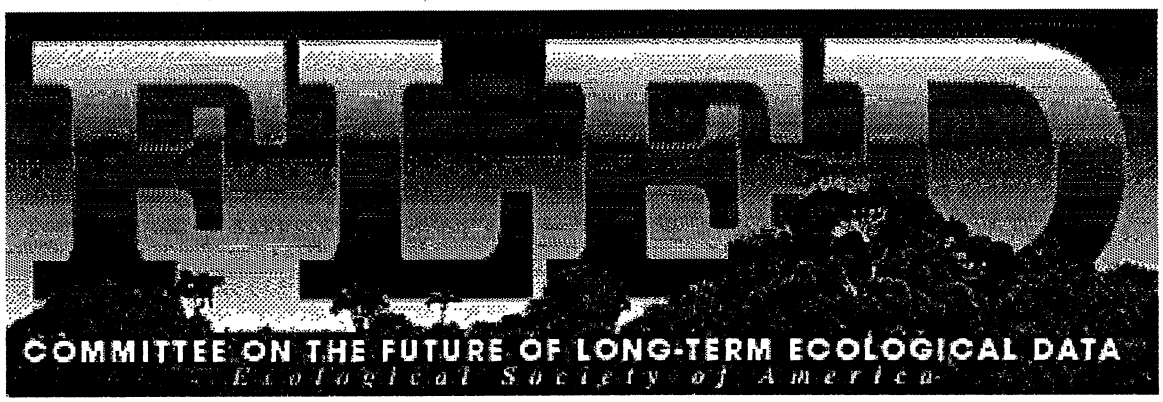
Final Report of the

Ecological Society of America

Committee on the

Future of Long-term Ecological Data (FLED)



submitted by

Katherine L. Gross, Chair

and

Catherine E. Pake, Research Associate

Volume I: Text of the Report

December 1995

Acknowledgements

The work of the FLED committee was supported by a grant from the Andrew W. Mellon Foundation. The Ecological Society of America (ESA) and the FLED committee gratefully acknowledge this support. The opinions and recommendations presented here represent those of the FLED committee and have been approved by the ESA Governing Board (October, 1995). Although the Mellon Foundation supported the work of the committee, they have not influenced, reviewed or endorsed any of its recommendations.

Final Report of the Ecological Society of America

Committee on the

Future of Long-term Ecological Data (FLED)

Volume I: Text of the Report

submitted by

Katherine L. Gross, Chair,

and

Catherine E. Pake, Research Associate

and the FLED Committee Members

Edith Allen
Caroline Bledsoe
Robert Colwell
Paul Dayton
Megan Dethier
John Helly
Robert Holt
Nancy Morin
William Michener
Steward T.A. Pickett
Susan Stafford

Project Assistants

Ann O'Neill
Justin Pittinaro

Address reprint requests to:

The Ecological Society of America
2010 Massachusetts Avenue, NW
Suite 430
Washington, D.C. 20036 USA

The Ecological Society of America

GORDON H. ORIANS, President
Department of Zoology, Box 351800
University of Washington
Seattle, WA 98195

Phone: (206) 543-1658
FAX: (206) 543-3041
e-mail: orians@zoology.washington.edu

October 23, 1995

Dr. Katherine L. Gross
W. K. Kellogg Biological Station
Michigan State University
Hickory Corners, MI 49060

Dear Kay,

At its October 20-22 meeting in Washington, D. C. the Governing Board of the Ecological Society of America accepted the Final Report of the ESA ad hoc Committee on the Future of Long-term Ecological Data (FLED) as summarized in the DRAFT "Executive Summary and Recommendations" dated October 16, 1995. We look forward to receiving a final copy of the full report in the near future.

You will be pleased to know that the Governing Board is already acting on many of the FLED Committee's recommendations. Steps are being taken to post funding opportunities related to the recovery, restoration, and use on long-term datasets on the ESA Homepage. Members will also be encouraged to write letters of support for these programs. We discussed the possibility of developing a special feature in Ecological Applications on issues surrounding long-term data sets.

The newly constituted special committee on Data Sharing and Archiving, on which you serve, will further analyze many of the recommendations of the FLED report, including collaboration with other organizations, examination of policies for citation of archived data and their sources, intellectual property rights and ethical responsibilities associated with archived datasets, and metadata that should accompany all archived datasets. Whether or not the ESA will establish a position of Internet Services Manager will depend on the results of the deliberations of the special committee and other inputs to the Governing Board. Meanwhile, some of the functions that would be performed by an ISM will be carried out by existing staff. Please communicate the appreciation of the Governing Board to all members of the FLED Committee for their excellent work on a very important topic.

Sincerely yours,

Gordon H Orians

Gordon H. Orians, President

cc: Nancy Huntly

TABLE OF CONTENTS

Report of the ESA *ad hoc* Committee on the Future of Long-term Ecological Data (FLED)

Executive Summary and Recommendations

Long-term studies are invaluable for addressing many questions in ecology. Because it is difficult to maintain such studies, and there currently exists no national repository for the data sets, many valuable long-term data sets are at risk of being lost. In recognition of the importance of long-term data sets and the need to develop mechanisms to promote their preservation, maintenance, and use, the Ecological Society of America (ESA) established an *ad hoc* committee on the Future of Long-term Ecological Data (FLED) in April 1993.

This report provides a summary of the work of the FLED committee and its recommendations on how to meet the challenges of preserving long-term ecological data through the use of information technology, the shaping of professional policy and the development of funding sources. The report recommends mechanisms for locating, restoring, preserving and accessing long-term ecological data.

The primary goal of the FLED committee was to develop recommendations for a process by which valuable long-term, or historical, ecological data sets could be identified, preserved and made accessible to the scientific community. Although our focus was on "long-term" studies (defined by the FLED committee to be 5 years or longer), we recognized that many of the issues raised and recommendations made apply to **all ecological data**. However, because many long-term studies were initiated before the advent of computerized data bases, many of these data sets exist only in paper form and are therefore particularly susceptible to loss and are less available than data stored electronically. In addition, critical information as to how the data were collected and recorded (the "metadata") is often scattered and must be retrieved from such diverse sources as field notes, logs and publications. Thus, restoration of many of these long-term data sets into a format that is accessible to the community is critical if these data sets are to be preserved and used.

The Committee recognizes that several of the recommendations it is making to the Governing Board of the ESA, will require a different type of commitment to facilitating scientific exchange than the Society has traditionally taken. However, the ESA is on the advent of a number of new ventures involving increased electronic communication that will likely revolutionize the way we interact professionally, communicate with each other and even publish our work. For example, the ESA, in collaboration with the Mellon Foundation, will soon republish electronically all past issues of its three research journals, and will move toward simultaneous electronic and paper publication of new issues of these journals. The ESA has also joined in partnership with the founder of a new, all-electronic journal called *Conservation Ecology*.

It is a logical next step for the ESA to take the lead in making information available to its members on sources of long-term (and other) ecological data, to facilitate the recovery and restoration of existing long-term data sets that are at risk, and to provide leadership and guidance in addressing the concerns of individual scientists that will arise with increasing access to and use of electronic sources of data. The ESA is recognized for its leadership and concerns about many of these issues. Thus, the recommendations of the FLED committee to the Governing Board of the ESA were made to provide both a means to ensure that valuable sources of ecological information are not lost, and to promote within the ESA continuing efforts to cooperate with scientists in other disciplines with similar interests and concerns.

Sources of Long-term Ecological Data

Long-term ecological data are collected by agencies, institutions, and individuals. The incentives for data collection vary from regulatory mandates to protect endangered species or human health to scientific curiosity. These data span a variety of scales of investigation from populations to ecosystems, from communities to landscapes. Long-term data exist on rare species, pest species, agricultural production, toxin levels, pesticide levels, climate, hydrology, water and air quality.

National agencies and organizations that collect long-term ecological data may be governmental or non-governmental. The FLED committee identified more than 25 national agencies and organizations that collect or maintain long-term ecological data. Government agencies, other than research laboratories, typically engage in monitoring rather than research. Monitoring data, however, may be valuable in testing hypotheses about long-term ecological processes.

Long-term ecological data are also collected and maintained by private research laboratories, environmental organizations, museums, and field stations. Many of these have curated, long-term data that are available, sometimes on the Internet, sometimes for a fee. Museums contain voucher specimens, as well as databases derived from specimens, and have had a long history of free and open exchange with researchers.

Data from individual researchers may be the most abundant type of long-term ecological data, and they are clearly the most diffuse. Individual researchers are typically associated with universities or research institutions, and their research activities are frequently funded through competitive grants. Access to these types of data generally depends upon an agreement between the user and the investigator. Sometimes granting agencies will specify conditions for data availability that will explicitly promote data sharing.

The long-term ecological data collected by these different sources fall along a continuum of risk of being lost. Data that appear the most secure are archived in a data bank with a history of solid funding, generally from a government source. Among the least secure are data in the hands of an individual researcher who has made little or no provision for long-term curation. However, data collected by a federally-funded agency are not necessarily secure either. Agency budgets diminish and mandates change, often leaving a data set with no curator. Examples of this type of data being thrown out during an agency housekeeping sweep are not uncommon.

Individual researchers at universities or research institutions, the bulk of the ESA membership, are the group most likely to have long-term data that are at the greatest risk of being lost. Although laws and regulations passed during the Reagan administration explicitly transfer ownership of intellectual property to the grantee (generally the University or institution, not the PI!), currently universities take little responsibility for long-term data curation. If a researcher retires, dies, has a break in funding, or loses interest in or the ability to continue a project there is great risk that the data will be lost. The FLED committee has learned of numerous incidences of researchers ending their careers without the resources to make provisions for the curation and maintenance of their long-term data sets.

Recommendations

To support the need of ecologists to locate, and to promote the use of, existing sources of long-term ecological data, we recommend that the ESA:

1. Establish and maintain a HomePage of Long-term Ecological Data that would provide pointers to sources of long-term ecological data. The information in this HomePage would initially include the summaries and descriptors provided in Volume II of the FLED Report as Directory 1 (United States Agencies and Organizations), Directory 2 (Individual Long-term Ecological Data Sets), and Directory 3 (Annotated Bibliography of Existing Catalogs, Directories and HomePages with Long-term Ecological Data.)

2. Cooperate with and support efforts of agencies and other scientific organizations at the national level for the establishment of a network of ecological and environmental data sources and exchange (e.g. the NBII--National Biological Information Infrastructure of the National Biological Service; and the NBIC--National Biological Information Center).

3. Educate and provide information to the membership on how to use and contribute to these national data exchange networks.

Options for Recovery and Restoration of Long-term Ecological Data

The difficulty of obtaining funding for the maintenance and conversion of long-term data sets to a more permanent (and easily accessible) format was one of the most frequently cited concerns of ecologists who maintain long-term data. It is also one of the major reasons why such data are at risk of being lost. In many of the cases we reviewed, individuals (either principal investigators or designated caretakers of the data) had relied on a variety of funding sources to restore and/or preserve data sets. These frequently involved the development of partnerships among several agencies or institutions with a common interest in the project. The success of these joint partnerships relied heavily on the commitment of one or a few individuals to seek and develop this funding. Unfortunately, generally the institutions or agencies which have supported many of these long-term projects do not have a mechanism for assuring their continuation or preservation, with appropriate documentation, to be maintained as a valuable long-term resource for other scientists.

The National Science Foundation (NSF) is important source of support for much of the ecological research conducted by individual ecologists in the United States. Many projects that are initiated as short-term experiments or studies become, by the nature of the complexity of the system and the ability of investigators to obtain continued funding, a long-term study. Recognition of the need to develop more sustained funding for long-term research led to the development the LTER (Long Term Ecological Research) program at the NSF to provide direct support for long-term ecological research. The LTER program currently supports research at 18 sites (16 in North America and 2 in Antarctica) and a Network Office which facilitates cross-site synthesis and integration. Hundreds of individual scientists are involved in the U.S. LTER program, and its success, nationally and internationally, is well-documented.

The Long-Term Projects in Environmental Biology (LTPEB) cluster in the Division of Environmental Biology at NSF includes two other programs that support long-term research in both systematics and ecology. The LTREB (Long Term Research in Environmental Biology) Program provides direct support to individual scientists to conduct long-term research (but at a considerably lower level of funding) outside of the LTER program. This program has supported 20-30 long-term research projects maintained by individuals. Appropriate data management is an important consideration in the review and funding of projects in this program. In many cases, LTREB funding has been used to restore (or convert to electronic format) data sets that were available only in paper format (e.g. Murray Buell's studies of old-field succession at the Hutcheson Memorial Forest).

The Research Collections in Systematics and Ecology (RCSE) at NSF is another possible source of funding for the restoration and preservation of long-term ecological data. Although it currently primarily supports projects to convert research collections (information from specimens) to electronic form, many of these collections have ecological and systematic value. Funding for this program, though limited, is considerably higher than that available for the LTREB program ($6.5 vs. $1.0 million in 1995). This opportunity for funding of long-term ecological research collections (data and items) should be promoted in the ecological community, particularly when there is an opportunity to work in collaboration with systematists to preserve information that is valuable to both disciplines.

Establishing criteria for determining which data sets should receive the time, energy, and funding needed for restoration and curation is an over-riding concern. The FLED committee acknowledges that every long-term data set probably has unique features which would make it valuable and therefore worthy of preservation. It is impossible to know the questions and hence the requisite data that ecologists will be evaluating in the future. For example, global warming and ozone depletion were not major issues of ecological concern 100 years ago. Nonetheless, data collections initiated then can provide important and otherwise unobtainable information on variables that we can now use to address these questions.

To maintain flexibility and maximize inclusion of the variety of long-term data sets that exist, we have developed some general guidelines to determine which data sets should have priority for the limited funds and energy that are available for restoration and/or inclusion in a ecological data archive. The actual selection of data sets (or projects) to receive funding for restoration or to be included in a national archive is best determined by a peer-review process.

Important criteria to include in this evaluation are:

1. Metadata. The quality of the metadata is the single most important component of this decision. The metadata (higher level information about the data set, why and how it was collected, its content, quality, and structure) determine the utility of the data set to other scientists.

2. Rarity. To some degree, every data set is unique. However, some data sets might be considered more rare than others. .

3. Length of record. Exceptionally long-term records are not only rare, they give us a dynamic picture of the variables measured.

4. Relocatability. Relocatability gives the data value for resampling.

While the value of a site that is not relocatable is reduced, it may still have value at a large scale.

5. Statistically analyzable. Those data that were collected in a way that is amenable to our present statistical techniques and standards are the most valuable and should have the highest priority for preservation. However, the lack of 'statistically valid design' based on current statistical standards should not be used as the sole criterion to exclude a data set from preservation or analysis.

6. Size of spatial domain. Data collected from large spatial scales (e.g., regional and global) represent a wealth of information similar to long-term data, albeit from a different perspective. These data may be treated as having a value similar to those of long temporal scales.

Recommendations

A critical issue that distinguishes long-term ecological data from other types of ecological data is that often the data need to be 'restored' or converted to an electronic form before they can be analyzed and made available to others. Sources of funds to restore long-term ecological data are limited, and new sources need to be promoted and supported. The FLED committee therefore recommends that the ESA:

1. Work with its membership to promote funding for the restoration of valuable long-term ecological data sets within existing programs at NSF. The ESA should particularly promote opportunities for collaborative efforts with systematists to utilize funds from the RCSE program.

2. Work to identify alternative sources of funding for restoration of long-term ecological data sets and keep the membership informed of these possibilities through the ESA HomePage.

3. Establish a "gatekeeper committee" in conjunction with the Long-Term Studies Section (LTSS) to develop criteria for prioritizing data sets to be restored and to be preserved in an ecological data archive. This committee could also serve in an advisory role for ecologists seeking appropriate archives for long-term data sets.

Demystifying Metadata

The important role of metadata (i.e., data documentation) in facilitating ecological research has been recognized since the 1980's and several practical approaches to metadata management have been presented, much of it attributable to the leadership of data managers associated with the LTER network in the United States. Significant progress has been made during the past decade in developing metadata standards for geospatial data. For example, in 1994, a comprehensive set of Content Standards for Digital Geospatial Metadata was released that defines standard geospatial metadata descriptors related to data availability and accessibility, determination of fitness for use, and processing and utilizing a set of data. However, metadata standards for non-geospatial ecological data currently do not exist in any standard format beyond individual studies and experiments.

The lack of adequate metadata is a primary reason that long-term ecological data are at risk of being lost. The very nature of many long-term studies makes them more susceptible to having inadequate or irretrievable metadata. Many processes can lead to the loss of information about a particular data set or project through time. These include technical concerns such as the gradual (and inevitable) degradation of storage media containing the data, obsolescence of storage technology, or the loss of storage media through catastrophic events. Also, the memory and notes of the investigator are often the primary source of much of the information required to document a data set. This source can be lost permanently (or be extremely difficult to retrieve) after project results have been published or the study has been terminated. Unfortunately, many of the specific details required to interpret a data set are lost when it is converted to an electronic format--data forms and field notes are often discarded when the data are digitally preserved.

