Chapter 16

Recommendations for Data Accessibility

C.M. Laney, K.S. Baker, D.P.C. Peters, and K.W. Ramsey

The EcoTrends Project was established to aid researchers and others interested in synthetic studies of long-term, continental-scale and national-level patterns in environmental drivers and ecological responses. Hundreds of standardized, documented datasets from many sites and scientific fields were synthesized to meet this goal. Generating comparative data at many sites across several organizational networks and finding novel solutions to technical, organizational, and communication challenges required ongoing collaborative work with all project participants, including researchers and information managers.

The lessons learned from this collaborative effort contributed to our understanding of contemporary ecological information management (that is, the management of digital ecological data via multifaceted, interdependent arrangements and systems). Drawing on these lessons learned by EcoTrends participants—project leaders, researchers, and network- and site-level information managers—we present 10 recommendations for site-level information management and for future synthesis projects. These recommendations for supporting synthesis projects are related to three broad categories:

- Data management and products
- Project design
- Information environments

Challenges

The collection, management, and sharing of ecological data are rapidly changing because of escalating advances in technology and in knowledge-sharing. Advances in automated, continuous collection of data from sensors are increasing the number of methods available to observe and measure the environment. These technologies and methods can generate data that span a wide range of spatial and temporal scales (see Porter et al. 2005, Collins et al. 2006, Benson et al. 2010 as examples). Management of data has evolved along with statistical software and database technologies. For example, quality checking of data for errors in values and formats was previously conducted manually by researchers or technicians but is now often performed using automated statistical software (for example, Michener and Brunt 2000). Data that were once stored in simple spreadsheets are now often stored in more complex relational databases. The sharing of data and knowledge has increased as more research sites post links to their data on web pages or make the data available via new web services. To aid in the sharing of data, data practices, policies, and documentation standards have been and continue to be developed among research communities (for example, Karasti and Baker 2008, Porter 2010, Vanderbilt et al. 2010).

Large synthetic studies of diverse ecological data have been greatly facilitated in recent years by advances in data collection, management, and sharing, which is exciting for the research community, but these new projects also pose new challenges. Comparing large amounts of data across diverse ecosystems can aid in understanding of ecological processes and the effectiveness of new research methodologies. When such analyses lead to new understandings about ecology and ecological data, the lessons learned can inform the next round of data collection, processing, analysis, and documentation. Thus, large synthesis projects have been increasingly popular over the past few decades (for example, Riera et al. 2006, Moran et al. 2008). However, new challenges have appeared with each large-scale project. Here, we describe the primary 10 challenges that the EcoTrends Project faced, grouping them into three categories.

The first category addresses data management and products. Ideally, datasets would be easy to find online and to incorporate into a well-defined workflow for databasing and analysis. However, as the EcoTrends project illustrates, the task of finding and creating comparable datasets from disparate sources can be challenging because of several underappreciated impediments, including—

- difficulties in finding data,
- inadequate data and metadata standards,
- inaccurate or incomplete data and metadata content, and
- complex datasets.

Similar issues have been identified in other environmental science synthesis projects (for example, Benson et al. 2005, Jones et al. 2006, Michener et al. 2007, Baker and Chandler 2008).

The second category addresses synthesis project design. There are many ways to start, design, and implement a synthesis project, and it is important to begin with well-defined goals, knowledgeable and enthusiastic partners, and a well-informed sense of the challenges that may be faced throughout the project. Challenges in this category include—

- data heterogeneity and scaling issues,
- planning flexibility into project design, and
- making decisions on how to best design and implement a project and its requisite information infrastructure.

Finally, the third category addresses information environments to support synthesis. Challenges include—

- working with and developing environments in which information is effectively shared among participants,
- finding motivation to continue the project over time, and
- encouraging involvement of a large number of research sites.

