#2393

# TRAINING ENVIRONMENTAL INFORMATION MANAGERS OF THE FUTURE*

S.G. Stafford
Department of Forest Science, Oregon State University
Corvallis, Oregon, USA

J.W. Brunt
Department of Biology, University of New Mexico
Albuquerque, New Mexico, USA

B.J. Benson
Center for Limnology, University of Wisconsin-Madison
Madison, Wisconsin, USA

## ABSTRACT

Complex issues confronting scientists and policymakers require interdisciplinary collaboration and synthesis at much larger spatial and temporal scales than are typical in traditional ecological studies. In this complex research environment, all scientists need some background in bio-informatics and computational ecology. Here, we focus on the more intensive training needed for environmental information managers. "Scientific information management" is emerging as a discipline that emphasizes the effective transformation of data into information that is used to address these complex issues. Synthetic, data-intensive projects will be even more common in the future and will require trained information managers who have skills beyond the custodial and archival functions that data managers have had in the past. We outline an environmental information management training program† that will focus on the link between science and technology. Students will complete core curriculum modules offered at the University of New Mexico and Oregon State University and will then do internships at participating Long-Term Ecological Research sites.

---

* Presented at Eco-Informa '96, Lake Buena Vista, Florida, 4-7 November 1996.

† Support for the conceptual development of this training program has been provided by National Science Foundation grants DEB-9011663, DEB-9411976, and DEB-9011660.

## 1.0 INTRODUCTION

Complex issues confronting scientists and policymakers (e.g., global change, sustainability, biodiversity, and emerging diseases) require interdisciplinary collaboration and synthesis at much larger spatial and temporal scales than are typical in traditional ecological studies. In order to effectively formulate questions, coordinate data collection, manage and access electronic information, and exchange complex data sets across disciplines, cultures, and scales, information managers must be both technologically adept and strongly grounded in environmental science. These individuals must be able to create a cohesive information management strategy that incorporates analytic and technical developments in science.

A sociology of information management is developing that is as critical as are hardware and software issues. We see "scientific information management" emerging as a discipline with both a management and a research component, emphasizing the timely and effective transformation of data into information that is used by scientists, managers, policymakers, and the public (Stafford et al., 1994).

### 1.1 ENVIRONMENTAL DATA

Environmental data have high volume and are structured, distributed, and heterogeneous (Silberschatz et al., 1990). Each of these properties contributes in its own way to making the representation and storage of environmental data very difficult. Massive volumes of data have been collected, and automated sensing technology has accelerated the rate of accumulation. For data to be accessible to a wide community of researchers and to be used and interpreted accurately, their supporting information (i.e., metadata) must be available. Distributed environmental data are typically collected and analyzed in separate locations. Environmental data also tend to be quite heterogeneous (Scheuermann and Yu, 1989) and include diverse data types as well as measurements at multiple scales. Even within a single research site, each Principal Investigator represents and processes data using his/her preferred hardware and software platform. One may prefer Excel on a PC, another may prefer to use SAS on a Mac, and still another may prefer S-plus on any platform.

### 1.2 THE CHALLENGES

Questions about global change, biodiversity, sustainability, and emerging diseases are being increasingly directed at ecologists, but such questions demand research and synthesis at larger scales than are typical in ecological studies. One way ecologists are beginning to address questions at larger scales is through collaborative, multidisciplinary research programs. Interdisciplinary cross-site research and global environmental issues have created new challenges for information management. For example, the Long-Term Ecological Research (LTER) Network*(http://lternet.edu)* consists of 18 widely geographically separated sites, each of which is concerned with the phenomena occurring at its individual site, but each of which also contributes to an overall understanding of ecological change in the US (LTER, 1988). The goals of the LTER program are to conduct long-term, continuous measurements and analyses of ecological patterns and processes at specific sites, to integrate and synthesize results within and among sites, and to seek ways to generalize these results over broader spatial scales.