Because the quality of the metadata is critical to the restoration of long-term ecological data and is a decisive criterion for data sets that are currently accepted into permanent archives run by other biological disciplines, members of the FLED committee devoted considerable effort to identifying the minimal documentation required for an ecological data set. They examined potential benefits and costs associated with developing and implementing metadata for non-geospatial ecological data. In the report, we propose a set of generic metadata descriptors which could serve as the basis for a "metadata standard" for the ecological sciences and present alternative strategies for metadata implementation that meet differing organizational or investigator-specific objectives.

It is important to stress that publications based on a long-term (or any) ecological study generally **do not** include sufficient information about how the data were collected to allow another scientist to use the data. The recommendations in the FLED committee report include recommendations as to the metadata format and structure that would be needed for different categories of users. Exchange with an expert colleague or collaborator, for example, requires considerably less detail than with a third party or an audit by an agency.

Stewardship and a continuing need for curation and maintenance of the data and metadata represent real cost burdens which are not often factored into project budgets. After a study is deemed completed, who bears the responsibility of informing the user community of changes to the data set and newly discovered anomalies? The role and appropriateness of funding for data originators as metadata consultants have received little attention. Also, the ESA needs to promote the development of appropriate incentives to encourage (and reward) individuals who carefully document data and make it available to others.

Recommendations

In recognition of the fact that thorough documentation of the context, content, quality, and structure of data sets is necessary for long-term ecological data to be preserved in a usable format, the FLED committee recommends that the ESA:

1. Educate its membership on the importance of adequately documented metadata and its necessity for maintaining the long-term utility of any ecological data set.

2. Develop minimal metadata standards as part of feasibility plans for an electronically accessible ecological data archive. Other parts of this plan include guidelines for submission and mechanisms for reviewing and accepting candidate long-term (and other) ecological data sets to an ESA-sponsored archive.

3. Examine and promote incentives for the expenditure of time and energy required to adequately document a data set. Possible incentives involve the development of an ESA registered (or endorsed) system of accession numbers for ecological data sets and establishing guidelines for citing such data sets in ESA publications (also grant proposals and vitae).

Data Sharing and Long-term Curation

The preservation and use of long-term ecological data brings up issues surrounding the sharing of data. Ecologists have relatively little experience with data-sharing, and that which does occur is typically between close associates. The notion of data preservation and long-term curation suggests to many a centralized or decentralized data archive. After data are electronically entered into such an archive, they can be transferred to users almost instantaneously. The ease with which colleagues can have access to each others' data is at once marvelous and frightening. It is a breakthrough for free and open exchange of information in the pursuit of knowledge. On the other hand, this innovation brings up the specter of greater potential for ethical misconduct.

The FLED committee considered a number of questions and concerns that are relevant to individual scientists regarding data sharing and sought the advice and experience of scientists in other biological disciplines that have developed successful (and failed) data archives and exchange networks. From this we have developed a model of what makes data-sharing successful and what are some of the long-term concerns that remain even within the successful data archive and exchange systems that have developed in other disciplines. Among the key issues that lead to successful data sharing and exchange is the support of the professional societies and the development of a reward and recognition system for contributions to (and use of) data in the network.

The increasing interest in synthesis and integration in ecology will often demand data from sources outside ecology. This will increase opportunities for data exchange groups with whom ecologists have traditionally collaborated, such as systematists, geosciences, and land managers as well as those with whom collaboration has been less frequent. Many of these disciplines have demonstrated leadership in the development of electronic data bases to support data-sharing and exchange. Ecologists, as individuals, and the ESA, as a Society, should collaborate with scientists in these other disciplines to work toward the establishment of fair, effective and efficient means for facilitating data exchange among and within disciplines.

Recommendations

To promote the long-term curation of ecological data sets and their use, we recommend that the ESA:

1. Support the establishment of a pilot or demonstration project for curating and archiving ecological data as a means of exhibiting the utility and structure of a permanent ecological data archive and exchange network. The Database Activities (DBA) Program in the Division of Instrumentation and Resources at the NSF is specifically designed to provide funding for the establishment of such activities.

2. Encourage the Ethics Committee to review the ESA Code of Ethics for explicit concerns regarding data sharing and the protection of intellectual property rights in the electronic age.

3. Promote interactions and collaboration between ecologists and scientists of other disciplines for the development of tools for successful data exchange among and within these disciplines by:

- co-sponsoring workshops and symposia that highlight the intellectual links between other disciplines and ecology and develop mechanisms for more effective data exchange among these disciplines;

- encouraging the ESA Editorial Board to work with authors to assure that papers published in ESA journals cite relevant data bases from data exchange networks.

- supporting efforts by other disciplines (particularly systematics) to increase funding for the maintenance of collections and the development of associated databases to assure that these valuable sources of 'incipient ecological information' are maintained and available to ecologists.

Conclusion

Despite the interest in and acknowledgement of the value, need, and potential loss of long-term ecological data sets, there is currently no national system that will serve as both an archive and source of long-term (or other) ecological data for the ecological community. It thus seems appropriate that the ESA continue its tradition of promoting opportunities and tools that will enhance the quality and impact of ecology as a scientific discipline by acting on the recommendations made in this report that will lead to the preservation and increased availability of long-term ecological data. The establishment of a national ecological data exchange network--focusing on the preservation of all types of ecological data, but with inclusion of data from long-term studies as a priority--would be valuable to the entire ecological community.

Preface

Long-term studies are invaluable for addressing many questions in ecology. Because it is difficult to maintain such studies, and there currently exists no national repository for data sets from these studies, many valuable long-term data sets are at risk of being lost. In recognition of the importance of long-term data sets and the need to develop mechanisms to promote their preservation, maintenance, and use, the Ecological Society of America (ESA) established an *ad hoc* committee on the Future of Long-term Ecological Data (FLED) committee in April 1993. This report provides a summary of the work of the FLED committee and its recommendations.

The primary goal of the committee was to develop recommendations for a process by which valuable long-term, or historical, ecological data sets could be identified, preserved, and made accessible to the scientific community. Although our focus was on "long-term" studies we recognized that many of the issues raised and recommendations made apply to **all ecological data**. However, because many long-term studies were initiated before the advent of computerized data bases, many of these data sets exist only in paper form. In addition, critical information as to how the data were collected and recorded (the "metadata") is often scattered and must be retrieved from such diverse sources as field notes, logs and publications. Thus, restoration of many of these long-term data sets into a format that is accessible to the community is critical if these data sets are to be preserved and used.

The final report of the FLED committee has been written as two volumes. Volume I contains the main text of the report and provides the rationale and background information in support of the recommendations we have made to the Governing Board of the ESA. It consists of four sections: Sources of Long-term Ecological Data, Options for Recovery and Restoration of Long-term Ecological Data, Demystifying Metadata and Data Sharing and Curation. This volume also includes as appendices a summary of various sources of long-term ecological data and lessons learned from other disciplines that have established data repositories and/or exchange systems.

Volume II of the report includes three directories to various sources of long-term ecological data. This includes information on national agencies and organizations that collect long-term ecological data, examples of long-term data sets maintained by individuals (some of which are 'at risk'), and an annotated list of existing directories and catalogs that include long-term ecological data. These directories were compiled to provide an initial road map to the variety of sources of long-term (and other) ecological data currently collected and maintained by a variety of sources. An increasing number of these directories (or the data sets themselves) are available on-line via the Internet. Where possible, we have provided the information on how to access these directories or catalogs electronically.

Scope of the Report

In initiating this project, the FLED Committee had to deal initially with several important issues regarding the scope of the project. The first of these was a consideration of two questions: what are data? and what is a long-term study in ecology? This issue might at first appear trivial, but in fact it was necessary for us to be able to determine how we would focus our efforts and attention. The second issue was the geographic scope of the project--how much effort should be put into international as opposed to national efforts to locate and preserve ecological data?

The discussion of the question "what are data?" revolved around the distinction between information that exists in the form of numbers and values--what we termed "existing data"--and such things as maps, field notes, photographs or slides and even samples that are often taken in conjunction with a ecological study - what we termed "incipient data." There is demonstrated value to having the original maps, documents or physical samples from a long-term study available in order to extract more information, e.g., soils from the Rothamsted plots (Leigh and Johnston 1994) and blood samples used to detect the Hanta virus in New Mexico (Parmenter et al. 1993). However, mechanisms for preserving and archiving the various sources of "incipient ecological data" are varied and complicated. Therefore the FLED committee decided early on that the main focus of our work would be on developing recommendations for the location, restoration and exchange of "existing data." Such data can be readily archived and retrieved with existing electronic exchange systems.

The question of what constitutes a "long-term study" in ecology obviously depends very much on the questions being asked and the organisms (or systems) being examined. Recognizing that any time-limit we proposed would be arbitrary, the FLED committee adopted the "five year" time scale as a definition of a long-term study in ecology. This is the time scale used by the National Science Foundation (and several other federal agencies) in determining what is a long-term project in ecology.

The second issue related to defining the geographic scope of the Committee's efforts, especially in terms of locating sources of long-term studies. We became aware of a number of international efforts to promote and develop mechanisms to enhance data exchange in ecology and related disciplines (e.g. the SCOPE

Committee on Data, CODATA). While we have tried to keep the scope of this report broad, we have had to limit its focus to existing ecological data primarily from individuals, organizations and agencies within the United States. This nationalistic perspective was not meant to be exclusionary--and in fact we followed up on all the contacts that we received, without regard to geographic boundaries. However, most of our contacts and queries regarding the existence of long-term ecological data sets were focused in North America, primarily the United States.

In developing this report the FLED committee held several meetings and workshops with a number of people with shared interests and concerns about preservation of long-term ecological data. Scientists outside the FLED committee who participated in these meetings and workshops are listed in Appendix B. Also, the committee contacted a number of people who provided valuable information about previous efforts to develop systems for an ecological information exchange and helped identify a number of valuable long-term ecological data sets. These individuals are listed as Special Contributors in Appendix B.

Background and Rationale for the Report

Value of Long-term Studies in Ecology

The value of long-term ecological data for ecological studies is well documented (Likens 1989; Risser 1991). Strayer et al. (1986) and Westoby (1991) have articulated the variety of ecological phenomena and processes for which long-term data are needed and why they are so important for ecological studies. Strayer et al. (1986) stressed the importance of long-term data to assess slow phenomena, rare events, subtle processes and complex processes in ecological systems. Westoby (1991) emphasized the value of long-term data for documenting the statistical relationships between processes that occur on different temporal scales and for the analysis of between year variation that is needed to determine return times for rare events.

A particularly compelling recent case demonstrating the value of long-term studies for revealing the complexity of ecological systems comes from the experimental work of Jim Brown and colleagues. Experimental removals of large kangaroo rats from permanent exclosures in the Chihuahuan desert of Arizona initiated in 1977 demonstrated the importance of competition in determining species abundances and diversity. Not only does species richness of other rodents increase as a direct effect of the competitive release in the exclosures, but after 12 years of rodent removal there was also an effect on the diversity and composition of the plant community (e.g. Heske, Brown and Guo 1993; Valone and Brown 1995).

Marine systems also provide excellent examples of how long-term data sets can be used to elucidate the roles of complex, slow, or rare processes or events. In some cases, events that would not have been observed (or understood) from short-term manipulative experiments can be interpreted from long-term observations of the system. For example, assessing the devastating impact of an introduced clam species to the community of San Francisco Bay required extensive pre-introduction data in order to understand the underlying dynamics of the natural system (Nichols et al. 1990). Also, Lively et al. (1993), studying a rocky shore community in the Gulf of California, found strong space-time interactions for most of the species (barnacles, mussels, algae, whelks). In particular, there was yearly amplification of spatial effects due to unpredictable differences in recruitment among years. They noted (p. 170) that, "controlled manipulative experiments, by ensuring the presence or absence of species in factorial combinations, elucidate mechanisms and determine the possible, while longer term studies help to determine the actual."

Long-term studies have also been important in testing fundamental questions and theories in ecology. Perhaps the most famous long-term study of plant communities is the body of work at the Rothamsted Experimental Station in England. Experiments were started there from the 1840's to the 1860's to test the effects of different manuring and fertilizer treatments on crop yield (Williams 1978, Johnston 1991). The experimental plots at Rothamsted have been used to test contemporary theories that were not even imagined when the experiments were begun. Examining plants under different fertilizer treatments has demonstrated population differentiation (Thurston et al. 1976). Analyses of the biomass data combined with climate records have permitted tests of the problematic relationship between diversity and stability (Silvertown et al. 1994).

One of ecology's central ideas is the concept of community succession. Although trends in succession are often safely inferred from the substitution of a spatial series of different ages for an actual series in time, the interaction between time and spatial variation makes this a risky assumption (Pickett 1989, Collins and Adams 1983). Therefore, long-term studies have been especially important in successional plant communities. Much has been learned from these studies that was invisible or incorrectly portrayed in chronosequences. The nature of species turnover through time, the role of spatial relationships, the significance of disturbance and episodic events in community organization, and the importance of priority effects, for example, have only been clarified as the number of long-term studies increased. The Buell-Small succession study, at the Hutcheson Memorial Forest Center, is described later in this report as an example.

Our understanding of freshwater systems has also benefitted from long-term studies. Repeated measurements in streams have yielded information on how they are organized by disturbance and succession (Reice 1994). Long-term studies in lakes have demonstrated the importance of trophic dynamics (Carpenter and Kitchell 1987), species introductions (Mittelbach et al. 1995), anthropogenic changes, and dynamics within watersheds (Butler and Malanson 1994). Because of the integrative nature of both lotic and lentic systems, long-term studies of these systems can indicate the dynamics of large areas (Likens 1992).

There are numerous examples from aquatic (marine and freshwater) systems of how long-term data have provided information on population- to ecosystem-level responses to anthropogenic effects. Two classic long-term studies that have documented these effects in lakes are W.T. Edmondson's study of Lake Washington (initiated in 1949) and C.R. Goldman's studies of Lake Tahoe (begun the late 1950s). Rapid changes occurred in these systems due to the increased sewage effluent and pavement runoff that accompanied development and human population increases. From theses studies we have a greater understanding of the impact of eutrophication (Goldman 1981, Edmondson 1991), food web dynamics (Murtaugh 1981), and the impacts of introduced species (Luecke 1990).

Perhaps the most pressing need for overall long-term data is in quantifying or detecting impacts of global warming. Barry et al. (1995) recently compared rocky shore communities in permanent transects that had been set up and surveyed repeatedly in the 1930's. Their data strongly suggest a shift to more southern-affinity species, corresponding with a gradual 2 degree warming in summer maximum sea temperatures. On a much larger scale, long time-series plankton sampling in the North Atlantic (the CPR program) and off the coast of California (the CalCOFI program which is discussed later) are providing significant insight into how climate affects oceanic systems, especially zooplankton (McGowan, 1990). For example, Roemmich and McGowan (1995) found a 70-80% decrease in the biomass of large zooplankton in waters off southern California since 1951, corresponding to increased surface temperatures and thermal stratification. Such a massive change must have far-reaching biological consequences that, without long-term data, might have been "blamed" on pollution or over-fishing.

Awareness of the importance of long-term studies to understanding ecological systems has led to increased funding and direct support for research programs that have long-term research as a primary focus. Perhaps the best known of these is the NSF LTER (Long-Term Ecological Research) Network which supports long-term research at 18 sites, 16 located throughout North America and two in the Antarctic. The ten-year review of the LTER program stressed the value of the LTER Network and encouraged a broadened base of support for the program to expand both the spatial scale (more regional emphasis) and the scientific issues (social and economic factors) that could be addressed (Risser and Lubchenco 1993).

Concerns about environmental quality, global change and biodiversity have increased the importance of long-term data for assessment and monitoring across broader geographic scales than is typically represented by existing studies. A number of new programs are being initiated or developed from existing programs to provide comparative data on a variety of environmental and ecological variables. In the United States this has involved a combination of new initiatives directions and expansion of existing monitoring programs that have been the tradition of many mission-oriented agencies (e.g., EMAP within the Environmental Protections Agency, National Biological Service within the Department of Interior, the Legacy Program within the Department of Defense). Within the NSF-LTER network there is a strong emphasis on expanding cross-site comparisons and synthesis to a regional scale. Some of these efforts have international linkages, especially to Canada and Europe. In Europe, several programs are being initiated that are specifically designed for monitoring environmental quality across a network of sites (e.g. the Environmental Change Network (ECN) in the United Kingdom; EuroMAB).

Special Challenges of Long-Term Studies

Despite the acknowledged value of long-term ecological data to both basic and applied research, it is extremely difficult to maintain funding for such studies. The challenges of conducting long-term studies fall into three broad categories: administrative problems, data management problems, and output problems. Individually these problems are not unique to long-term studies, but their importance is compound over time in a long-term study which makes such studies increasingly difficult to maintain.