Over the course of the EcoTrends project, participants accumulated a rich body of experience with data processes and collaborative data practices. While large datastreams and technology configurations have prompted a variety of large-scale program endeavors, the EcoTrends project is unique as a multisite, multinetwork activity involving ecological data that span biological, chemical, and physical realms. The project simultaneously informed development while coordinating site- and network-level information environments.

In the next section, we provide recommendations related to the challenges listed above. For each recommendation, we first provide specific examples of the challenges that EcoTrends faced, then the lessons that we learned, and then explain the recommendation that may help address the challenge in future projects. These recommendations are expected to resonate with researchers and information managers, who work together as a cohesive, integrated team at both research sites and in multisite comparative studies of ecological data.

Recommendations for Data, Metadata, and Derived Data Products

1. Make data easily accessible online to researchers.

Locating data for the EcoTrends Project was a time-intensive exercise. A small, but significant, portion of datasets were not stored online, but were submitted via email by individual researchers or information managers. Moreover, online long-term datasets were often difficult to find within extensive catalogs of datasets on the webpage for each research site. Occasionally, when a research site updated its webpage, the link to a dataset changed, and the dataset would have to be relocated by EcoTrends personnel. These challenges were met by contacting researchers and information managers at each research site in order to solicit data that were not online, locate data that were online but difficult to find, and find datasets when they had been moved.

We recommend that research sites be supported in developing practices and procedures to make high-quality, well-documented datasets publicly available online as soon as possible. For example, the Long Term Ecological Research (LTER) program data policy, based on guidelines from the National Science Foundation, states that data should be posted within 2 years of being collected, with a few exceptions. In addition, we recommend that each dataset be assigned locally a unique identifier code, or accession number, that does not change over time. This identifier would make it easier for a synthesis project to more easily find a dataset that has been moved. Dataset titles are often used as identifiers, but these titles are subject to change when datasets are reorganized or displayed at different Internet locations.

2. Implement and develop metadata standards at the site and community levels.

The metadata documentation format was highly variable between research sites. At some research sites, each researcher documented datasets in a format unique to his or her personal standards of completeness. Other sites maintained site-level standards, such as filling out specific fields in a text document. Data downloaded from national repositories usually adhered to the standards created or adopted by that particular repository. For example, metadata from the Climate and Hydrology Databases Project reports metadata for each dataset via a standardized form, the completeness of which varies between participating sites. The LTER sites (approximately half of the participating research sites), however, recently adopted a standard metadata protocol, the Ecological Metadata Language (EML). This specification documents datasets with information such as study location, data collection methods, data policies, and descriptions of data table elements. It also includes community-defined lists of terms, or ontologies, to aid standardization. With EML only recently adopted by the LTER community, many LTER datasets were not yet fully documented and many documentation best practices are still in development.

As a result, the metadata documents that EcoTrends personnel worked with were highly variable between datasets and were error-prone, such that time was spent trying to understand the data. In metadata documents, the locations where data collection took place were often missing. We found that a lack of variable naming conventions (for example, primary productivity may be labeled "primprod" in one table, and "PP" in another table—even within the same study) made data processing difficult. Species names were often recorded as codes in data tables, yet in many cases, the codes contained typographical errors or were not adequately documented in the metadata. In other cases, a lack of detail in the methods led to misinterpretations of how the data were collected. Discussions between the EcoTrends Project Office (EPO) and the lead researcher of the study became a necessary component in processing the data correctly.

EML was developed for a large, diverse community that intended to share data using standards that support consistent data packaging and routine update of datasets over time. The EcoTrends Project found that source datasets with EML documentation were often easier to understand and process than those without such documentation, thus the Project used EML to document every derived dataset that the project generated. These metadata documents contain information about the source dataset (including ownership and a link to the original metadata) and about the EcoTrends Project as well as definitions of the associated data table.

However, while the EcoTrends Project attempted to support the existing EML standards as thoroughly as possible, the resulting documents were incomplete. For example, the methods used to calculate the derived data from the source data are not included in the EML because a standard does not exist for this information. Derived datasets on the EcoTrends website may thus be misinterpreted, and the source data should be examined before proceeding with further analysis.