As another example, **MODIS** Land Science Team & **L**ong-Term **E**cological **R**esearch Network **S**ynthesis (MODLERS) is a NASA-funded project that was initiated to validate Earth Observation System-era global data sets at the local scale*(http://atlantic.evsc.virginia.edu/~jhp7e/modlers/)*. Scientists at each of 14 LTER Network sites are developing local maps of three biosphere variables (landcover class, leaf area index, and aboveground net primary productivity) for a 100-$km^2$ area at a fine-grain size of 25 m; this effort incorporates extensive ground data sets, ecosystem models, and remotely sensed imagery. The fine-grain site maps are spatially aggregated to a coarse grain (1 km) so that they can be compared to coincident portions of the MODLAND Science Team's global maps of the same three biosphere variables. The MODLERS project presents an opportunity to grapple with one of the most vexing current problems in ecology: how scaling from a fine grain to a coarse grain affects estimates of important biosphere variables. Comparisons are being made among the multiple sites and biomes and between the MODLAND maps and the site maps at each grain size.

These types of synthetic, integrative, data-intensive projects, which are dependent upon technology and a working, open information system, will be more common in the future and will require us to retrain our information managers.

## 1.3 INFORMATION MANAGERS

As projects on ecological issues become more complex and the supporting computer technology rapidly advances, the role of the information manager is changing. In the past, data management was viewed as primarily custodial, and data managers were primarily librarians and archivists. Increasingly, however, the focus is on information, not only data sets, and information managers have become research facilitators. Information managers must be trained in methods of analysis of scientific data: statistics, modeling, image analysis, and data visualization. In addition, they must draw upon a whole suite of skills if they are to be effective. These skills involve different kinds of responsibilities and tasks, including study design, data structures, QA/QC, information access involving multiple scales and data types, metadata, security, intellectual property rights, database integrity, backup procedures, long-term access, and an impressive array of technical skills such as GIS, GPS, and internet tools.

Long-term ecological research requires these different information skills, which emphasize the link to science and ecology rather than computer science in a vacuum. Bridges must be built between researchers and data managers to close the gaps between science and technology. Being able to formulate the question(s), collaborate with colleagues, and manage information at larger scales requires a new breed of scientist — "interpreters" (Stonebraker, 1994) — who specialize in communicating about data and methodologies across spatial scales, disciplines, and cultures. Technically adept individuals who have a grounding in both the science as well as the technical aspects of information management are in short supply.

## 2.0 TRAINING PROGRAM

Currently there is no one place to learn and build the requisite set of skills to tackle environmental information management. The LTER Network is well-positioned to provide a training ground: information management at each site has objectives in common with other sites in the network, yet is subject to a range of site-specific conditions. We outline a formal curriculum that capitalizes on the strength and depth of disciplinary expertise across the LTER Network and that produces a suite of information managers with skills required to function in this new science environment.

### 2.1 CURRICULUM DEVELOPMENT

We outline a training program with two broad multidisciplinary themes:

- *Bio-informatics:* Strategies to accommodate the ever-increasing influx of information and to manage research information for the long term. Topics include information science, connectivity, database development, and geographic information systems.
- *Computational Ecology:* Techniques to answer questions at increasingly larger scales, such as modeling, temporal/spatial scaling, remote sensing, and statistics — data visualization technologies.

The curriculum will be designed to train information managers on the increasingly technical aspects of long-term environmental research. The training program will center around a series of core curriculum modules that focus on the link between science and technology. The curriculum will include two 11-week courses already being offered at the University of New Mexico (UNM) and Oregon State University (OSU). UNM's Research Information Management course includes the following topics: Statistics (to be further developed at OSU); Introduction to Practical Data Management; Data Base Management Systems; Programming Languages; Statistical Languages (SAS to be further developed at OSU); Field Practicum; Automated Data Collections; QA/QC; and Computational Workshop.

The second module of the core curriculum is offered at OSU in a course called Natural Resources Data Analysis, which includes the following topics: Terminology and Basic Concepts; SAS (further developing Statistical Languages from UNM); Research Information Management (highlight of key points from UNM); Regression Analysis; Principles of Sound Experimental Planning; Further Experimental Designs; Separation of Means; Testing Assumptions and Alternative Methods; Analysis of Covariance; Repeated Measures; and Presentation of Results.