Administrative problems include the expense of collecting, archiving and analyzing samples. This is especially true of physical samples such as soil or water, but it also includes collection, verification and curation of museum and herbarium specimens. All these activities are central to the success and utility of a long-term data set. Even the simple act of maintaining plot markers may involve considerable expense and effort. Similarly, training and retention of personnel capable of or interested in continuing a long-term study is an administrative burden.

Administrative problems also include the need to maintain funding for the project over long periods. Because long-term studies cannot continuously be "new" they may suffer due to fashion or a general bias towards and real need for new research. The desire of many ecologists to be entirely experimental or primarily theoretical may also work against continued support for long-term studies, although there are notable examples of the contribution of long-term data to both of these ecological approaches.

The second category of problems is the seemingly simple act of maintaining the data set. Many long-term data sets exist only on paper and one of the first tasks in maintaining a long-term data set is to generate secure electronic copies of the data. Error checking and correction become large tasks in transferring many data sets from paper to computer. Equally important to the data themselves are the metadata that describe them. Without metadata specifying exactly where and how the data were collected, the initial and on-going process of editing and correcting the data, the history of the data set and methodology that gave it rise, the data set itself can become worthless. Also, having a project evolve to include improved or more tractable methods is actually a potential threat to its continuity and comparability. Cross calibration, documentation, and temporal overlap of old and new methods is one solution to the need for evolution in a long-term study (Strayer et al. 1986).

Transfer of the data and metadata to new storage media as technology evolves is a major management problem for long-term studies. Often it is necessary to restore data sets and metadata from corrupted records and from obsolete storage technologies. Often the metadata and the data become separated, and if the separation is irreconcilable, it reduces the value of both to mere scrap. Later in this report we will point toward ways to avoid these fatal problems. The value and expense of maintaining long-term data sets suggest that careful attention to data management is of the highest priority. The time and effort involved in maintaining (and updating) data from a long-term study, however, needs to be considered in determining which studies should (and can) be maintained.

The final category of problems in maintaining long-term studies is output. Because the value of long-term studies is often only realized after some period of maturation, the significant output in terms of publications, raw data available to the user community, or conceptual and theoretical insight will be delayed compared to shorter term studies. Judging long-term studies on the basis of number of published works per unit time may show them to be rather inefficient in such simple terms. The value and efficiency of long-term data is of a rather different sort than the usual short term study.

The Role of the ESA

Science in the Electronic Age

Ecology traces its roots to early accounts of natural history and descriptions of patterns in natural landscapes. As the discipline has matured over the decades, this tradition has been transformed in many ways by both conceptual and technical advances. Nonetheless, most of the key concepts that have shaped the intellectual history of the discipline have arisen as interpretations of data, rather than by deduction from first principles (Schrader-Frechette and McCoy 1993). The modern emphasis on quantification, estimation, replication, experiments, and hypothesis testing has not changed this fact.

What has changed is the volume of data that ecologists produce and the means by which data are collected, as well as an increasing focus on synthesis and integration. Not only has the number of professional ecologists increased greatly, but automated acquisition and archival technology exponentiate the size and complexity of many ecological data sets. Anagenetic evolution from note cards and lab record books to punch cards and then to electronic and optical data storage technologies have made it increasingly possible to acquire and store large masses of data. First mainframes then mini- and micro-computers have allowed statistical analysis and parameter estimation for modeling from data sets that are far larger and more complex than anyone could have dealt with before the age of computers.

Continuing developments in electronic communications are producing revolutionary changes in the way we interact with one another professionally. This revolution promises to transform the way we publish our work and how we access and build on the published work of others. For example, the Ecological Society of America, in collaboration with the Mellon Foundation, will soon re-publish electronically all past issues of its three research journals, and will move towards simultaneous electronic and paper publication of new issues of these journals. Meanwhile, the ESA has joined in partnership with the founders of a new, all-electronic journal called *Conservation Ecology* (Colwell, 1995).

One might imagine that, by their very nature, long-term ecological data might require different approaches to data management than data collected for a short-term study. In fact, information never intended to become part of a long-term study often turns out to be essential for such studies, either as baseline or reference data or as a component of time-based data sets compiled from diverse sources (Strayer et al. 1986; Franklin 1989). The historical value of "slice-of-time" data sets often proves far greater than originally imagined. For this reason, observations and arguments regarding preservation and access apply with nearly equal force to all valuable ecological data sets, whatever the original intentions of their collectors.

This report presents the FLED committee's understanding of the challenges confronting ecologists interested in long-term ecological studies. It offers specific recommendations and alternative options on how to meet these challenges through the use of information technology, the shaping of professional policy, and development of sources of funding are offered.

Opportunities for Synthesis

Over the past decade the ESA has taken an increasingly active role in promoting and supporting initiatives that encourage synthetic activities to address both basic and applied questions. Prominent among these is the SBI (Sustainable Biosphere Initiative) in which the ESA, for the first time in its history, articulated a research agenda (Lubchenco et al 1991). The SBI report calls for more research in specific areas, including global environmental change, biological diversity, and sustainable ecological systems. Such research will require increasing cooperation among ecologists and scientists from other disciplines to develop the synthesis of data and expertise necessary to address these complex issues.

The extent to which synthesis and integration are recognized critical to the science of ecology is demonstrated by efforts that have been made over the past decades to establish an ecological synthesis center (AERC 1989, Brown and Carpenter 1993). The National Center for Ecological Analysis and Synthesis (NCEAS) was established in May 1995 at the University of California - Santa Barbara. The NCEAS was founded with the goal of promoting basic and applied collaborative research on the structure and dynamics of ecological systems. The synthetic activities needed to study these questions will require the integration of information across disciplines and scales and often will require use of long-term ecological data.

The report of the workshop that developed the plan for the NCEAS specified that major functions of the proposed Center should include means to (text relevant to long-term ecological data presented in bold-face for emphasis):

> provide a forum and mechanisms that facilitate new synthetic insights about ecological systems, and **to organize ecological information** in ways that make it more useful to decision makers concerned with problems such as biodiversity, global change, and sustainability,

and

> maximize the use of **the nation's data sets** to address ecological problems through synthesis;" (Brown and Carpenter 1993, p. 10).

The report further recognized the need for access to ecological data and stressed that research activities at the Center would include:

> in-house efforts followed by subsequent electronic access via high-speed networks **to access databases**, methodologies, and technological tools newly available to the national research community. " (Brown and Carpenter 1993, p.11).

The need for and concerns about accessing and using data from multiple sources for synthetic activities at the Center was also stressed in the report of the "Herndon Workshop", which developed a design study focused for the technical needs of the Center. The report specifically expressed concern about the need for:

> **Data access and analysis:** New paradigms in accessing and processing data are required. **There is not a national system for cataloging ecological data.** Hence finding and acquiring existing data sets is time consuming and expensive...."
>
> **Telecommunications:** .... Through its ongoing scientific activities, the NCSE (=Synthesis Center) can lead the evaluation of emerging information technologies and bring their benefits to the ecological community." (Anonymous, 1993; p. 8)

The establishment of the NCEAS provides the opportunity--and underscores the need-- to develop specific means for locating, restoring and preserving long-term ecological data. The call for proposals for the first two years includes as a major focus, spatiotemporal dynamics and specifically seeks proposals that involve,"The compilation, synthesis, and analysis of existing long-term data on population numbers, biomass, production or other ecological processes."

The ESA can--and should--play an important role in the development and establishment of a national system for cataloging and exchanging ecological data. In this report we specify what roles we feel the ESA--both at the Governing Board and membership level--can play in promoting and developing the establishment of such a national ecological data exchange system. The FLED committee emphasizes the critical importance of initiating this system for the preservation of long-term ecological data sets. In this report we outline recommendations for implementing mechanisms for locating (Section I), restoring (Sections 2 and 3) and preserving and accessing (Section 4) long-term ecological data.

The FLED Committee Findings and Recommendations

1. Sources of Long-term Ecological Data

Long-term ecological data are collected by **agencies, institutions, and individuals**. The incentives for data collection vary from regulatory mandates to protect endangered species or human health and scientific curiosity. As a consequence, the nature and extent of the long-term data collected by various sources is diverse and variable. Observations exist on rare species, pest species, agricultural production, toxin levels, pesticide levels, climate, hydrology, water and air quality. These data are taken and compiled at a range of spatial scales, from populations to communities to ecosystems and landscapes.

National agencies and organizations that collect long-term ecological data may be governmental or non-governmentally funded. Summaries of more than 25 national agencies and organizations that collect and/or manage long-term ecological data are given in Volume II, Directory 1 (see Table 1). Studies by government agencies have been central to our ability to address many long-term ecological questions. For example, the need to detect and explain changes in water quality has resulted in long-term monitoring and research projects such as the San Francisco Bay program. This monitoring program has been supported by the United States Geological Survey since 1967 and has provided critical baseline data to assess the relationship of human activity (e.g., dredging, commercial shipping, agricultural

Table 1. List of national agencies and organizations with long-term ecological data that are summarized in Directory 1 (in Volume II of the Report).

- I. U.S. Department of the Interior
  - A. National Biological Service
  - B. National Park Service
  - C. U.S Fish and Wildlife Service
  - D. Bureau of Land Management
  - E. U.S Geological Survey
  - F. Bureau of Indian Affairs
  - G. Bureau of Reclamation
  - H. Minerals Management Service
  - I. Office of Surface Mining Reclamation and Enforcement
- II. U. S. Department of Agriculture
  - A. U.S. Forest Service
  - B. Natural Resources Conservation Service
  - C. Agricultural Research Service
  - D. Economic Research Service
  - E. National Agricultural Statistics Service
  - F. Cooperative State Research, Education, and Extension Service
- III. U.S. Department of Commerce
  - A. NOAA (National Oceanic and Atmospheric Administration)
- IV. NASA (National Aeronautics and Space Administration)
- V. EPA (Environmental Protection Agency)
- VI. U.S. Department of Defense
  - A. Army Corps of Engineers
- VII. U.S. Department of Energy
- VIII. The Nature Conservancy
- IX. Pacific States Marine Fisheries Commission
- X. National Science Foundation

practices) to phytoplankton blooms and species invasions (e.g., Powell, et al. 1989a, 1989b, Nichols, et al. 1990, Cloern 1991). Another example of ecological research that capitalizes on agency data, are studies documenting population trends using the Breeding Bird Survey from the NBS (Bohninggaese et al. 1994, Taper et al. 1995, Robbins et al. 1989, Peterjohn and Sauer 1994). Long-term data from public agencies have also been used to examine the population dynamics of rare and pest species, community responses to grazing or fire, and abiotic factors that describe ecosystem dynamics.

Other institutions involved in collecting long-term ecological data include **private research laboratories, environmental organizations, museums, and field stations**. Many of these have curated, long-term data that are available, sometimes on the Internet, sometimes for a fee. Museums contain voucher specimens, as well as databases derived from specimens, and have a long history of free and open exchange with researchers. (see Section 4 for their data sharing experiences and Volume II, Directory 3 for Internet addresses of specimen-based databases.) Natural Heritage Programs have been established in most states, in conjunction with The Nature Conservancy (see The Nature Conservancy in Volume II, Directory 1), that maintain data on species distributions and abundance trends. Ecological research laboratories such as the J. W. Jones Ecological Research Center (Newton, GA), the Marine Biological Laboratory Ecosystems Center (Woods Hole, MA), and biological field stations also are important sources of long-term ecological data. (see Section 4 for data sharing experiences and Volume II, Directory 2 for data sets archived by biological field stations.)

Projects initiated and conducted by **individual researchers** may be the most abundant source of long-term ecological data, and they are clearly the most diffuse. Individual researchers are typically associated with universities or research institutions, and their research activities are frequently funded through competitive grants. Access to these types of data frequently depends upon an agreement between the user and the investigator. Sometimes the conditions put forth by the granting agency explicitly promote data sharing. For example, individuals involved in the NSF LTER program are requested to make their data available on the Internet one year after the project has ended (see Section 2). For all investigators awarded general grants, NSF expects "...significant findings from research and education activities it supports to be promptly submitted for publication, with authorship that accurately reflects the contributions of those involved. It expects investigators to share with other researchers, at no more than incremental cost and within a reasonable time, the data, samples, physical collections, and other supporting materials created or gathered in the course of work. It also encourages grantees to share software and inventions or otherwise act to make the innovations they embody widely useful and usable." (National

Science Foundation 1994, p.21).

Determination of "At Risk" Data

The long-term data collected by different sources fall along a continuum of risk of being lost. Although data collected by a govenment agency would appear to be the most secure, there are numerous examples of data sets being abandoned (or lost) as budgets diminish and agency priorities change. Among the least secure are data sets collected by individuals who have made little or no provision--or know of no option--for the long-term curation of their data. The following sections describe a variety of situations in which existing long-term data sets can become at risk in public agencies, in private institutions, and in individually-owned archives.

Public Agencies

The most secure long-term data are archived in a data bank with a history of solid funding, generally from a government source. Data held by the International Tree Ring Data Base and the American Pollen Database, both funded by NOAA's Paleoclimate Program, fall into this category (see Section 4). For these data, government support was essential to create the current system of archiving and disseminating data. At this point in time, were support discontinued, the currently archived data could be preserved at minimal expense.

On the other hand, data maintained by a government agency can become at risk for political or funding reasons. For example, hydrological monitoring at the San Dimas Experimental Forest near Los Angeles, California that began in 1937 was discontinued in 1960. Discontinuation of this project occurred in part because Forest Service data had shown that the local watershed could not provide the amount of water needed for a rapidly growing population. The geoecology database (see Olson in Volume II, Directory 2) provides another example. This database contains over 1000 variables (soil, climate, species, etc.) collected by many different agencies on every county across the conterminous United States. The data were merged to provide a composite picture of environmental conditions on regional scales. However, due to changes in agency focus, this massive database is no longer receiving funding to maintain or update it.

Data related to species monitoring are under variable degrees of risk depending on the degree of public support. For instance, the Breeding Bird Survey data are managed by the U.S. Fish and Wildlife Service, but are less likely to be lost if agency support is withdrawn because so many volunteers (individuals and organizations) are committed to collecting the data. On the other hand, weakening of the Endangered Species Act, which mandates monitoring of rare, threatened, and endangered species (much of which is carried out by the state--and The Nature Conservancy--supported Heritage programs), would result in a withdrawal of federal funds. It is unlikely that all of these species, especially those that are less charismatic, would be adopted for monitoring by private citizens or organizations.

Agency data are also vulnerable due to lack of consistency in agency administration and staff turnover. Data may be lost because the original data collector has left the agency. A smooth transition between data caretakers, including adequately documented metadata, is required for the value of the data to be realized. Even if the mandate still exists to continue data collection, a new researcher or new office director may change the emphasis or methods, often in an effort to make data compatible with modern statistical methods, but also making older datasets incompatible with newer ones.

Museums, Field Stations, and Private Research Institutions

The degree to which data held by museums, field stations, or private institutions are at risk of abandonment depends upon the structure and policies of the organization. Data managed by entities that have central curation for data, specimens, or photographs tend to be more secure than data managed by organizations that do not. While museums have a history of centralized curation, the specimens upon which their databases are derived can degrade or be lost due to lack of space or funding.

Biological field stations usually expect individual researchers to curate their own data. Deposition of data at the field station is typically not required, although there are notable exceptions. The Organization of Tropical Studies in Costa Rica and several field stations associated with the LTER program have specified data management policies and to varying degrees maintain data sets collected by investigators at these sites. In addition, field stations often have long-term research programs that are overseen by station directors (see Volume II, Directory 2).

Private organizations vary in their structure, policies, and methods of data curation. Perhaps the largest and best known of these is The Nature Conservancy which maintains three databases of interest to ecologists: The Conservancy Central Conservation Database, The Heritage Database, and the Vegetation Classification System (see Volume II, Directory 1 for more details). TNC frequently works in conjunction with other organizations to establish ecological data bases. The Heritage Database, for example, is available through the Natural Heritage Programs run by most state fish and game departments. As mentioned above, the Endangered Species Act creates a mandate for the Natural Heritage Program, subjecting this portion of The Nature Conservancy data to political whims.

Individuals

Among the least secure long-term data sets are those held by an individual researcher, many (most?) of whom have made little or no provision for long-term curation. Consequently, when this key figure retires or dies, the data may be lost due to inadequately documented metadata and the lack of time and energy required for others to resurrect the data. Individual researchers at universities and research institutions are the bulk of the ESA membership. Education of this group regarding the importance of documenting these data sets and making arrangements for their long-term maintenance is imperative if we are to assure that these valuable long-term ecological data sets are not lost.