EcoTrends work brought the concept of derived data to the foreground. The issue of data misinterpretation was discussed with the broader community, prompting discussions about how to best accommodate this level of information within future EML schemas.

EML content standards are still in development, which means that a number of data comparability issues remain undefined. LTER information managers have been prominent advocates for improvement of EML, thereby benefiting the ecological research community. EcoTrends contributed to the development of site-level conventions and to the enactment of metadata standards by reporting documentation errors to site personnel. Specifically, benefits included prompting sites either to create EML for their historical data or to improve on what was available; to standardize attribute, unit, and taxonomic codes and names; to flesh out methods sections; and to provide stable Internet addresses (preferably with dataset accession numbers) for each dataset over time.

We recommend that research sites implement community-wide metadata standards, such as EML, and become involved in the process of refining existing standards and developing new local standards when community standards are not adequate for local research. Implementing local procedures with reference to community standards helps maintain data integrity at both the site and project levels. Standards that guide the documentation of a scientific study, its methodology, and the resulting data tables, can promote responsible sharing and use among researchers by clearly representing dataset origin and can make data more discoverable via online searches.

3. Develop and use standard data practices to create "clean" data.

Data lose their integrity if there are errors. We consider "clean" or quality-controlled data to be free of typographical or value errors and to be easily importable into a spreadsheet, a statistical program, or a database. In practice, there were frequent errors found in the source data that significantly hindered analysis and synthesis. For example, time-series data often had unexplained gaps. Occasionally, incorrect values, such as outliers or incorrectly labeled data (for example, mean temperature labeled as maximum temperature) were found by the EPO during the data processing or during data checking by site personnel. Outliers often existed in the data early in the study when techniques were new and the collection process had not been thoroughly tested. Where data and metadata gave no indication of poor quality or missing value assignment, problem data were inadvertently used in the initial analyses and corrected in the final analyses and graphs.

There are several plausible reasons for a lack of data integrity. Long-term data, assumed to be "clean" due to the long period of time that they have been maintained and their availability on the Internet, may actually suffer from neglect. Legacy data practices such as short and nondescriptive variable names or inadequate software tools for checking are often an issue. Alternatively, when delivery of data from site changes (for example, becomes updated, semiautomated, or automated), quality control, and other site-level analysis work may not be carried out or may not be adequately incorporated into the dataset.

By presenting source data in a recast form on a website, EcoTrends focused the attention of site participants on quality-checking of those datasets. Frequently during the site data checking process in 2008-2009, site personnel noticed erroneous data points in the annual summaries of their datasets, attributable to poor-quality primary data or to erroneous summarization of the data. Many source datasets and EcoTrends-derived datasets were corrected following discussions about data practices that occurred with individual researchers and at larger meetings.

While good data practices goes beyond the scope of this chapter, *we recommend* that sites act upon the developing resources available in the literature at the community level (Michener and Brunt 2000, Cook et al. 2001, Baca 2008, Borer et al. 2009) and the national or international level (NISO 2004, Van den Eynden et al. 2009). Data processing is an iterative exercise involving multiple facets, from sample analysis and measurement calibration to data analysis, quality control, statistical analysis, comparative study, and visualization. All of these activities can occur at both the site level, driven by scientific inquiry for a specific use of the data, and at the multisite or network level, driven by new, often synthetic uses of the data. Site-based analyses to scrutinize the data are needed before data can be used effectively by others. Development of good information-management practices must include ways to prevent misuse and/or misinterpretation of data.

