Oracle export utilities can be used to backup the database and protect against accidental deletion or incorrect updating of a table or, if necessary, be used to restore the entire database.

Quality control mechanisms have been established including random blind samples and replicate analyses. The data entry software has some built-in error checking, and a two-person team proofreads entered data. Finally, researchers review summary tables and further error checks

are performed, such as ion balances and calculation of critical parameters, from a redundant data set.

## FUTURE DIRECTIONS AND CHALLENGES

The NTL-LTER project like many other ecological research programs is being challenged by new types and expanded volumes of data as the scope of research expands and new technology affects measurement. The expansion of the research program to include human interactions with lake ecosystems is generating the need to incorporate new types of data sets into the database. The data management staff is interacting with social scientists to design database tables and provide metadata for a growing collection of new data sets including land ownership, census, and attitude survey data. The increased volume of spatial data, especially from satellite-based sensors, requires that the ecological science community be prepared to use these data and that appropriate data management be in place.

The use of the WWW to distribute data will expand as we continue to construct query functionality from the WWW to the Oracle database. We also plan to use the WWW interface for data entry directly into the database. This interactive data entry will be designed to provide immediate feedback on entry errors.

## ACKNOWLEDGMENTS

Financial support was provided by the National Science Foundation's Long Term Ecological Research Program, grant # DEB-96-32853.

## LITERATURE CITED

- Benson, B. J. 1996. The North Temperate Lakes LTER research information management system. Proceedings of the Eco-Informa Workshop, Global Networks for Environmental Information, November 4-7 1996, Lake Buena Vista, Florida, 11: 719-724.
- Callahan, J. T. 1984. Long-term ecological research. *BioScience* 34:363-367.
- Celko, J. 1995. *SQL for smarties: advanced SQL programming*. Morgan Kaufmann Publishers, Inc. San Francisco, CA.
- Date, C.J. 1997. *A guide to the SQL standard fourth edition*. Addison-Wesley Longman.
- Ladanyi, H. 1997. *SQL unleashed*. Sams Publishing. Indianapolis, IN.
- Magnuson, J. J., T. K. Kratz, T. M. Frost, C. J. Bowser, B. J. Benson, and R. Nero. 1991. Expanding the temporal and spatial scales of ecological research and comparison of divergent ecosystems: roles for LTER in the United States. Pages 45-70 in P.G. Risser, editor. *Long-term ecological research*. John Wiley & Sons Ltd. New York, NY.
- Stubbs, M. and B. J. Benson. 1996. Query access to relational databases via the World Wide Web. Proceedings of the Eco-Informa Workshop, Global Networks for Environmental Information, November 4-7 1996, Lake Buena Vista, Florida, 11:105-109.
- Swanson F. J. and J. F. Franklin. 1988. The long-term ecological research program. *Eos* 69(3):34, 36, 46.