Laws and regulations passed during the Reagan administration explicitly transfer ownership of intellectual property to the grantee (generally the University or institution, not the Principle Investigator!), however, currently universities take little responsibility for long-term data curation. If a researcher retires, dies, has a break in funding, or loses interest in or ability to continue a project, there is great risk that the data will be lost.

There are a few heartening stories of long-term data sets begun by one researcher and later followed by another, who continued data collection, curation, and analysis (see Section 2 for case studies). For example, Forest Shreve and colleagues established permanent plots for monitoring saguaro populations at Tumamoc Hill near Tucson, Arizona, in the early 1900s. In the 1950s, Ray Turner of the University of Arizona and the U.S.G.S. relocated and followed many of Shreve's original plots. Turner and Elizabeth Pierson, also of the U.S.G.S., have continued to maintain this legacy data set. Their publications provide valuable insight into saguaro recruitment and population fluctuations. However, despite wide interest in this remarkably long-term data set, it still is at risk because there is currently no funding to support this project (see Pierson, Volume II, Directory 2).

Searching for Long-term Ecological Data

Until recently, search strategies for locating long-term or other ecological data were very rudimentary. By and large, finding out about who had particular data sets was limited to access through formal and informal professional networks, library searches of scientific journals, or scouring through internal reports of federal agencies. Now, with the advent of the Internet, which allows rapid electronic communication among scientists, and the increasing availability of data on the Internet, more options exist (see Box 1). Internet access to data provides both the opportunity for ecologists to locate and use a variety of sources of ecological (and related) information but also the responsibility to properly acknowledge these sources of information. However, locating useful sources of information on the Internet is currently time-consuming and difficult because few directories exist and search tools are rudimentary. We expect that search strategies and tools for accessing information on the Internet will evolve rapidly. In its recommendations, the FLED Committee proposes that the ESA take an active leadership role to assure that tools useful to ecologists are developed.

Search Strategies for the Novice (An area of evolution)

Assume that you are interested in the impact of introductions on the species diversity of the freshwater fish community in the Western U. S. You have already done the classic library keyword search using on-line sources to publications (Sabio, Biosis, Agricola, Bioabstracts, Science Citation Index, Current Contents, etc). But you suspect that some valuable data related to your topic are not published. For example, relevant data may be on note cards in a box in someone's office or may be part of the grey literature generated by an ongoing monitoring program of a Federal or State agency. Besides word-of-mouth, what do you do? (Box 1).

> BOX 1: TODAY'S SEARCH STRATEGY.
> Get on the Internet! Right now, you may learn about some agencies and individuals that have long-term data by browsing the Internet. This semi-random process requires you to have access to an Internet provider and Netscape, Mosaic, or another Internet browser. Internet browsers are Network navigators that allow retrieval and viewing of Internet documents. With these tools you may search under the names of agencies you suspect have the type of data you seek: U. S. Fish and Wildlife, U. S. Forest Service, and NOAA, to name a few. Many of these national agencies have put or are in the process of putting actual data on-line and downloadable from the Net (but with no guarantees about the quality of the metadata!). In addition, many state and private agencies that collect long-term data are, or may soon be, on the Internet. While browsing an agency HomePage, look for appropriate clues about relevant research and monitoring programs and the names, phone numbers, and e-mail addresses of people to contact. In addition, you may hit upon one of several environmental clearinghouses that provide links to environmental data sources from federal and state agencies.

The unfortunate situation today is that the current search strategies on the Internet are generally limited to the hit and miss process of "surfing the Net", a time-consuming and redundant exercise. Despite the existence of several environmental clearinghouses, information on ecological data and agencies that collect and maintain data of interest to ecologists is diffuse and difficult to locate. While agencies have a mandate to make their data publicly available, and a growing number are posting it on the Net, what is lacking is a comprehensive catalog of long-term ecological data maintained not only by agencies, but by individuals who would like to foster communication and scientific collaboration. Such a catalog or network could serve to enhance the possibilities for synthetic studies, for the "adoption" of abandoned data sets, or for otherwise coordinating the efforts of researchers.

A first step toward remedying this situation is the publication of this report, which includes several Directories (in hard copy; see Volume II) to long-term ecological data. For maximum utility, these directories of long-term ecological data should be accessible on the Internet and searchable electronically.

Recommendations

To support the need of ecologists to locate, and promote the use of, existing sources of long-term ecological data, we recommend that the ESA:

1. Establish and maintain a HomePage of Long-term Ecological Data that would provide pointers to sources of long-term ecological data. The information in this HomePage would initiallly include the summaries and descriptors provided in Volume II of the FLED Report as Directory 1. (U. S. Agencies and Organizations), Directory 2 (Individual Long-term Ecological Data Sets), and Directory 3 (Annotated Bibliography of Existing Catalogs, Directories and HomePages with Long-Term Ecological Data).

2. Cooperate with and support efforts of agencies and other scientific organizations at the national level for the establishment of a network of ecological and environmental data sources and exchange (e.g., the NBII--National Biological Information Infrastructure of the NBS; and the NBIC--National Biological Information Center).

3. Educate and provide information to the membership on how to use and contribute to these national data exchange networks.

2. Options for Recovery and Restoration of Long-term Ecological Data

If long-term ecological data is to be a resource for persons other than those who collected the data, the information must be documented and stored in a manner retrievable by others. This leads us to ask: What options are available to support the restoration and storage of a long-term data set? What funding sources are available? What will it cost? What can we learn from those who have visited this territory before us?

The following case studies represent various ways in which data managed by individuals and organizations have been restored and/or are currently being curated with existing resources. They provide the backdrop for a discussion of potential funding sources for data restoration and what criteria can be used to determine which ecological data sets are most deserving of restoration.

Case Studies: Data Sets Managed by Individuals

The successful restoration of a data set requires the sustained interest, enthusiasm, and perseverance of one or a few individuals. In some cases, individuals who collect these data have temporarily solved the problem of long-term archiving and curation by passing their data sets on to others, not necessarily with funding. In other, rarer cases, individuals have acquired funding to preserve their data and established means for the studies to be maintained and/or the data made available to others. Nonetheless, all of these data sets are at some degree of risk because there is no permanent repository for ecological data.

The Lilac/Honeysuckle Phenology Data Sets

Background. In 1957, Dr. Joseph Caprio (Montana State University) established a network of sites to see whether plant phenology was an integrated descriptor of local climate. Data were collected on timing of leaf emergence, flowering, fruiting, and plant height of lilacs planted at AES weather stations. The project was supported by funding from the U.S. Agricultural Experiment Station/Western Region. His program developed into a network of over 2500 volunteers who supplied data from specific sites. In 1961, W.L. Colville (University of Nebraska) established a similar network in the Eastern Region. Both the Eastern and Western networks were supported with US-AES funding to several individual researchers in the U.S. and Canada. In 1971, the United States regional Agricultural Experiment Stations expanded the project to include data on both lilac and honeysuckle. At its peak, in 1973, data were supplied from over 300 stations. U.S.-AES support fort the project declined over the next decade and ended in 1986.

Current Restoration and Curation Activities. In 1987, Mark Schwartz (then at San Francisco State University, now at University of Wisconsin-Milwaukee) contacted the project leaders for contact former observers, obtain funding, and reinstate the phenology network. About 50 Eastern observers responded and currently continue to supply data to Schwartz. However, until recently the only support for the project was for mailing expenses provided by Schwartz's department at the University of Wisconsin-Milwaukee. However, in August 1995, Schwartz received a grant from the NSF Climate Dynamics Program (Division of Atmospheric Sciences, Directorate of Geosciences) for a 3-year project using these data ("Connecting Satellite and Surface Measures of Spring's Onset"). The NSF grant provides a small amount of funding ($500/year) for support and maintenance of the Eastern network; the Western network is still inactive. Data from the Eastern network are listed in NASA's Global Change Master Directory, and are freely available to interested researchers.

Studies of a Rocky Intertidal Community, Tatoosh Island, Washington

Background. Since 1968, Bob Paine and his students have been studying a variety of components of intertidal community structure on Tatoosh Island, Washington. Long-term data gathered include: measurements of zonation patterns of common plants and sessile animals; abundance of key predators and grazers; presence of specific long-lived organisms. In addition, shorter term experiments involving many elements of the community (plant and animal) have been conducted. Because Paine is now nearing retirement, he is concerned that: 1) his data be put into formats accessible and useful to others and 2) there be successors to the research effort on Tatoosh to provide continuity to the long-term data.

Current Restoration and Curation Activities. Paine recently obtained funding from the Andrew Mellon Foundation for a 3.5 year Experimental Transition Grant. The grant will support four activities necessary for the long-term preservation and maintenance of this data set. Funds will be used to purchase a computer and image-analysis software to archive more than three decades of photographic records of the study sites. Photographs will be projected and parallax-adjusted; the images will be stored on CD-ROM (30-50 yr shelf life) for archiving and use by others. Research materials such as field journals and experimental "tools" will be archived, probably in a University of Washington facility. Paine plans to devote considerable time in the next several years to clarifying and cataloguing the journals to make them easier for others to interpret. Paine also will create a detailed map of the island to facilitate geo-referencing of study sites and photographic sites. Lastly, Paine will train two 'successors' (former students, T. Wootten and C. Pfister) to continue some of the standard measurements taken over the preceding decades on Tatoosh. Both of them conduct their own research on Tatoosh Island, and it is likely that they would opt to continue working there in the future, thus maintaining continuity of data collection.

Studies of Old Field Succession, Hutcheson Memorial Forest, New Jersey

Background. In 1958, Drs. Murray and Helen Buell and John Small, of Rutgers University, began a long-term study of post-agricultural succession in old fields to test controversial hypotheses concerning mechanisms by which succession occurred. The Buells' wished to determine whether the Initial or the Relay Floristics Hypothesis of succession was correct (Egler 1952). In addition, the predominant mode for studying old-field succession at the time was to examine fields of different ages that presumably had the same successional trajectory (e.g., Bard 1952, for the same region), an assumption that was untested and not always justified (Pickett 1989). The study began in 1958 with the establishment of two old-fields and ended in 1966 after a total of 10 fields had been abandoned in pairs at 2 year intervals. The fields represented a variety of crop histories and abandonment histories. Each field has a total of 48 permanent plots (0.5 x 2.0 m). Percent cover for all species, and stem density for forest tree species, were recorded for each plot. From 1958 until the original investigators died, the study was maintained and supervised by Murray Buell (until 1975), John Small (until 1977), Helen Buell (until 1990). In 1978, Steward Pickett moved to Rutgers and joined the group and has maintained supervision for the study to date. The overlap between Pickett and Helen Buell was essential for successful continuation of the project. Data were stored only on paper until 1971 when the Buells' were convinced by Tom Siccama (Yale University) to enter the data (1958-1971) on IBM punch cards (Small et al. 1971).

Current Restoration and Curation Activities. In 1990, with support from an NSF LTREB (Long-Term Research in Environmental Biology) grant, Pickett (currently at IES, Institute for Ecosystem Studies, Millbrook New York) consolidated the field sampling regime, converted the data still on paper to electronic form, and established a data entry and quality control routine (with the help of Dan Wartenberg, University Medicine and Dentistry of New Jersey). Ultimately, data will be stored at two sites (Rutgers University and IES) as ASCII formatted data sets in a generalized data management system. The raw data on the original field sheets temporarily reside at IES, where current data entry and quality control take place. Photocopies exist at IES as well as at Rutgers (c/o E. W. Stiles, Director, Hutcheson Memorial Forest Center, Department of Biological Sciences). Raw data will eventually be returned to Rutgers for archiving at Hutcheson Memorial Forest Center.

Problems involved in the curation and recovery of this data set result from inconsistent formats for the field sheets through time, sparsely documented metadata, perishable field plot marking materials (until 1990), the long gap between when data were collected and when they were entered into electronic form, inconsistent taxonomy due to sampling the vegetation at a time of year when not all species are flowering, and the need to sample plants difficult to identify (e.g., seedlings, stressed and senescent plants).

Vegetation Composition and Gradient Analysis

Background. Beginning with his thesis work in the 1940s, over his career Robert H. Whittaker established a series of transects and quantified vegetation composition along elevational gradients at several sites in the United States and South Africa. The data from these studies include estimates of biomass and productivity for forests in the Great Smokey Mountains (Tennessee), the Santa Catalina Mountains (Arizona; in collaboration with Neiring), the Sierra Nevadas (California; with Westman) and several sites in New England, including Hubbard Brook. Numerous papers were published from this work (see Peet 1985) and he is still recognized as a leader in plant community ecology, particularly because of his contributions to gradient analysis, ordination techniques, community dynamics, and species diversity.

Current Restoration and Curation Activities After Whittaker's death in 1980, colleagues at Cornell (Peter Marks and Dick Root, in particular) sought a permanent repository for his research data and field notes. They worked with a graduate student in the History of Science (Bill Kimler) to assure that the material was properly preserved and accessible. The notebooks of raw data along with other material are maintained in the Cornell Archives. The collection is called the "Robert Harding Whittaker Papers" (Collection No. 4248). The Whittaker accession is largely paper copy stored in 3-ring notebooks that contain his field data, plot by plot, transect by transect. The notebooks require over 22 cubic feet and includes the raw field data for the Smokies, Santa Catalinas, and Siskiyous, as well as miscellaneous correspondence. Additional information on how to access the data can be obtained from the Library, Rare and Manuscript Collections at Cornell University Library, Ithaca, New York, 14853-5302 (Ph: 607 255-3530).

Kelp Forest Population Study, Point Loma, San Diego, California

Background. The Giant Kelp *(Macrocystis)* forest off Point Loma, San Diego County was first mapped at the turn of the century; commercial harvest began in the 1930s. In response to a marked decline in the kelp forest, in the mid 1950's Wheeler J. North began a research program. In the late 1950s during a strong

El Niño, the forest virtually disappeared; recovery began by in the mid 1960s. North continued his studies until 1970, when P. Dayton and M. Tegner initiated research which still continues to date. Since the data were collected over time periods which include 2 major El Niño events and several storms, they are useful for examination of both normal patterns and rare events. The data set includes various biological measurements: a) kelp densities from the center of the forest (1971-present) and in other areas (1980-present); b) invertebrate densities; c) sea urchin size frequency information. Physical measurements include: a) benthic temperatures (mid 1980s--present); b) benthic irradiance (early 1990s-present).

Current Restoration and Curation Activities. This active research program is not at risk, although funding is year-to-year. Data are archived in electronic form in Dayton's laboratory at Scripps Institute of Oceanography and are available with appropriate acknowledgment.

Case Studies: Data Sets Managed by Organizations and Agencies

Generally long-term studies initiated and maintained by an organization or agency are secure as long as the data collection meets an agency mandate or priority. However, as agency mandates and funding priorities change, often extensive long-term data sets are abandoned and become at risk of being lost. As these examples illustrate, frequently several organizations have worked together to restore and preserve long-term ecological data sets that belong to an institution.

Vegetation Transects, Jornada Experimental Range, Las Cruces New Mexico

Background. A 1992 survey by United States Department of Agriculture of data sets relevant to global change research revealed that one USDA site, the Jornada Experimental Range in New Mexico, had a large number of long-term data sets at risk of being lost to the public and the scientific community. For example, a 64 year record (1915-1979) of maps of vegetation basal area, used to assess effects of grazing and climate variation (primarily yearly precipitation variability), are stored on a single set of paper maps. In another experimental study, begun in 1938, long-term vegetation growth was established to evaluate changes after removal of creosote-bush shrubs, exclusion of rabbits, or seeding of plots. The data from these studies are recorded on disk and on paper. The data that were most at risk were a set of reduced-scale vegetation maps recording locations of grasses and shrubs in fixed plots between 1915 and 1979, for which only one paper set of records exists.

Current Restoration and Curation Activities. With funding from CIESIN (Consortium for International Earth Science Information Network) a data rescue project was initiated to preserve the data from these vegetation maps in a digital format. The maps were scanned and hand-digitized to create over 8000 graphic images that are accurately scaled to reflect plant size. For archiving, a computer system was designed to convert the map images to 3-D images that can be viewed as a time-series. Personnel involved in the data recovery include local scientists with knowledge of the sites and systems and computer scientists that developed the computer system and user interfaces. It cost $160,000 over a 2-year period to rescue this data set. The intent is to eventually have these data available over the Internet (see the Jornada Data Rescue Project website; http://shamu.psl.nmsu.edu/Jornada.html).

Studies at The Santa Rita Experimental Range, Arizona

Background Established in 1903, the Santa Rita Experimental Range (SRER) is the oldest range experimental station in the world. Research was primarily conducted by the USDA Forest Service until 1987, after which the University of Arizona College of Agriculture took over administration of the site. The 90+-year accumulation of data includes weather records; livestock use; vegetation change; land treatment studies; short-term, usually small scale tests or observations; and over 350 permanent photography stations. Access to these records at present is limited to on-site visitation and study of handwritten records stored in file boxes and cabinets.