4. Provide well-documented derived data for use by local and remote researchers.

In many cases, the source data were complex and difficult to process correctly due to unique collection and analysis methods. A goal of the EcoTrends Project is to create derived data products whose format is much simpler than the way the data were originally collected in order to ensure that a broad range of users can understand the data. The EPO, in consultation with the science advisory committee, aggregated data using methods commonly used by ecologists. Most of the time, these methods worked well. However, in some cases no matter how well documented and how cleanly represented in data tables, the complexity of the dataset was the main barrier to synthesis. Biotic datasets were particularly challenging, with numerous species and different kinds of measures. In many cases, the Project Office needed to discuss with the lead researcher the suitability of a dataset for a particular aggregation effort.

We recommend that research sites create and post online derived data products as long-term, signature datasets. These types of derived data products are not typically posted online, though they are often created and used for in-house analysis. There are two main reasons for our recommendation.

First, creating derived datasets provides a mechanism for performing regular checks on the integrity of the data, a procedure that helps ensure "clean data" (see recommendation 3). If the data are kept up-to-date in a standard format, then statistical programs can be written to periodically recheck the format of the data tables themselves, check the data table contents against what is recorded in the metadata, check for errors in

the data, and produce visualizations of the data that an experienced researcher could quickly check for anomalies. This recommendation would increase the integrity of the data and increase the stature of the dataset as other researchers use the data over time.

Second, posting in-house, high-quality derived data could have great benefits for collaborative research by assuring the use of appropriate and accurate derivation methods. Moreover, when routinely available, derived data become a shared product that may prompt dialogue among researchers. Several discussions were initiated between the Project Office and sites when datasets were complex and the data aggregation or summarization approach was unclear. For example, while implicitly known as being important at the site level, month-long oceanographic cruises carried out three times a year are rarely integrated to give annual estimates. In general, a check on the regularity and frequency of sampling is required before annual estimates are calculated. Researchers used to working with terrestrial data may inadvertently create annual summaries of the data, not being aware of the issues associated with the logistics of cruises and oceanographic sampling. However, if derived data were made available, along with links to the source data from which they were created and the methods with which they were derived, including algorithms and scripts, they would provide a standard in data quality and use and would increase the integrity of the dataset in its entirety.

Recommendations for Project Design

5. Plan for data heterogeneity and "complexities of scale."

Data are collected, quality-checked, and organized in various ways depending on the phenomena sampled (such as bird counts or wind measurements), the spatial distribution (for example, single vs. multiple locations), frequency of sampling (for example, daily vs. quarterly), regularity of sampling (missing days in a daily record, for example), and methods of data collection (for instance, an observer vs. an instrument). Heterogeneity in data management methods adds to the challenge of producing comparable data. For the EcoTrends Project, we focused on time-series data of specific variables which mitigated some

effects of incoming data heterogeneity. However, no single programming solution could be developed to automate data handling; programming solutions were developed for single datasets or clusters of similar datasets. To share standardized derived data on a website, data summarization and organization were optimized for display of single variables over specific time aggregations (for example annual bird counts or monthly wind speed). Decisions made to simplify website development, such as only graphing variables through time in the EcoTrends Project, resulted in limitations in the current underlying data structure.

Data are also collected and aggregated at different temporal and spatial units under a variety of circumstances. Scaling from small to large regions and from short to long time periods can involve complex processes. For example, sites collect weather data using a varying number of stations distributed across the land. The EcoTrends Project asked each site to identify "representative" weather datasets from their site. For some sites, particularly those that have relatively flat surfaces, choosing data from site headquarters was sufficient because differences between stations were relatively small. At other sites, however, particularly those with major elevation differences within a small area, choosing a "representative" dataset was difficult. If the EcoTrends Project was expanded to use long-term data from all weather stations at each site, this quandary would be side-stepped only to introduce scaling issues due to an increase in the number of datasets to be handled.

The multiple options for presentation of data also introduce complexities of scale. The initial plan—for a website with static content containing data shown graphically in this book—changed to planning for dynamic data delivery and visualization. The Technical Committee recommended structuring the data and database to support automated metadata generation for derived datasets using existing tools that were under develoment (EML for documenting derived datasets and Metacat for cataloging the resulting EML documents) and tracking data provenance and versioning. This proved to be a significant increase in project scope and requirements for information system design and infrastructure building.