Additional modules would encompass study planning, "messy data" statistical techniques, hands-on experiences with other statistical software packages (S-plus), experience analyzing research data, distributed databases, relational data structures, GIS, GPS, networking, and wide-area information servers. The remainder of coursework would be structured to best achieve the students' scientific objectives. Students without a strong background in ecology will need to take coursework in this area. Students in this program will emerge with a strong foundation in ecology, computer science, and statistics, as well as with a suite of skills that will allow them to specialize in scientific and technical aspects of information management.

We envision providing training in a variety of formats, tailoring the instructional approaches both to the content of the course as well as to the needs of the students. For example, lectures may be combined with interactive computer exercises and electronic exchange of information as well as group sessions that stress critical thinking and problem-solving skills. Courses may range from one quarter to only a few days in length and may be offered during the week, in the evening, or on the weekend as well as on-campus, on-site, off-site, or electronically.

## 2.2 THE LTER NETWORK: THE TRAINING GROUND

The proposed training program will be built on the knowledge and experiences of scientists associated with the US LTER Program. The mission of the LTER Network is to conduct and nurture ecological research by:

- Understanding general ecological phenomena that occur over longer temporal and spatial scales.
- Creating a legacy of well-designed and well-documented long-term experiments and observations for use by future generations.
- Conducting major synthesis-testing hypotheses across regions and building theoretical models to explain empirical results.
- Providing information to identify and solve societal problems.

The LTER Program provides an excellent training ground in which to study issues related to the management of long-term environmental information. Research information management at the US LTER sites is carried out with common objectives, but the approaches taken to meet these objectives vary among the heterogeneous sites. Meeting standardized goals with a variety of solutions has built strength into this distributed system of information management.

It is essential that students also have a strong grounding in the application of information technology to ecological research. Therefore, after completing the core curriculum at UNM and OSU, students would do internships in their choice of ecological projects at any participating LTER site or other collaborating site, including the San Diego Supercomputer Center, other long-term ecological research sites, Long-Term Research in Environmental Biology (LTREB) sites, and International Long-Term Ecological Research (ILTER) sites. The internships would give students experience in applying information management skills in an ecological research setting within the US or internationally. A "collaboratory" using white boards and video-conferencing techniques would be used to share ideas and keep in communication with one another. Students would work with the designated LTER data manager and other interested scientists at each site. In addition, students would participate in design meetings, workshops, and symposia at their site as well as other training sessions sponsored by the training program organizers.

## 3.0 REFERENCES


LTER, *1990s Global Change Action Plan Utilizing a Network of Ecological Research Sites*, Report of the Long-Term Ecological Research Network Workshop, LTER Network Office, University of Washington, Seattle, WA, 1988.

P. Scheuermann and C. Yu, editors, *Report of the Workshop on Heterogeneous Database Systems*, NSF Report, Northwestern University, Evanston, IL, 1989.

A. Silberschatz, M. Stonebraker, and J.D. Ullman, editors, *Database Systems: Achievements and Opportunities*, The Report of the NSF Invitational Workshop on the Future of Database Systems Research, Palo Alto, CA, 1990.

S.G. Stafford, J.W. Brunt, and W.K. Michener, "Integration of Scientific Information Management and Environmental Research." In *Environmental Information Management and Analysis: Ecosystems to Global Scales*, eds. J.W. Brunt and S.G. Stafford, Taylor & Francis, London, pp. 3-19, 1994.

M. Stonebraker, "Sequoia 2000 — A Reflection on the First Three Years." In *Seventh International Working Conference on Scientific and Statistical Database Management*, eds. J. French and H. Hinterberger, Charlottesville, VA, pp. 108-116, 28-30 September 1994.

*Proceedings of*

# Eco-Informa '96

***Global Networks for Environmental Information***

**Volume 10**

4-7 November 1996
Lake Buena Vista, Florida
USA

Published by
Environmental Research Institute of Michigan (ERIM)
P.O. Box 134001, Ann Arbor, MI 48113-4001, USA

The papers appearing in this two-volume set constitute the proceedings of *Eco-Informa '96: Global Networks for Environmental Information*. They reflect the authors' opinions and are published as received. Their inclusion in this publication does not constitute endorsement by the Environmental Research Institute of Michigan.



ISBN 0-9603590-5-2 (Volume 10)

ISBN 0-9603590-7-9 (2 Volume Set)

Library of Congress Catalog Card Number 96-86738

Printed in the United States of America