Current Restoration and Curation Activities. To conserve and archive the data while collaborating with ongoing efforts to develop national standards for data management, creation of a geographic information system (GIS)-based archive of SRER studies has been proposed. This effort will rescue information about previous studies, assist the development of new studies and the SRER long range plan, attract additional research interest, and facilitate extramural research proposals. The projected budget for database design (including data structuring, standards, searching techniques, etc.) and the digitizing of 30 data sets (the conversion of tabular data and associated metadata from paper to digital files) is $286,000.

Several groups have pledged enough monetary and database support to begin the project of data restoration. In 1995, the University of Arizona and the United States Forest Service made contributions ($30,000 and $15,000 respectively), with the likelihood of continuing support. The EPA has expressed interest in funding digital conversion of the photographic record. The Agricultural Research Service (ARS) plans to dedicate about $10,000 to prepare roughly 20 years of SRER watershed data. In addition, an NSF proposal ($185,000) to support a cooperative effort of data restoration between SRER and the Jornada

LTER is in review. The next phase of the project will include application program interface efforts to improve subject searches and graphic user interface efforts to facilitate GIS-based spatial analyses. The goal for the SRER searchable index is to be compatible with library search software. Finally, Internet design will be developed to facilitate world wide access to the data. No cost estimate has been attached to this second phase as of yet.

Physical and Biological Surveys of the California Current: CalCOFI

Background. The decline of the California sardine fishery in the late 1940s stimulated the realization that successful management requires greater scientific understanding of the resource as it interacts with its physical and biotic environment. The California legislature approved a tax on sardine landings to study the causes of the sardine fishery decline. These funds were turned over to an industry-science Marine Research Committee that later established the California Cooperative Fisheries Investigation (CalCOFI) enterprise, a coordinating body for research by United States Fish and Wildlife Service, Scripps Institution of Oceanography, the California Fish and Game Commission, Stanford University, and the California Academy of Sciences.

While the sardine crisis catalyzed research interest, awareness that a fishery must be studied in the context of broad scale oceanography led to a push for integrated research that transformed the scope of ocean science (Scheiber 1990). A critical factor was the coordination of many agencies and disciplines dedicated to a holistic understanding of an offshore oceanographic system, the California Current. They established an extraordinary study area of some 670,000 square miles at several depths for detailed sampling of chemical and physical properties of the water and plankton abundance.

Current Restoration and Curation Activities. CalCOFI continues today as a partnership among three agencies: the California Department of Fish and Game, the Southwest Fisheries Science Center of the National Marine Fisheries Service, and the Marine Life Research Group of the Scripps Institution of Oceanography at the University of California San Diego. The backbone of the program continues to be regular surveys of the California Current including measurements of temperature, salinity, dissolved oxygen, nutrients, chlorophyll, current speed and direction, primary production, subsurface light, surface light, zooplankton (preserved samples and biomass indexes), larval fish, neuston, weather, seabird abundances, and phytoplankton (preserved samples). In response to extensive cutbacks in support, the surveys have varied in scope and resolution. However, time series data from 1949 include tens of thousands of measurements of environmental characteristics and a library of more than 43,000 samples of zooplankton and fish. A World Wide Web site (http//:www.scripps.edu), makes this and all Scripps data available. Data are at risk since funding is uncertain, despite a recent assessment that: "The CalCOFI program has produced one of the most consistent and highest-quality long-term data sets available for investigating the relationships between biological and physical processes." (National Resources Council 1995a, p. 79)

Sources of Funding for Restoration of Long-term Ecological Data

The difficulty in obtaining funding for the maintenance and conversion of long-term data sets to a more permanent (and easily accessible) format was one of the most frequently-cited concerns of ecologists who maintain long-term data (ESA Long-Term Study Section survey 1990, unpubl.). It is also one of the major reasons why such data are at risk of being lost. The successful avenues of support in the above case studies included some that kept data in the hands of individuals, while others institutionalized the data, sometimes involving several different organizations in partnership. Below we highlight current funding options that are available for caretakers of long-term ecological data. Later, we recommend ways for the ESA to enhance these options.

<u>The National Science Foundation (NSF)</u>. The NSF is an important source of support for much of the ecological research conducted by individual ecologists in the United States. Through continued funding, many projects that were initiated as short-term experiments or studies become, by the nature of the complexity of the system and the ability of investigators to obtain continued funding, long-term studies. Recognition of the need to develop more sustained funding for long-term research led to the development of the LTER (Long-Term Ecological Research) program at the NSF. The LTER program currently supports research at 18 sites (16 in North America and 2 in Antarctica) and a Network Office which facilitates cross-site synthesis and integration. Hundreds of individual scientists are involved in the United States LTER program and its success, nationally and internationally, is well-documented.

The Long-Term Projects in Environmental Biology (LTPEB) cluster in the Division of Environmental Biology at NSF includes two other programs (LTREB and RCSE) that support long-term research in both systematics and ecology. Both data management and data sharing are important components in the review and funding of projects in all of these programs. The LTREB (Long-Term Research In Environmental Biology) Program supports long-term research projects conducted by individual scientists outside the LTER program. In past years, it has supported 20-30 individual projects. In many cases, LTREB funding has been used to restore (or convert to electronic format) data sets that were available only in paper format (e.g., Murray Buell's studies of old-field succession at the Hutcheson Memorial Forest).

The Research Collections in Systematics and Ecology (RCSE) is another possible source of funding for the restoration and preservation of long-term ecological data. Although it is currently used primarily to support projects that convert information from natural history collections into electronic format, many of these collections have both ecological and systematic value. Funding for this program, though limited, is considerably higher than that available for the LTREB program ($6.5 vs. $ 1.0 million in 1995). This opportunity for funding of long-term ecological research collections (data and items) should be promoted in the ecological community, particularly when there is an opportunity to work in collaboration with systematists to preserve information valuable to both disciplines.

Other sources. Inadequate funds in NSF (and other federal agencies) to support the restoration and maintenance of long-term ecological data has forced ecologists to seek alternative sources of funding. Too often, restoration efforts are "bootlegged" from other projects or supported with personal funds. While there are private agencies that will support data maintenance or archiving of a valuable long-term ecological study (e.g. the Mellon Experimental Transition Grant to Bob Paine described above), these opportunities are currently very limited. However, private foundations and agencies could be receptive to initiatives from the ESA (or motivated individual ecologists) for the development of a grant program to support the restoration and preservation of long-term ecological data.

In another situation described above, several investigators took an interest in the SRER legacy data sets and approached multiple sources for funding. It is notable that most of the funding to this date is from the University of Arizona, and that these funds are now being used as leverage to attract funding from other agencies that have had some historic relationship to the project (e.g. USDA, USFS).

While our case studies deliberately highlighted the more promising solutions, some projects are struggling to stay afloat. For example, the Principal Investigator of the 27 year Hudson Bay Project have not been able to secure long-term funding to maintain or preserve these data. The data set includes information on 45,000 nests and 115,000 marked individual lesser snow geese and document the trophic cascade and habitat destruction resulting from over consumption by increasing goose populations (see Rockwell in Volume II, Directory 2). The Principal Investigator on this project (Fred Cooke) has moved to a new position, and the project is no longer eligible for NSERC (National Science and Engineering Research Council of Canada) funds. Although the current team of investigators (R. F. Rockwell, City University of New York; R. L. Jeffries, University of Toronto; and K. F. Abraham, Ontario Ministry of Natural Resources) is seeking more permanent sources of- funding to maintain the project (including NSF), for the interim they have put together a patchwork of support from a variety of sources, e.g., Central Flyway Council, Ducks Unlimited, National Rifle Association, and many small sportsmen's groups.

Selection Criteria for Restoration and Archiving of Long-term Ecological Data Sets

Establishing criteria for determining which data sets should receive the time, energy, and funding needed for restoration and curation is an important component for the development of any program that would potentially fund these efforts. The FLED committee acknowledges that every long-term data set probably has unique features which would make it valuable and therefore worthy of preservation. It is impossible to know the questions and hence the requisite data that ecologists will be evaluating in the future. For example, global warming and ozone depletion were not major issues of ecological concern 100 years ago. Nonetheless, data collections initiated then can provide important and otherwise unobtainable information on variables that we can now use to address these questions.

To maintain flexibility in evaluating the variety of long-term data sets that exist, we have developed general guidelines for determining which data sets merit priority for data restoration and/or inclusion in an ecological data archive (Section 4). We recognize that the development of these criteria must be an on-going process, as there will always be special cases and unique features of data sets that could be used to argue for their preservation. We propose the following guidelines and recommend that a peer-review process be established to prioritize individual data sets for restoration and preservation.

Important criteria to include in this evaluation are:

1. Metadata. The quality of the metadata is the single most important component of this decision. Higher level information about the data set, why and how it was collected, its content, quality, and structure will determine its ultimate utility. Relatively less detail is required for collaboration with an expert colleague, while more detail is needed for searchable third-party use (see Section 4).

2. Rarity. To some degree, every data set is unique. In fact, systematic observations about anything do not comprise a very large part of human activity, and by that yardstick most scientific data sets can be considered rare. However, some data sets might be considered more rare than others. For example, at the turn of the century, few scientists were making demographic observations on desert perennials, so a 90-year data set on a saguaro population turns out to be one-of-a-kind.

3. Length of record. The cost, in terms of money, time, and effort of recollecting a long-term data set may be exorbitant. In most cases they are simply irreplaceable. Exceptionally, long-term records are not only rare, they give us a dynamic picture of the variables measured. On the other hand, relatively shorter data sets that provide a snapshot in time still have value: rather than a dynamic picture, they provide us with baseline data.

4. Relocatability. Relocatability gives the data value for resampling. While the value of a site that is not relocatable is reduced, it may still have value on a large scale. For example, herbarium curators use older specimens, the collection points of which are not precisely known, to create presence/absence maps, or county dot maps, on a landscape scale. Likewise, a great deal of older plot data collected by agencies can only be approximately relocated. Plots from the Vegetation Type Map data collected in the 1920-30s by the U. S. Forest Service were located to within one hectare in southern California. This provided sufficient resolution to document the decline of native shrubs over hundreds of plots between 1920 and 1990 (Rich Minnich, pers. comm.).

5. Statistically analyzable. Data that were collected in a way that is congruent with our present statistical standards are the most valuable and deserve the highest priority for preservation. However, the lack of a 'statistically valid design' based on current statistical standards should not be used as the sole criterion to exclude a data set from preservation or analysis. Frequently, these standards did not exist at the time the study was designed, and the data in question may be the best there are. Statistical tools have been and will continue to be developed to deal with the types of problems that long-term data present (e.g., time series, Milliken and Johnson 1984, Barnett 1994, Barnett et al. 1994). Such breakthroughs have added value to the Rothamstad data, one of our longest and most classic ecological records. Finally, even if data are not "statistically analyzable" in the strictest sense, they can be invaluable for model parameterization and validation.

6. Size of spatial domain. Data collected from large spatial scales (e.g., regional and global) represent a wealth of information similar to long-term data, albeit from a different perspective. These data may be treated as having a value similar to those of large temporal scales.

Recommendations

A critical issue that distinguishes long-term ecological data from other types of ecological data is that often the data need to be 'restored' or converted to an electronic format before they can be analyzed or made available to others. Sources of funds to restore long-term ecological data are limited, and new sources need to be promoted and supported. The FLED committee therefore recommends that the ESA:

1. Work with its membership to promote funding for the restoration of valuable long-term ecological data sets within existing programs in NSF. The ESA should particularly promote opportunities for collaborative efforts with systematists to utilize funds from the RCSE program.

2. Work to identify alternative sources of funding for restoration of long-term ecological data sets and keep the membership informed of these possibilities through the ESA HomePage.

3. Establish a "gatekeeper committee" in conjunction with the Long-Term Studies Section (LTSS) to develop criteria for prioritizing data sets to be restored by scarce funds and to be preserved in an ecological data archive. This committee could also serve an advisory role for ecologists seeking appropriate archives for long-term data sets.

3. Demystifying Metadata *

The important role of metadata (i.e., data documentation) in facilitating ecological research has been recognized since the 1980s and several practical approaches to metadata management have been presented, much of it attributable to the leadership of data managers associated with the LTER network in the United States. Significant progress has been made during the past decade in developing metadata standards for geospatial data. For example, in 1994, a comprehensive set of Content Standards for Digital Geospatial Metadata was released that defines standard geospatial metadata descriptors related to data availability and accessibility, determination of fitness for use, and processing and utilizing a set of data. However, metadata standards for non-geospatial ecological data currently do not exist in any standard format beyond individual studies and experiments.

The lack of adequate metadata is a primary reason that long-term ecological data are at risk of being lost. The very nature of many long-term studies makes them more susceptible to having inadequate or irretrievable metadata. Many processes can lead to the loss of information about a particular data set or project through time. These include technical concerns such as the gradual (and inevitable) degradation of storage media containing the data, obsolescence of storage technology, or the loss of storage media through catastrophic events. Also, the memory and notes of the investigator are often the primary source of much of the information required to document a data set, and this source can be lost permanently (or be extremely difficult to retrieve) after project results have been published or the study has been terminated. Unfortunately, many of the specific details required to interpret a data set are lost when it is converted to an electronic format--data forms and field notes are often discarded when the data are digitally preserved.

It is important to stress that publications based on a long-term (or any) ecological study generally **do not** include sufficient information about how the data were collected to allow another scientist to use the data. The recommendations in this report include recommendations as to the metadata format and structure that would be needed for different categories of users. Exchange with an expert colleague or collaborator, for example, requires considerably less detail than with a third party or an audit by an agency.

Because the quality of the metadata is critical to the restoration of long-term

---

* Based on a manuscript submitted by: Michener, W. K., J. W. Brunt, J. Helly, T. B. Kirchner, and S. Stafford. 199x. Non-geospatial metadata for ecology. Ecological Applications (submitted, October 1995)

ecological data and is a decisive criterion for data sets that are currently accepted into permanent archives run by other biological disciplines, members of the FLED committee devoted considerable effort to identifying the minimal documentation required for an ecological data set. Here we: (1) examine potential benefits and costs associated with developing and implementing metadata for non-geospatial ecological data; (2) propose a set of generic metadata descriptors which could serve as the basis for a "metadata standard" for the ecological sciences; and (3) present alternative strategies for metadata implementation that meet differing organizational or investigator-specific objectives. Finally, we conclude with several recommendations related to future development and implementation of ecological metadata.

Benefits and Costs Associated with Metadata Implementation

Metadata, i.e. data documentation, may be defined as representing the higher level information or instructions that describe the content, context, quality, structure, and accessibility of a specific data set. Ideally, metadata comprise all information that is necessary and sufficient to enable long-term secondary use (re-use) of the data set by the original investigator(s), as well as use by other scientists who were not directly involved in the original research efforts. Thus, objectives for metadata implementation include facilitating: (1) identification and acquisition of data for a specific theme, time period, and geographical location; (2) determination of the suitability of data for meeting a specific objective; and (3) data processing, analysis, and modeling.

Scientists often refer to the rows and columns of numeric or encoded observations as raw data. Raw data are useful only when they can be framed within a theoretical or conceptual model. Relating raw data to underlying theoretical or conceptual models requires understanding the types of variables that were measured, measurement units, potential biases in the measurements, sampling methodology, and other pertinent facts not represented in the raw data, i.e., the metadata. The combination of raw data and metadata within a conceptual framework produces information.

Technological advances can make data collected in earlier times obsolete. Automated data collection procedures can now overwhelm our ability to effectively store, retrieve, manage, and analyze data (Stafford et al. 1994), which has sometimes necessitated the implementation of procedures to purposefully discard some data. It is the premature loss of useful data, such as long temporal sequences or irreplaceable data, that is a major scientific concern. The preservation of metadata is particularly problematic because metadata encompass a diverse and variable collection of facts that are often not recorded in any systematic way, including some facts which may reside only within the minds of the researchers.

There are many processes which can lead to the loss of information through time (Figure 1).