We recommend that, before a multisite synthesis project is completely planned and started, the project leaders recognize and consider carefully the project scope, accounting for the variety and complexity of the source data as well as the constraints associated

with their management. Such advance planning is key to adequate and appropriate information management for such synthesis projects. We also recommend that project leaders consider how to best present their data before implementing information management solutions. For example, will the data be presented, as in EcoTrends, as time series? Or will it be expected that different variables will be compared against one another or against non-time-series data? Planning for additional functionality after the project has begun may require changes in how datasets are organized. Therefore, accounting for data heterogeneity and scaling complexity, both in the source data and the resulting data, before the project begins is important. Information specialists trained in both economies of scale and complexities of scale can add insight to project planning (Baker and Chandler 2008).

6. Iteratively design and assess project processes and systems.

Interdependent information environments existed at research sites EPO and LNO. Work at the interfaces of these environments involved an unanticipated amount of coordination and design work as well as mediation, negotiation, and decisionmaking.

The EcoTrends Project started with a linear workflow (traditional for many data management processes), but the workflow rapidly evolved into a cyclical set of processes using feedback from participants to inform further development. Just as the scientific process often does not proceed linearly, there was value in envisioning the data processes as a complex set of interdependent systems, sometimes operating on differing time scales. In the case of the EcoTrends Project, feedback from discussions among various groups subsequently informed further development.

Similarly, data handling cannot be solved by a single technical solution, but rather requires ongoing redesign. *Our recommendation* for improving data handling and information management is to plan for modifications, whether in the short term or the long term, according to insights gained and lessons learned throughout the process. For example, when initial assumptions about the readiness and easy access of long-term data and metadata from site web pages proved to be incorrect, the science advisory committee was formed to inform the process of identifying the variables and datasets of interest and the common aggregations to be performed.

The project coordinator position was developed to work directly with site personnel to obtain, correct, and understand their data in preparation for inclusion as derived data products and to ensure that committee decisions were followed. As the volume and complexity of the data increased, new communication systems evolved, including ways to share derived data with site contributors. The project coordinator position expanded into an interactive role in both assembling data and creating the derived products needed for the EcoTrends Project and in providing feedback to site personnel on the quality of their data and metadata. Iterative modification of a project may include striving to refine conceptual models of how data are stored and related, continuing design of information systems, working iteratively in phases, and incorporating inquiry-based collaborative learning.

7. Involve advisors from fields who reflect the breadth of the project and who are experienced with information management.

Science-driven ecological synthesis projects may be either narrow, focusing on a single variable over space or over time, or broad with respect to space, time, and/or variables. In either case, advice from experts in the fields that the project embraces is highly useful. The breadth of the EcoTrends Project mandated the collaboration of experts in different fields without which EcoTrends would have fallen short of its goals. When EcoTrends was first started, communications regarding project development were principally between two scientists and site principal investigators because it was thought that the data of interest would be easily accessible online. When it was discovered that the data were difficult or impossible to find, the project was formulated more formally. The science advisory committee was formed to widen the breadth of scientific knowledge and the technology committee was formed to inform technological development (chapter 2). Communications were then expanded to first include researchers from each site, then information managers. The LNO formally became involved when supplemental funding from the National Science Foundation became available.

The combined advice from a wide range of expert contributors had a profound effect on the success of the project. *We recommend* for a new synthesis project that the project leader(s) recruit experts whose knowledge spans the breadth of the anticipated project and that

they be involved at the start of project planning. This expansion should include not just experts in the focal science but also experts in roles necessary for the implementation of the project, such as information systems designers, information managers, and statisticians.

Recommendations for Improved Information Environments To Support Synthesis Products

8. Focus on development of both local and network information environments.

An "information environment" is a collection of scientists, information managers, and analysts and of the technology needed to manage and share data. Effective information environments involve development of shared language, conventions, and practices for communication among people from different backgrounds. These environments exist at both site and network levels. They include development and use of technical, organizational, and social work processes to manage multiple types of data and the translation of science. Comparing data from multiple sites can stimulate new information management activities and approaches; however, work on collaborative data activities must be constantly balanced with the need to meet site requirements.

The EcoTrends Project needed an effective information environment to successfully manage data and communications. The environment established included a technological system to track, process, and manage data and a communications system to support collaboration and decisionmaking among participating scientists, information managers, and developers. These systems had to develop iteratively with lessons learned from one iteration informing the development of the next. Specifically, these systems promoted understanding of technical and cultural issues regarding data; informed decisions on how data should be selected, processed, and shared; and provided feedback on data handling. Time invested in identifying, developing, and using coordination mechanisms accounted for a large amount of unplanned time that was ultimately recognized as well spent.

We recommend that sites that already have information environments continue to invest in their multifaceted growth and ongoing redesign and that sites without a formal environment dedicate time to developing strategies for creating one, even if resources are scarce. The rewards of a smoothly operating set of practices and systems more than compensate for the cost.

9. Combine long-term data handling with short-term scientific products and data checking procedures.

Throughout the several years that the EcoTrends Project needed to produce its intended products—this book and a complementary website—it was important to keep participants engaged with the project and to share preliminary products. EcoTrends generated both short-term scientific products and periodic data checks requested by the participating sites. The scientific products included papers written by the 2009 scientific working groups. These prompted review of the website content and accessibility, fostered new ideas for future website features and content, and motivated supporters of the project. EcoTrends also developed a data quality report when requesting sites to check their derived data. Created as a spreadsheet and distributed easily by email, this file provided a much needed feedback mechanism for sites and provided a useful, albeit improvised, approach to recordkeeping. Each round of responses from the sites after a data-checking session generated improvements to the report. In the long term, however, a more sophisticated online solution may be more robust, transparent, and user-friendly.

Balancing long-term goals with short-term actions is central to development of a contemporary information environment. Juxtaposing the fulfillment of immediate tasks within a well-defined long-term project creates an environment in which design can be proactive planning for the future while meeting immediate needs. Short-term scientific products, such as papers that examine the data, can justify the usefulness of the project, motivate participants to continue with further development, and inform future development. Data-checking events can validate data processing, elicit feedback from the supporting community, and generate enthusiasm for the project. However, short-term products may require the development of new methods or work-arounds to create them, potentially involving new analysis procedures, communication mechanisms, or types of collaborative

activities. These methods or work-arounds can be very useful, but they should inform long-term project development.

10. Develop and maintain transparency by fostering communication and feedback.

Project transparency refers to making participation, processes, and systems accessible and clear for both those closely involved and those casually connected to the project. Transparency requires constant attention to ensure availability of information and openness of the decisionmaking process. While the original intent of the EcoTrends Project was to be open and inclusive, identifying and developing mechanisms for collaboration and documentation took time. Initially, the existing LTER community networking infrastructure—from listservs to use of regular LTER community meetings and monthly information management video conferences—served the project well. However, there was a persistent push to create and continue collaborative activities that would open up discussions concerning data by EcoTrends committees or individual research sites to a public arena that could engage a full spectrum of data providers and users.

The EcoTrends Project Office communication systems evolved in response to projects' and participants' needs. For example, an initial group email request for data submission was followed by individual site communications; committee work with individual hardcopies of graphs evolved to presentation of graphs on an internal website. Presentations at community events improved multisite awareness and engagement. Initial contact with principal investigators and selected members of committees eventually broadened to include information managers and eventually the LTER information management community. The development of a site-specific spreadsheet summarizing dataset submissions created much needed feedback to sites and a coordination mechanism for joint recordkeeping, both within a site and between sites and the Project Office. Graphical representations were referenced online to allow sites to check their contributions.