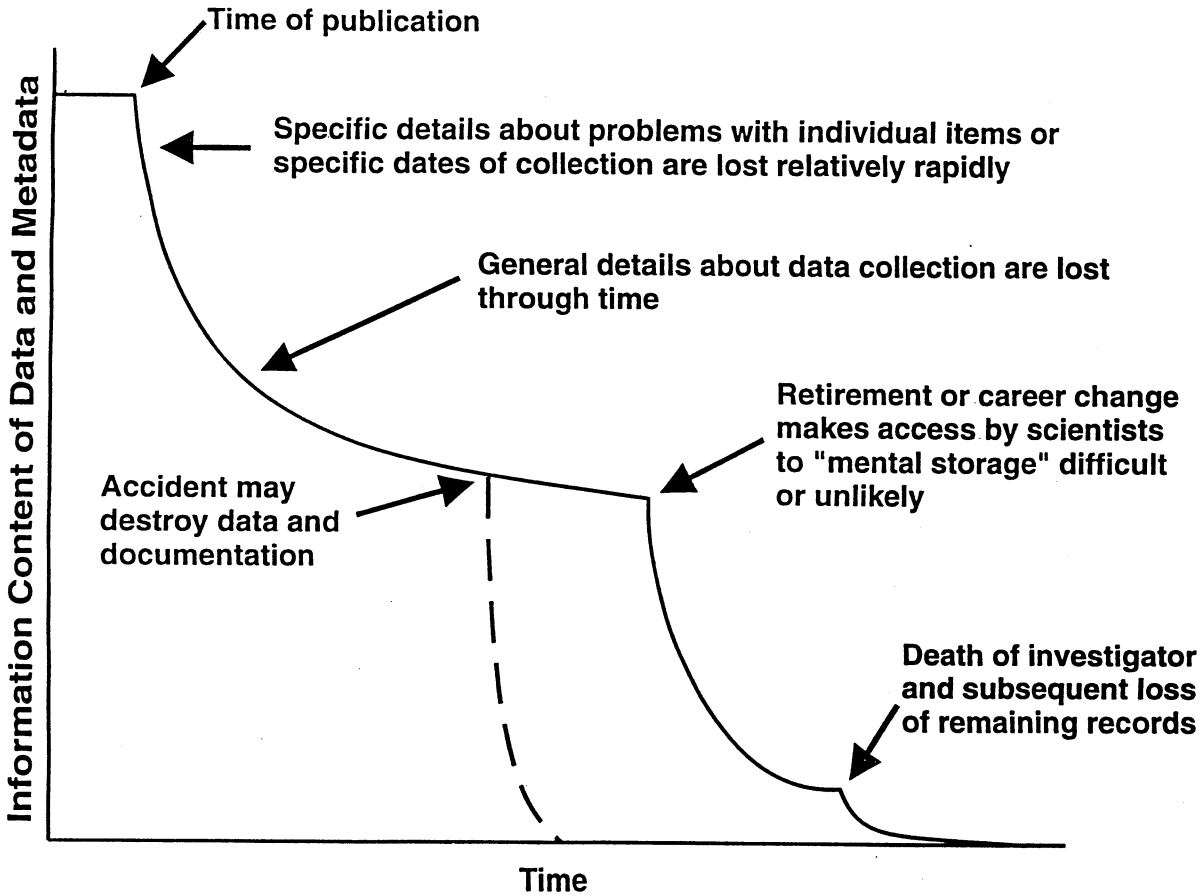


Figure 1. Decline in database metadata information content after project completion.

Some of these processes operate continuously, such as the gradual degradation of storage media containing the data, whereas others can be categorized as discrete events, such as the retirement or death of the scientist who collected the data, obsolescence of storage technology, or the loss of storage media through catastrophic events. Although loss of metadata can occur throughout the period of data collection, the rate of loss is likely to increase after project results have been published or the study has been terminated. Specific details are likely to be lost first due to abandonment of data forms and notes in lieu of digitally preserved data and through loss from the memory of the investigator. Over longer time periods degradation of storage media and further memory losses can reduce the information about general details not covered in relevant publications. Retirement or other major career changes may lead to the physical loss of records and hamper access to the investigator's recollections regarding data. Bowser (1986), for example, documents many of the problems associated with archival, metadata, and quality assurance that were encountered during attempts to re-analyze data collected from 1926 to 1941 by E. A. Birge, C. Juday, and collaborators in Wisconsin lakes.

Ecologists also lose information through the loss of conceptual models used to help interpret the data. Such models are often simple and can be expressed using statistical models to represent relationships between variables. However, some data sets, particularly long time series, are associated with hypotheses involving complicated non-linear relationships that are best represented by complex simulation models. Thus, preservation of the information about a data set may also involve the preservation of the simulation model and its associated input and output files (Kirchner 1994). Peer-reviewed publications featuring simulation models tend to focus on the results and the conceptual and mathematical foundations for the model. Because simulation models tend to be modified through time, preservation of the model code and input files is critical if model experiments are ever to be truly reproducible.

Both benefits and costs accrue during the development, implementation, and maintenance of metadata. In the following discussion, we present some of the benefits and costs that are associated with metadata implementation. An example from the International Biological Program (IBP) illustrates many of the difficulties encountered in attempts to reinvigorate extant data sets and highlights the importance of well-conceived and adequately maintained metadata.

Benefits The most important reason to invest the time and energy in developing metadata is that **human memory is short**. If data are to undergo any secondary usage, then adequate metadata will be required even if that secondary usage consists of re-use by the data originator. Scientists have long recognized the importance of preserving information but have often focused only on preserving the results of their synthetic activities through publication. Publication typically preserves some of the metadata but often only a subjectively selected portion of the metadata needed to relate the data to a specific hypothesis. To aggravate this scenario, ecological data sets are often extremely complex. Missing values, mid-course modification of sampling or laboratory procedures, addition or deletion of study parameters, personnel turnover, plot or habitat modification by disturbances (natural and anthropogenic) or changing environmental conditions, and numerous other factors leading to data anomalies are commonplace. Adequate documentation (metadata) of sampling and analytical procedures, data anomalies, and data set structure will help ensure that data can be correctly interpreted or reinterpreted at a later date. Twenty years is often established as the objective for having data usable by scientists that are unfamiliar with the data and their collection ('the 20-year test'; Webster 1991, Strebel et al. 1994).

In addition to the limitations of human memory, significant changes in the scope of ecology further underscore the critical role of metadata in supporting science. For example, the life span of a typical ecological data set that was collected ten years ago may have been very short, lasting from data set conception to publication, roughly corresponding to the average funding cycle of two to three years. At best, many such data sets met their resting place as dusty file folders of poorly documented data relegated to the bottom drawer of a filing cabinet. History and personal experiences are ripe with examples wherein data became useless because relevant metadata were missing or unavailable (National Research Council 1995a). More recently, however, increased interest in long-term ecological research (Franklin et al. 1990), comparative studies (Pace 1993), and expansion of the spatial, temporal, and thematic scales of basic and applied ecological studies (Levin 1992) have resulted in data sets being used for multiple purposes, often repeatedly over long periods of time.

Metadata provide the information that is critical for expanding the scales at which ecologists work. Comparative studies including temporal comparisons among sites, statistical replication, and comparisons within and among sites all depend upon the availability of sufficient metadata. For example, calibration and intercalibration (measurements of similar parameters by different methods or instruments) of methods and instruments should be well-documented in order to confirm data integrity and proper use of experimental methods and data acquisition. Similarly, ground-based reference data from multiple sites are frequently used to calibrate or support analyses of remotely sensed data, thereby expanding the spatial domain from the site to the landscape, region, or globe. Metadata are critical for combining physical, chemical, and biological data sets that contain different parameters, but share common spatial or temporal domains. Many short-term studies serendipitously evolve into long-term studies. In some cases, relatively short- to medium-term time-series data (possibly from different investigators or research programs) are integrated into a single long-term record. Metadata are essential for maintaining a historical record of long-term data sets that have resulted from such integration efforts, as well as documenting changes in personnel, methods, and data anomalies in ongoing long-term projects. Over the course of a long-term project, field and laboratory equipment are frequently replaced by other instruments that offer better precision or improved data acquisition methods (e.g., remote data loggers, etc.). Bowser (1986) and Strayer et al. (1986) discuss the importance of method intercalibration, quality assurance, and metadata for supporting reliable long-term data sets.

Synthesis and modeling projects are often hindered by the lack of high quality data and metadata. For example, ecological modelers routinely extract parameters from publications. Frequently, publications do not provide sufficient information pertaining to the data distribution, requiring many assumptions by the modeler about data ranges, frequency distributions, percentiles, etc. Ideally, raw data would be available for the modeling project, as well as the metadata that are critical for describing data collection objectives and methods, scale relevance of the data, and other potential limitations for secondary usage. For example, data collected under the auspices of IBP at the Andrews Experimental Forest in the 1970's were published as data summaries in internal technical reports (Emmingham and Lundburg 1977). Currently, entire Long-Term Ecological Research (LTER) data sets collected at the Andrews Experimental Forest are accessible on-line via the World Wide Web.

Costs High costs, primarily in terms of personnel time, can be associated with initially developing metadata. For short-term projects, the metadata file size and level of effort expended in developing metadata may exceed the physical size of the raw data file and the efforts expended in data collection. Real costs are associated with editing data and metadata and making them available in hard copy or electronic formats to the scientific community. Research grants and other existing funding mechanisms are often insufficient to support development of a comprehensive set of metadata.

Stewardship and a continuing need for curation and maintenance of the data and metadata represent real cost burdens which are not often factored into project budgets. After a study is deemed completed, who bears the responsibility of informing the user community of changes to the data set and newly discovered anomalies? Furthermore, critical details can often be overlooked in even the most comprehensive metadata. However, the role and appropriateness of funding for data originators as metadata consultants have received little attention and should be considered.

Data and Metadata Entropy Through Time: An Example

Consider the history of data collected for the IBP Grassland Biome. Data were collected by investigators at several sites within the biome and sent to the Natural Resources Ecology Laboratory (NREL) at Colorado State University (CSU) for processing and management. Most of the data were recorded on standardized paper forms that were color coded for convenience of the investigators. Data were transcribed from the forms to punched cards by professional keypunch operators. Much of the data was then transferred to 7-track magnetic tapes for storage. Metadata were distributed among technical manuals, which specified methods for data collection; peer reviewed publications; and the data form, which specified items such as units and species codes and included room for comments that were typically not keypunched when the forms were transcribed. The cards, tapes and many data forms were preserved by the NREL after the end of the IBP project but without the benefit of active maintenance, such as the periodic replacement of magnetic tapes. When the CSU computer center made the transition from 7-track to 9-track tapes, scientists obtained funding from the National Science Foundation to transfer as much data to the new tapes as was feasible without extensive effort. Those tapes which could not be read were abandoned and an attempt was made to recover the data from the card decks. However, the combination of old card decks and antiquated card readers meant that some data could not be recovered. In addition, some data sets were stored in machine-specific packed binary files that could no longer be decoded. Card decks were discarded after the CSU computer center abandoned card readers.

At the start of the Central Plains Experimental Range LTER project an attempt was made to recover data stored on the 9-track tapes and to preserve the IBP data forms using microfiche. Budgetary limitations restricted preservation activities to data specific to the Pawnee Site, although some data from other sites were preserved as well if it was convenient or of immediate interest to an LTER scientist. Some of the tapes had degraded to a point where they could no longer be read. Numerous problems were encountered with the microfiche process, including the fading and bleeding of inks on the forms and difficulty in getting clear photographs from the darkly colored forms. Although attempts were made to assemble metadata, the metadata for any specific data set were often incomplete or absent since publication histories were not linked to the data sets, and many of the original investigators were no longer able to provide requisite information. Thus, although most of the raw data associated with the CPER site were successfully preserved after considerable effort, it is also true that critical metadata were lost, thereby reducing the value of many of the historic data sets.

Standard Ecological Metadata Descriptors

The previous discussion highlighted the importance of metadata for providing scientists with the information that is necessary to re-use previously collected data. The IBP example further documented many problems associated with preserving data and metadata over a single decade. Assuming that some ecological data sets are inherently valuable and that we desire to preserve them and provide the relevant metadata required for sound secondary usage, we are still left with the problem of determining what metadata are essential for supporting data re-use. Fortunately, the process that scientists might normally follow in acquiring and utilizing existing data sets provides a guide to what metadata may be required. For example, once a scientist recognizes the need for specific data, several questions (or steps) would need to be addressed in an orderly fashion: (1) What data sets exist that might meet specified objectives?; (2) Why were those data sets collected and are they “fit” for my particular use?; (3) Can and, if so, how can these data sets be obtained?; (4) How are the data organized and structured?; and (5) Is there any additional information available that would facilitate my use and interpretation of the data?

The five steps hypothetically followed during identification, acquisition, and utilization of data serve as the basis for the five classes of metadata descriptors that are listed in Table 2 (based on Michener et al. 1987, 1990; Federal Geographic Data Committee 1994; Kirchner et al. 1995). [It should be noted that the term 'data set' as used in this discussion is synonymous with 'data table' which frequently appears in the computer science literature.] Class I includes basic attributes of the data set (data set title, associated scientists, abstract, and keywords) that are frequently included in hardcopy and electronic data set catalogs (e.g., Michener et al. 1990). The purpose of Class I descriptors is to alert potential secondary users to the existence of data sets that fall within specific temporal, spatial, and thematic domains. Preliminary determination of fitness-for-use by secondary users can be facilitated by incorporating a brief discussion of the scientific context of the data and a description of potential uses of the data set into the abstract. In many cases, a short summary of the "data set usage history" (Table 2, V. G.; especially the data request history and questions and comments from secondary users) could be used to identify potential uses of the data and highlight major strengths and weaknesses.

Class II includes all relevant metadata that describe the research leading to the development of a particular data set. Two sub-categories of research origin descriptors are necessarily included. The first sub-category includes a description of the broader, more comprehensive project (e.g., LTER program at a specific site) which may have led to numerous spin-off projects from which individual data sets emerged (e.g., climate, primary production, decomposition, etc.). The purpose of the "overall" project description is to provide the broader scientific context for an individual study. If an individual data set emerged from a stand-alone project, then the "overall" project description is superfluous. The second sub-category includes all pertinent information related to the research origins of a specific data set. Site characteristics, experimental or sampling designs, research methods, and project personnel are described in detail. Two additional descriptors may be essential for some data sets. First, permits are required for research and collecting on public lands within the United States and for importation of specimens into the country. Thus, permit history may be especially critical for museums that archive physical specimens, especially if museum personnel were not involved in the research and do not have the permit records. Second, many environmental monitoring and compliance data sets are generated in response to legal and organizational requirements. In such cases, it is important to document relevant laws, decision criteria, compliance standards, and other factors that may affect study design and data collection (Eagan and Ventura 1993).

Class III metadata describe the status of the data set and associated metadata, as well as information related to data set accessibility. Data set accessibility is affected by numerous factors which should be fully documented in the metadata.

Table 2. Standard ecological metadata descriptors and examples (based on Michener et al. 1987, 1990, FGDC 1994, Kirchner et al. 1995).

I. Data set descriptors	
A. Data set identity	Title or theme of data set
B. Data set identification code	Database accession numbers or any site-specific codes used to uniquely identify a data set
C. Data set description	
1. originator(s)	Names and addresses of principal investigator(s) associated with data set
2. abstract	Descriptive abstract that summarizes research objectives, data set contents (including temporal, spatial, and thematic domain), context, and potential uses of the data set
D. Keywords	Location (spatial scale), time period and sampling frequency (temporal scale), theme or contents (thematic scale)
II. Research origin descriptors	
A. "Overall" project description	{Note: this section may be essential if data set represents a component of a larger or more comprehensive database; otherwise, relevant items may be incorporated into II.B.}
1. identity	Project title or theme
2. originator(s)	Name(s) and address(es) of principal investigator(s) associated with project
3. period of study	Date commenced, date terminated or expected duration
4. objectives	Scope and purpose of research program
5. abstract	Descriptive abstract that summarizes the broader scientific scope of the "overall" research project

6. source(s) of funding — Grant and contract numbers, and names and addresses of funding sources

B. "Specific sub-project" description

1. site description

a. site type	Descriptive (e.g., short-grass prairie, blackwater stream, etc.)
b. geography	Location (e.g., latitude/longitude), size
c. habitat	Detailed characteristics of habitats sampled
d. geology, landform	Soils, slope/elevation/aspect, terrain/physiography, geology/lithology
e. watersheds, hydrology	Size, boundaries, receiving streams, etc.
f. site history	Site management practices, disturbance history, etc.
g. climate	Descriptive summary of site climatic characteristics

2. experimental or sampling design

a. design characteristics	Description of statistical/sampling design
b. permanent plots	Dimension, location, general vegetation characteristics (if applicable)
c. data collection period, frequency, etc.	Information necessary to understand temporal sampling regime

3. research methods

a. field/laboratory	Description or reference to standard field/laboratory methods
b. instrumentation	Description and model/serial numbers
c. taxonomy and systematics	References for taxonomic keys, identification and location of voucher specimens, etc.

d. permit history	References to pertinent scientific and collecting permits
e. legal/ organizational	Relevant laws, decision criteria, compliance standards, etc.
4. project personnel	Principal and associated investigator(s), technicians, supervisors, students

III. Data set status and accessibility

A. Status

1. latest update	Date of last modification of data set
2. latest archive date	Date of last data set archival
3. metadata status	Date of last metadata update and current status
4 data verification	Status of data quality assurance checking

B. Accessibility

1. storage location and medium	Pointers to where data reside (including redundant archival sites)
2. contact person(s)	Name, address, phone, fax, electronic mail
3. copyright restrictions	Whether copyright restrictions prohibit use of all or portions of the data set
4. proprietary restrictions	Other restrictions which may prevent use of all or portions of the data set
a. release date	Date when proprietary restrictions expire
b. citation	How data may be appropriately cited
c. disclaimer(s)	Any disclaimers which should be acknowledged by secondary users

5. costs	Costs associated with acquiring data (which may vary by size of data request, desired medium, etc.)

IV. Data structural descriptors

A. Data set file

1. identity	Unique file names or codes
2. size	Number of records, record length, total number of bytes, etc.
3. format and storage mode	File type (e.g., ASCII, binary, etc.), compression schemes employed, etc.
4. header information	Description of any header data or information attached to file {Note: header information may include elements related to 'variable information' (IV. B.) and, if so, could be linked to the appropriate section(s)}
5. alphanumeric attributes	Mixed, upper, or lower case
6. special characters/ fields	Methods used to denote comments, 'flag' modified or questionable data, etc.
7. authentication procedures	Digital signature, checksum, actual subset(s) of data, and other techniques for assuring accurate transmission of data to secondary users

B. Variable information

1. variable identity	Unique variable name or code
2. variable definition	Precise definition of variables in data set
3. units of measurement	Units of measurement associated with each variable
4. data type	
a. storage type	Integer, floating point, character, string, etc.
b. list and definition	Description of any codes that are associated with variables

c. range for numeric values	Minimum, maximum
d. missing value codes	Description of how missing values are represented in data set
e. precision	Number of significant digits
5. data format	
a. fixed, variable length	
b. columns	Start column, end column
c. optional number of decimal places	
C. Data anomalies	Description of missing data, anomalous data, calibration errors, etc.

V. Supplemental descriptors

A. Data acquisition	
1. data forms or	Description or examples of data forms, automated data acquisition methodsloggers, digitizing procedures, etc.
2. location of completed data forms	
3. data entry verification	Procedures employed to verify that digital data set is error-free
B. Quality assurance/ quality control	Identification and treatment of outliers, description of quality assessments, calibration of reference standards, equipment performance results, etc.
C. Related materials physical specimens, field notebooks	References and locations of maps, photographs, videos, GIS data layers, comments, etc.
D. Computer programs and data processing	Description or listing of any algorithms used in deriving, processing, or transforming data algorithms

E. Archival

1. archival procedures	Description of how data are archived for long-term storage and access
2. redundant archival sites	Locations and procedures followed

F. Publications and results Electronic reprints, listing of publications resulting from or related to the study. graphical/statistical data representations, etc.

G. History of data set usage

1. data request history	Log of who requested data, for what purpose, and how it was used
2. data set update history	Description of any updates performed on data set
3. review history	Last entry, last researcher review, etc.
4. questions and comments	Questionable or unusual data discovered by secondary from secondary users, limitations or problems encountered in specific applications of the data, unresolved questions or comments

In some cases, copyright or other legal restrictions (e.g., state or Federal laws restricting access to maps of endangered species locations, etc.) apply to specific data sets. In other cases, various proprietary restrictions may apply. For example, research laboratories, universities, and funding agencies frequently require that the grant which funded the research or that the institution or site where the research was performed be appropriately cited.

Class IV metadata describe all attributes related to the structure of the data file. All variables should be labeled, defined, and characterized as to data type and format. Finally, all known data anomalies (e.g., missing data, etc.) are fully documented.

Class V metadata include all other related information that may be necessary for facilitating secondary usage, publishing the data set, or supporting an audit of the data set. In some cases, for example, a scientist may find it necessary to review raw data forms, quality assurance/quality control procedures, computer programs and algorithms, and publications resulting from the data set. In addition, it may prove necessary to examine existing field notebooks or physical specimens. Pertinent data for physical specimens may include references to accession records/numbers (e.g., the transfer of a group of voucher specimens to a museum), specimen numbers assigned by the collector or the collection, and linkages among different forms of physical vouchers (e.g., sound recordings, chemical analyses, etc.) with different parts of physical specimens. Archived maps and photographs may facilitate resampling of a specific site.

Metadata may also serve as a vehicle for user feedback and data anomaly reporting. A "data set usage history" (Table 2, V. G.) may greatly facilitate long-term utilization of important data sets. There is no unique minimal and sufficient set of metadata for any given data set since sufficiency depends on the use(s) to which the data are put. Since uses may vary, problems with data and metadata should be recorded and retained as a usage history of the data. This can be thought of as attaching 'post-it notes' to the data to alert subsequent users to idiosyncrasies within the metadata or to anomalies within the data. However, direct modifications of the data should only be made by the data set owner/originator.

Metadata Implementation

A primary objective of metadata development and implementation is to facilitate data re-use by the data originator as well as support research activities by other scientists (Briggs and Su 1994). Figure 2 illustrates how metadata may evolve during the course of a specific project, as well as how secondary users (a modeler, in this example) might interact with the data and associated set of metadata. Specifically, hypotheses/questions and generic descriptions of the experimental or sampling design may be incorporated into the data set description (I.). Other more detailed aspects of the project design, including field and laboratory procedures would be included in the "specific sub-project" description (II.B.). Information related to data collection, data entry, and QA/QC is relevant to data set status and accessibility (III.), data structural descriptors (IV.), and supplemental descriptors (V.). Descriptors related to the "complete" digital data set are relevant to all metadata classes except perhaps those related to the research origin (II.). Finally, information obtained during analysis, synthesis, modeling, and publication may be incorporated into the supplemental descriptors (V.) in order to facilitate secondary utilization.

Subsequent to completion of the original project, a modeler may become aware of the existence of a particular data set via perusal of an organization's metadata database or data catalog which may contain "generic" data set descriptors (Figure 2, I.). The decision to acquire and understand a particular data set as well as mechanisms for doing so would require that the modeler have access to all relevant metadata that describe the origin of the data set (II.), it's status and accessibility (III.), structure of the data set (IV.), and possibly other miscellaneous details (V.). During the course of model execution, hypothesis testing, and model validation new information may come to light about the data set which could benefit other secondary users. This information plus a listing of publications resulting from secondary use of a data set would ideally be incorporated into the data set usage history (V.G.) in order to facilitate additional secondary usage.

The development and maintenance of metadata can be a very costly endeavor. Thus, it may be important to attempt to match the level of metadata <u>content</u> and <u>format</u> to the needs of anticipated users. As an example, we have identified three levels or types of secondary data utilization (Table 3; also see Kellogg Biological Station 1982). This categorization is derived from the identification of at least three types of data re-users and the recognition that the content of the metadata must increase at each level. The first, a Level I data re-user, may be a colleague or collaborator with technical expertise in the subject area and adequate knowledge of data collection, analytical, and processing procedures. Thus, such an individual may only require a basic description of the data set, as well as more detailed data structural descriptors in order to make effective use of the data set. A Level II data re-user may be someone who is searching a metadata catalog for reference or comparative data and would be using the data 'in-the-blind' (i.e., without direct contact with the originators of the data). In addition to data set and data structural descriptors, such an individual would also require much more detail about research origins and data set status and accessibility. Finally, a Level III data re-user might be conducting an audit of the data for ethical or environmental litigation or conducting a peer review for a citable publication involving second-party reproduction of computational results.

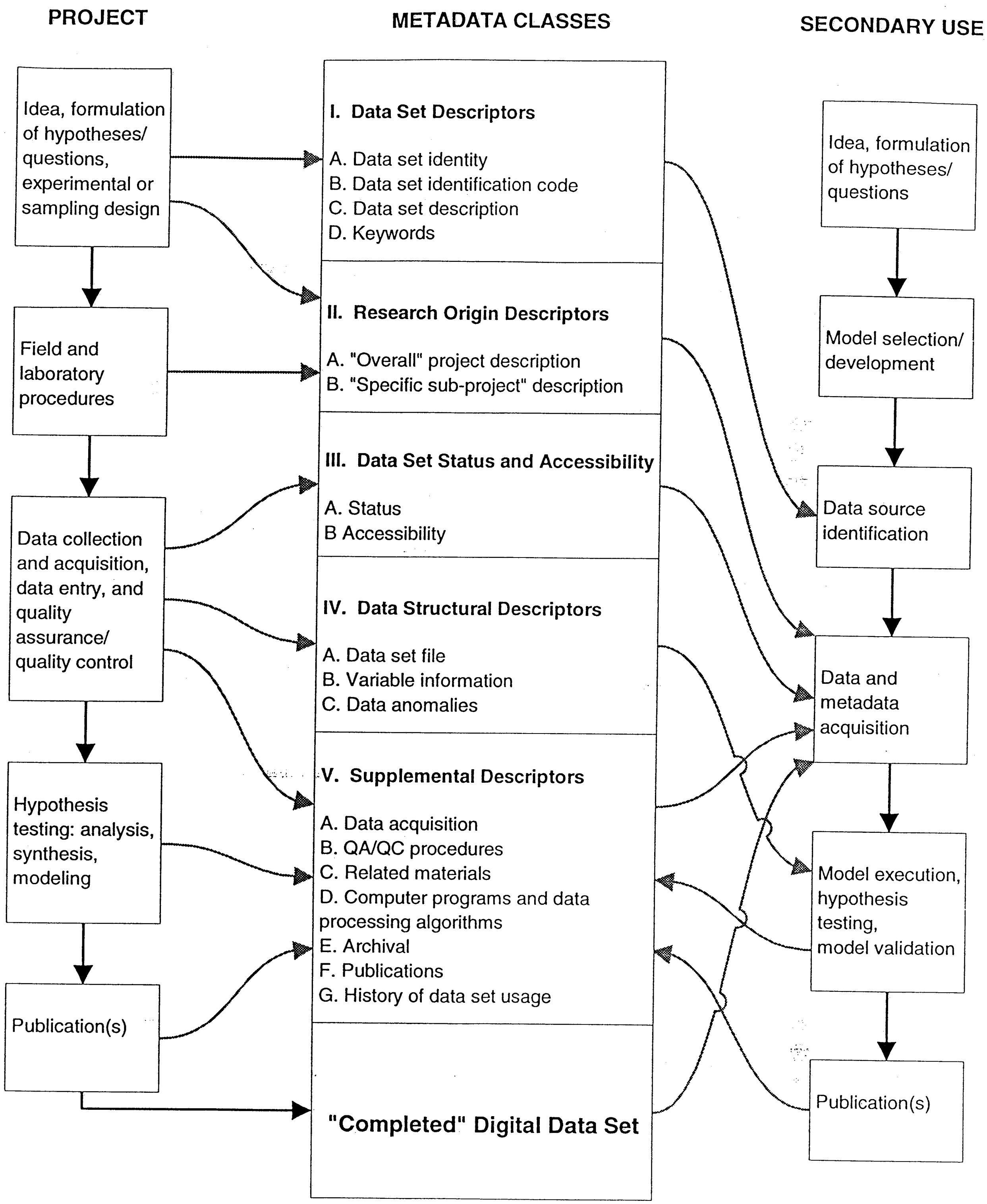


Figure 2. Metadata development (middle column, refer to Table 2 for more information) in relation to project design and implementation (left column), and relationship of project metadata with subsequent secondary data set utilization in a modeling project (right column)

Satisfying this objective may require that the individual have access to the most comprehensive set of metadata, including all supplemental descriptors. These examples illustrate the relationship between several types of secondary usage and the variability in requisite metadata content.

With high levels of projected or actual secondary data usage and as metadata content increases, the utility of data is improved by adding structure to the metadata. Expanding upon the example presented in Table 3, we define three levels of format or structure (low, medium, high) that also roughly correspond to the degree of formalization and the level of effort involved in adding that structure (Table 4). The lowest level of metadata structure may simply consist of a hard copy document or free format ASCII text in narrative form. This low level of metadata structure may be suitable for exchange with expert colleagues, but inadequate for electronic data set publication or other uses. A medium level of structure may encompass mixed format or partially parameterized information fields that could be easily searched (electronically) by a third party. For example, a medium level of structure may minimally support search and retrieval of Level I descriptors. High levels of structure may be used to store information in fixed format or highly parameterized fields such as those associated with the more sophisticated database management systems (DBMS). Some DBMS software supports development of executable and searchable metadata databases. Although this high level of structure is good practice for projects that require periodic data audits, it may be excessive for other objectives.

Increased metadata structure is beneficial for at least two reasons. First, the checklist character of structured metadata provides a memory-aid for the data originator about what is important to record about the data to enable his/her own re-use as well as facilitate utilization by others. Secondly, increased structure facilitates verification of results and development of searchable catalogs and database interfaces, making the data available to a larger potential population of users with a wider range of processing software. As an example, the NOAA Earth System Data Directory (Barton 1995) and the U. S. Geological Survey's Global Land Information System (Scholz and Smith 1995) utilize Directory Interchange Format (DIF) as a mechanism for ensuring that a minimum level of metadata are available during searches for data sets.

Although increasing metadata structure (i.e., format definition) reduces the burden on data re-users, it significantly increases the burden on the data originator. While one may argue that the burden should be on the data re-user to ferret out the relevant details this is frequently impossible. Thus, data re-use is frequently based on intelligent and well-intentioned guesses. For example, if the data originator is still alive it may be that he or she does not remember what quality assurance procedures or analytical algorithms were used since the relevant information was never documented, or programmers or knowledgeable technical personnel have since left the project. Ultimately, the burden falls on both the data originator and secondary users to apply good practices and minimize the propagation of error arising from unintentional misuse of the data. The Carbon Dioxide Information Analysis Center, for example, emphasizes the value-added component of data sets resulting from joint participation of

Table 3. Content of metadata (refer to categories in Table 2) associated with three levels of secondary data utilization

Metadata Descriptors (see Table 2)	Levels of Secondary Data Utilization and Associated Metadata Content		
Major Categories	Level I: Exchange with Expert Colleague	Level II: Searchable and Third Party Data Re-use	Level III: Publishable and Auditable
I. Data set descriptors	●	●	●
II. Research origin descriptors		●	●
III. Data set status and accessibility		●	●
IV. Data structural descriptors	●	●	●
V. Supplemental descriptors			●

Table 4. Degree of metadata format/structure sufficient for three levels of projected secondary data utilization.

Planned Use		LOW {Free format, ASCII, narrative or hard copy}	MEDIUM {Mixed format, partially parameterized}	HIGH {Fixed format, highly parameterized, executable, language dependent}
III	Publishable and Auditable	**Inadequate**	**Minimal**	**Good Practice**
II	Searchable and 3rd party Re-use	**Minimal**	**Good Practice**	**Excessive**
I	Exchange with Expert Colleague	**Good Practice**	**Excessive**	**Excessive**

Amount of Structure {Formalization, level of effort}

scientists and users in metadata preparation, rigorous QA/QC processing, peer-review of data and metadata, "beta testing" of data sets prior to general release, and incorporation of user feedback into its data packages (Boden 1995, National Research Council 1995a)

Conclusion

Basic and applied ecological research depend upon the availability of high quality data. If *a priori* consideration is paid to the development of high quality data sets and accompanying metadata, then individual scientists and organizations can focus valuable time and effort on performing appropriate analyses with the requisite high quality data. As metadata and metadata standards are developed and implemented, individual scientists and organizations can further benefit by being able to easily re-use data developed for other applications.

Further progress in development, adoption, and implementation of non-geospatial ecological metadata standards depend upon data and metadata being recognized as representing an integral component of the scientific process. Study repeatability; comparative ecological studies; attempts to scale up domain-specific studies to broader spatial, temporal, and thematic domains; ecological simulation modeling and model validation; and more applied ecological research (e.g., restoration ecology, ecological risk assessment, research into sustainable development, etc.) all depend upon the availability of archived data and, equally importantly, upon the ability to understand those data via the metadata. Data are more frequently being re-used by data originators and being utilized by other scientists that were often not involved in the data collection efforts. Thus, the scientific value of being able to reuse data and to utilize data for multiple objectives that may not have been foreseen by the data originator(s) may far exceed the perceived value associated with publications resulting from the original study.

All data should be accompanied by metadata. The completeness of the metadata governs the length of time and the extent to which data can be reused by the original investigator(s) and utilized by other scientists, resource managers, decision-makers, and other potential users. Just as the data and information contained in a manuscript support peer-review of the publication and the conclusions reached therein, metadata support peer-review of the data and facilitate secondary utilization. Ideally, the metadata should be physically linked as closely to the data as possible. For example, non-imagery data and associated metadata collected under the auspices of the United States Department of Energy's Atmospheric Radiation Program are integrated and stored in Network Common Data Format (netCDF) structure (Melton 1995).

If state and federal agencies, scientific societies, and academic institutions perceive the value of data sets collected by grantees or members of their organizations, then appropriate value should be placed on the publication of data and metadata, in addition to more traditional peer-reviewed publications (National Research Council 1995a). Perhaps data and accompanying metadata for "irreplaceable" or otherwise valuable ecological data sets could be published in an as yet to be developed electronic journal and then submitted to a data archive. Such data sets would then be citable in the scientific literature. Ultimately, however, successful incentives will rely upon organizations placing appropriate value on data/metadata publications during the scientific merit review process.