Attention to project transparency improved both quality and quantity of data submitted, influenced the practice of collaborative science, and promoted buy-in to the EcoTrends Project by participants at all sites.

We recommend that future projects assess the needs of their stakeholders as involved and engaged participants and plan accordingly for project transparency. Research into existing communications systems and online networking tools may help. In addition, we recommend that the project be poised to evolve their communication systems as further needs are perceived.

Conclusions

In this chapter, we presented key lessons learned and recommendations for future synthesis projects from the perspective of a distributed information management team tasked to support network-level ecological research. Alternatively, a site-based research scientist using the data from such a project might have further recommendations on how to best expand analysis teams and develop software routines to statistically explore the data. A software or database developer might have further insights in framing unique, iterative design situations for use in dynamic synthesis environments. Successful planning of any large data synthesis project can be significantly enhanced by the perspectives and knowledge of people from diverse backgrounds and experience.

The EcoTrends Project can be considered a success for the following reasons:

- First, this book, with a diverse array of summarized long-term data collected from 50 sites, and an associated website with some searching and data exploration functionalities fulfill the initial goals of the project.
- Second, EcoTrends contributed significantly to both individual- and community-level understanding of multilevel information management by providing hands-on experience with multisite data integration.
- Third, the EcoTrends Project was unique in carrying out a data production process in a collaborative, interdisciplinary setting with a well-established information management community and in having the information system work distributed between two geographically distinct, but communicating centers (EcoTrends Project Office in Las Cruces, NM, and LTER Network Office in Albuquerque, NM). These arrangements reveal a number of underappreciated dimensions of the work involved in creating comparable data.

In addition to the highlighted successes, the EcoTrends Project demonstrates the importance of addressing and supporting knowledge production, data production, and infrastructure growth within a single framework. The project also highlights the importance of broadening participants' perspectives over time via transparent processes and communication. Specifically, the perspectives of EcoTrends Project participants broadened from simply defining digital products and a single companion workflow to eventually envisioning multiple interdependent data processes and information environments. These processes and environments included not only a technical infrastructure but an array of organizational and social arrangements. Besides just considering the data and the individual work arenas, participants learned to consider the variety of participant roles and activities that tied them together. Iterative, collaborative learning throughout a project and planned flexibility to react to new ideas were important elements of the EcoTrends Project and may well serve any new multisite synthesis project.

References

Baca, M. 2008. Introduction to Metadata. 2nd ed. Getty Research Institute, Los Angeles, CA.

Baker, K.S., and C. Chandler. 2008. Enabling long-term oceanographic research: changing data practices, information management strategies and informatics. Presentation at Fall American Geophysical Union Conference, Dec 15-19, 2008, San Francisco, CA.

Benson, B.J., P. Hanson, J. Chipman, et al. 2005. Breaking the data barrier: research facilitation through information management. *In* JJ Magnuson, TK Kratz, and BJ Benson, eds., Long-Term Dynamics of Lakes in the Landscape, pp. 259-279. Oxford University Press, Oxford, United Kingdom.

Benson, B.J., B.J. Bond, M.P. Hamilton, et al. 2010. Perspectives on next-generation technology for environmental sensor networks. Frontiers in Ecology and the Environment 8:193-200.

Borer, E.T., E.W. Seabloom, M.B. Jones, et al. 2009. Some simple guidelines for effective data management. Bulletin of the Ecological Society of America 90:205-214.

Collins, S.L., L.M. Bettencourt, A. Hagberg, et al. 2006. New opportunities in ecological sensing using wireless sensor networks. Frontiers in Ecology and the Environment 4:402-407.

Cook, R.B., R.J. Olson, R. Kanciruk, et al. 2001. Best Practices for Preparing Ecological Data Sets to Share and Archive. Bulletin of the Ecological Society of America 82:138-141.

Jones, M., M. Schildhauer, O.J. Reichman, et al. 2006. The new bioinformatics: integrating ecological data from the gene to the biosphere. Annual Review of Ecology, Evolution, and Systematics 37:519–544.

Karasti, H., and K.S. Baker. 2008. Digital data practices and the Global Long Term Ecological Research Program. International Journal of Digital Curation 3:42-58.

Michener, W.K., and J.W. Brunt. 2000. Ecological Data: Design, Management and Processing. Blackwell Science, London.

Michener, W.K., J.H. Beach, M.B. Jones, et al. 2007. A knowledge environment for the biodiversity and ecological sciences. Journal of Intelligent Information Systems 29:111-126.

Moran, M.S., D.P.C. Peters, M.P. McClaran, et al. 2008. Long-term data collection at USDA experimental sites for studies of ecohydrology. Ecohydrology 1:377-393.

NISO [National Information Standards Organization]. 2004. Understanding Metadata. NISO Press. http://www.niso.org/publications/press/UnderstandingMetadata.pdf. Bethesda, MD.

Porter, J., P. Arzberger, H. Braun, et al. 2005. Wireless sensor networks for ecology. BioScience 55:561-572.

Porter, J.H. 2010. A brief history of data sharing in the U.S. Long Term Ecological Research Network. Bulletin of the Ecological Society of America 91:14-20.

Riera, J.L., T.K. Kratz, and J.J. Magnuson. 2006. Generalization from intersite research. *In* J.J. Magnuson, T.K. Kratz, and B.J. Benson, eds., Long-Term Dynamics of Lakes in the Landscape: Long-Term Ecological Research on North Temperate Lakes, pp. 107-120. Oxford University Press, New York, NY.

Van den Eynden, V., L. Corti, M. Woollard, et al. 2009. Managing and Sharing Data: Best Practice Guide for Researchers. www.data-archive.ac.uk/media/2894/managingsharing.pdf. United Kingdom Data Archive, University of Essex, Wivenhoe Park, Essex, United Kingdom.

Vanderbilt, K.L., D. Blankman, X. Guo, et al. A multilingual metadata catalog for the ILTER: issues and approaches. Ecological Informatics 5(3):187-193.
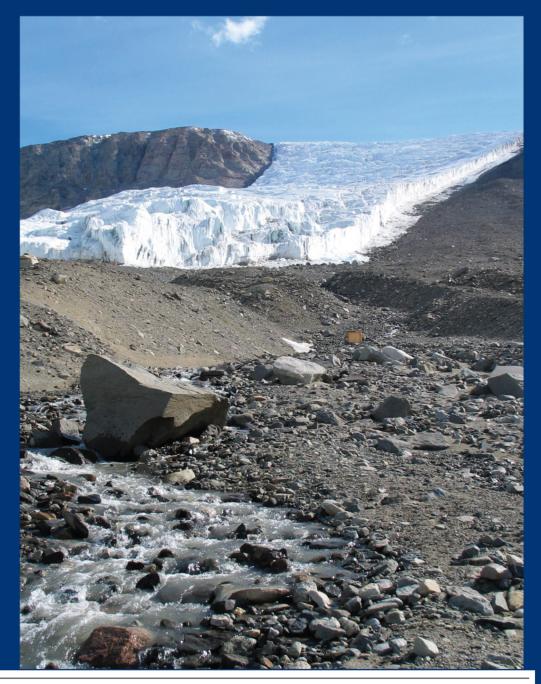
Long-Term Trends in Ecological Systems:
A Basis for Understanding Responses to Global Change

Editors:

Debra P.C. Peters, Christine M. Laney, Ariel E. Lugo, Scott L. Collins, Charles T. Driscoll, Peter M. Groffman, J. Morgan Grove, Alan K. Knapp, Timothy K. Kratz, Mark D. Ohman, Robert B. Waide, and Jin Yao