Agencies and scientific societies should promote metadata development and metadata standardization (National Research Council 1995a). For example, the geographical sciences community has developed spatial data transfer standards (National Institute of Standards and Technology, 1992) and metadata standards (Federal Geographic Data Committee, 1994) that are widely endorsed by Federal and state agencies, scientific societies, and academic institutions. The National Biological Service (American Institute of Biological Sciences 1995) and a sub-committee of the Federal Geographic Data Committee (M. Nyquist, personal communication) are developing additional extensions that would comprise supersets of the existing geospatial metadata content standards (Federal Geographic Data Committee, 1994) and would be appropriate for biological resource, cultural, and demographic metadata. When appropriate, existing metadata standards (e.g., spatial metadata standards) should be endorsed and promoted by the ecological sciences. Where metadata standards are incomplete or do not exist, attention should focus on developing, endorsing, and adopting appropriate standards.

Data and metadata should be independent of hardware and software to the fullest extent possible (Conley and Brunt 1991). Proprietary data storage formats inevitably change through time or are replaced by new formats. Thus, the life span (long-term utility) of data and metadata may be severely degraded when data/metadata conform to a proprietary standard as opposed to a more generic "industry-wide standard." Agencies and institutions may find it beneficial to collaborate in development and support of digital library services for data and metadata archival.

Funding agencies, scientific societies, and research institutions should recognize that there are costs, as well as benefits, associated with archiving data and developing and maintaining the requisite metadata. Thus, enhanced levels of funding to support these ancillary activities should be recognized as being necessary and appropriate. Similarly, funds would be required to resurrect valuable historic data and metadata. However, it should also be understood that historic data are frequently more readily retrieved (resurrected) than are the essential metadata, as demonstrated by the IBP example.

As ecologists address the complex issues associated with metadata standardization, long-term data and metadata archival, and secondary data utilization, a cautionary note from the geographical sciences may be in order. Specifically, Chrisman (1994) asserts that "all the standardized procedures in the world cannot ensure that the product actually satisfies the user's needs." He emphasizes the joint responsibilities of users and providers in relation to spatial data use and documentation, the need to incorporate spatial statistics more fully into GIS, research leading to a better understanding of error propagation in GIS and, importantly, the critical need to develop "procedures that can handle large differences in resolution, accuracy and other key properties."

Ecology, like geography, is by its very nature interdisciplinary. However, a review of interdisciplinary environmental research and assessments conducted by the Committee for a Pilot Study on Database Interfaces concluded that "the existing missions and attendant reward systems of research organizations act to inhibit the data sharing, mutual support, and interdisciplinary mindset needed for successful data interfacing" (National Research Council 1995a). The increasing reliance on long-term, broad-scale, and multi- and interdisciplinary data to address issues related to global change, biodiversity, sustainability, and other societal concerns highlights the need for retaining important ecological data sets in an accessible and understandable form. Increased attention to developing high quality data sets and their attendant metadata; understanding how uncertainty, error propagation, and research and statistical assumptions affect the "fitness" of data sets for intended and unintended uses; and promoting a sense of stewardship for ecological data will certainly enhance the interdisciplinary nature of ecology as a science.

Recommendations

In recognition of the fact that thorough documentation of the context, content, quality, and structure of data sets is necessary for long-term ecological data to be preserved in a usable format, the FLED Committee recommends that the ESA:

1. Educate its membership on the importance of adequately documented metadata and its necessity for maintaining the long-term utility of any ecological data set.

2. Develop minimal metadata standards as part of feasibility plans for an electronically accessible ecological data archive. The plan should include guidelines for submission, and mechanisms for reviewing and accepting, candidate long-term (and other) ecological data sets to an ESA-sponsored archive.

3. Examine and promote incentives for the expenditure of time and energy required to adequately document a data set. Possible incentives involve the development of an ESA registered (or endorsed) system of accession numbers of ecological data sets and establishing guidelines for citing such data sets in ESA publications (also grant proposals and vitae).

4. Data Sharing and Long-term Curation

The desire for long-term preservation, continuing use and access to ecological data raises issues regarding the sharing of data. Ecologists have relatively little experience with data sharing, other than between close associates. The increasing ease with which scientists can access each others' data in the age of electronic communication is both marvelous and problematic. It is a boon for the free and open exchange of information and provides motivation for the preservation and increased utilization of long-term data sets, but it also presents new challenges in terms of data management, methodology, and intellectual property rights.

Data sharing raises ethical and practical concerns because of the uncertainties that can arise in sharing data with a broader group of scientists and the work required to document data sets so they can be used by others. However, data sharing is a well-established tradition within a number of scientific disciplines with whom many ecologists interact. The fields of physics, dendrochronology, paleoecology, molecular biology, and systematics, for example, all rely heavily upon data exchange. Lessons learned from these groups on how to develop systems for data-sharing and exchange--and to demonstrate the value of a data exchange system--can provide valuable insights to ecologists how we might develop and use an ecological data archive and exchange system.

The Value of Data Sharing

Efforts to extend the spatial, temporal, and functional scope of ecological research require combining contemporary and historical data sets from multiple sources. For example, acquiring the temporal perspective and resolution necessary to discern significant declines in species diversity, population changes, fluctuations in age structure, or responses to anthropogenic factors often requires the use of data collected by previous investigators--and from sources outside ecology. For example, understanding the processes controlling changes in the spatial and temporal patterns of biological diversity involves the use of specimen-based survey data as well as the related voucher specimens found in systematics databases, remote-sensing imagery, and other mapping data.

Databases compiled from a variety of sources that have been reformatted, verified by quality assurance and quality control procedures (QA/QC), integrated with ancillary data, and well-documented have been, and continue to be, valuable to efforts to extend the spatial, temporal, and functional scope of ecological studies. Some examples of these are the Geoecology Database which compiles over 1000 variables by counties of the lower 48 states, including data from the United States Forest Service Forest Inventory (see Olson in Volume II, Directory 2), the Breeding Bird Survey maintained by the United States Fish and Wildlife Service, and the Natural Resources Conservation Service (formerly Soil Conservation Service). The paleoclimatology data exchange network funded by NOAA (described below) provides another example. Several other examples are listed in Directory 3 in Volume II of this report.

Concerns about Data Sharing

The primary concerns of individual scientists regarding data sharing focus on issues of data ownership, protection of intellectual property rights, acknowledgement, and attribution. Such concerns can--if unrecognized--undermine an atmosphere of free and open exchange within and among scientific disciplines. Other concerns about data sharing are largely logistical--maintenance of data integrity, adequate documentation for effective reuse of data, acquisition of funding, management of data transfer contracts, and so forth. These are important, but secondary, concerns for an effective data sharing network. Workable solutions to many of these problems that have been developed in other disciplines could serve as models for ecologists (see below).

The concerns of an individual about data ownership are analogous to a firm's proprietary interest in commercially valuable data. Thus, there are parallels--and relevant legal discussion--in principles established under copyright law. It is important to recognize, however, that these principles have been developed to handle problems related to commercial interests and are not necessarily consistent with the basic principles and spirit of science. The legal machinery associated with civil litigation in intellectual property rights, nevertheless, does provide a frame of reference (see below).

As yet the commercial issues surrounding data ownership have been important in only a few disciplines in the biological sciences, notably medicine and molecular biology. However, with increasing electronic access to--and demand for--ecological data, ecologists and their colleagues in related disciplines will soon be confronting these issues. Problems have already begun to emerge with the restrictive practices associated with the licensing of remote sensing data. There is also the possibility that fees may be charged for some types of ecologically important data. For example, meteorological data may be sold by European organizations rather than freely provided as it now is.

Data Ownership and Copyright. In the commercial world data, ownership is usually achieved by either creating or buying data (Gordon, et al. 1994). In contrast, in the scientific community the prevailing mode of achieving data ownership is through the collection of data, although contractual acquisition of data also occurs and is critical to ecology. A good example of the latter is the case of remote sensing data, which are usually licensed rather than sold. Long-term ecological data sets are frequently collected and maintained by a succession of curators. Consequently, ownership of long-term or historical ecological data is sometimes achieved a third way: inheritance (or adoption).

A copyright provides legal standing to prosecute the unauthorized copying of a protected work and could potentially be applied to data. Copyright law determines the primacy of claims of authorship based on the registration of a creative work with the Library of Congress. At present, the Library is developing an Electronic Copyright Management System with three components: (1) a registration and recordation system; (2) a digital library system; and (3) a rights management system. While this automation effort will improve the ability to examine existing copyrights, it will do little to affect the basic problem with copyright: it is too expensive for most individuals to enforce.

A central issue regarding data ownership and protection against unauthorized copying is that databases generally are collections of facts, and **facts cannot be copyrighted**. Copyright protection is, on the other hand, afforded to the arrangement of facts as long as the collection of data are selected and coordinated in such a way as to represent an original work of authorship (Gordon, et al. 1994). Most databases can be readily modified into a form that is nearly identical to an original and yet not be the same thing in terms of strict copyright protection. Clearly this poses problems in terms of establishing the ownership of the data. Parenthetically, it also becomes a configuration management nightmare in maintaining synchronization across research activities if researchers are using similar--but different--copies of the same data set.

<u>Data Ownership for Individual Researchers</u>. For the most part, issues of law are not directly relevant to concerns about individual ownership of scientific data. This is because much scientific data is generated by public monies and is therefore subject to disclosure under the FOIA (Freedom of Information Act). Although grants from funding agencies support scientific research, these agencies do not own the data. A 1980 Supreme Court decision found that "...because an investigator is neither carrying out a protocol designed by a federal funding agency nor providing data directly for use by the agency, work supported by a federal grant is not the possession of the funding agency" (Bailar, et al. 1990, p. 13). Thus, from a legal perspective, neither federal funding agencies nor individual researchers who fund their data collection through federal grants own the data collected. Although public access to these data is required by the FOIA--and such access is wholly in accord with the spirit of science--it is not clear who has the responsibility to maintain and support access to such data.

Scientists generally consider themselves as having the entitlement, authority, right, freedom, and power to use the data they collect using a federally funded grant--but they do not own the data. The perception of individual entitlement to data is widely respected and acknowledged by funding agencies and they expect disclosure of the data only after a "reasonable time'" (e.g., NSF Guidelines quoted in Section 2). However there are few guidelines for what constitutes "fair use and access" to data after this time has passed. Since scientists are expected to be reasonable stewards of the data they collect-- and to share it in a timely fashion--this can create conflicts of interest if researchers are expected to maintain (and share) data that they do not own in a legal sense.

Data Sharing Ethics. There are a number of ethical conflicts and concerns involved in sharing scientific data. While there are scientific and legal interests in promoting free and open exchange of data, the system of research credit works against openness. There are a number of ways that scientists can get the data they need--but who receives credit when data from others is used? And who should have priority in publishing data collected by others?

As an example of this tension, one of the key questions addressed by journal editors is whether a secondary author (i.e., one not involved in the original research) should be allowed to publish before the primary authors do (Bailar et al. 1990). However this is resolved, scientists generally care more about getting their work published and what journal editors will allow, than in assuring free access to their data. This has led several disciplines to develop incentives for data sharing such as requiring submission of data to a data bank at the time a journal article is published (e.g., DNA data banks; described below).

Although a complexity of forces are operating on data acquisition and intellectual property rights in science, the foregoing discussion primarily describes how things have been. With rapid developments in the ability to exchange data world-wide, the potential exists for data to be shared among groups (or individuals) with little knowledge of each other and potentially no direct contact with the data originator. For ecologists taking advantage of this new technology, the future promises to be a period of increased data sharing and integration that will provide greater spatial and temporal resolution of the dynamics of natural systems. As data sharing becomes more impersonal, there will be increasing ethical responsibility on all researchers to properly attribute authorship of data as well as to effectively document data for appropriate use by others. The Ecological Society--and other scientific societies--has a responsibility to review their Code of Ethics to assure that it deals explicitly with evolving issues related to data sharing in the electronic age (e.g., Garte 1995, Mitchum 1995, Sieber and Trumbo 1995).

Experiences from Groups that Share Data

Many of the concerns ecologists have about establishing a data exchange networks are not unique to ecological data. In the belief that there are important lessons to be learned from other disciplines that have developed data exchange networks, the FLED committee looked into various models for data exchange. Our intent was to learn under what conditions initiatives to create data banks succeeded--or failed--and what tactical approaches had been taken to alleviate concerns about data sharing. In the following summaries, we highlight findings from discussions with individuals in a number of disciplines with which ecologists have in the past and are increasingly interacting that have developed data sharing and archival systems.

Summary of Groups That Share Data

Paleoclimate Data: ITRDB and NAPD/EPD. The International Tree Ring Data Bank (ITRDB) was established in 1974, by individuals who were motivated by the need to share data to address questions about past climate. Climate operates over large geographic scales, so tree ring data must be shared nationally and internationally. The ITRDB initially depended on the energy and grants provided by a single committed individual (Harold Fritz) and the support of successive Directors of the Laboratory of Tree Ring Research (University of Arizona). After over a decade of operation, NOAA's Paleoclimate program took over routine operation and maintenance of the ITRDB and provided funds for an advisory committee drawn from the worldwide contributing community.

All data deposited in the ITRDB are available at cost by anonymous FTP. Searches are aided by a search engine on the World Wide Web (http://www.ngdc.noaa.gov/paleo/treering.html). Researchers are expected to acknowledge the source of the data (contributor to ITRDB) in publications based on data from the ITRDB. There have been only a few cases when the original data contributor has not been acknowledged. This creates a disincentive for submission of data sets to the database and the ITRDB is considering establishing a mechanism for peer-reviewed, citable publication of data sets such as that being developed by the American Geophysical Union.

The North American Pollen Database (NAPD) and the European Pollen Database (EPD) were established in the early 1990s in response to a need for data for comparative analyses of climate and vegetation dynamics over large spatial and temporal scales (Anonymous, 1993). Another motivating factor was the need for access to the raw data themselves, rather than the summaries that were generally published in journals. Prior to the establishment of the NAPD and the EPD, several individual researchers had compiled substantial independent databases: particularly, Thomas Webb III at Brown University (COHMAP members 1988) and Brian Huntley and John Birks in Europe (Huntley and Birks 1983). While contributors to these databases could use other data sets, the effort to maintain these databases grew beyond what individuals could support. Today, the NAPD is funded by NOAA and all data are freely available by anonymous FTP (ftp.ngdc.noaa.gov:/paleo/pollen) and on the World Wide Web (http://www.ngdc.noaa.gov/paleo/ pollen.html). Close collaboration between the NAPD and the EPD has enabled the development of compatible file structures across the two databases. Modifications to either database are made only after joint consultation. Attitudes in regard to these shared databases has evolved from skepticism and reluctance on the part of a few, to contributors feeling honored to see their data being used by others to further science. This has led to an open atmosphere of generosity, and even a feeling of obligation to contribute to the archive.

<u>Molecular Data: Human Genome Project, ACeDB, and Others</u>. Archives of molecular data began in the 1970s for the purposes of comparative work. Publication of extremely long sequences became cumbersome, leading to the development of DNA data banks. Again, these early archives were the result of efforts by a few committed individuals; particularly, Margaret Dayhoff (proteins) and Walter Goad (DNA sequences). In 1982, Genbank, a centralized nucleotide sequence archive, was started as a multi-agency project (NIH, DOE, and others). The Genbank program initially gathered and compiled information from journal publications. Eventually, they asked journal editors to require authors to file diskettes of sequence data directly with Genbank when papers were published. Genbank provided free software to expedite contributions of data. Similar projects were developed in Europe (EMBL)and Japan (DDBJ). Genbank is now operated by the National Center for Biotechnology Information (NCBI) at the National Library of Medicine. Genbank's descendant, the Genome Sequence Data Base (GSDB), is operated by NCGR (National Center for Genome Resources) in Santa Fe, New Mexico. The information contained it these databases is available--currently without charge--to the research community by both GSDB and Genbank, which exchange information nightly.

Another example of a molecular database is the nematode AceDB (A *Caenorhabditis elegans* Data Base) project. This project was spearheaded by a few dedicated investigators (Thierry-Meig at Montpellier and Durbin at Cambridge), who also provided free software to facilitate use and distribution of shared data. Initially the project was tightly coupled with the *C. elegans* Genome Project. Two other successful projects, the *Arabidopsis* and *Drosophila* databases, are based on the ACeDB example.

Not all attempts to develop a data exchange network in molecular genetics have been successful. For example, attempts to establish a database for *E. coli* sequence data have failed, in part, because of competing interests of users and contributors. Those who were responsible for making genetic maps, defining gene symbols, etc. (stock centers) were opposed to the establishment of a data bank because they feared it could lead to pieces of their not-fully-synthesized product being made available.

Natural History Collections. Systematists and curators of natural history collections have had a long history of data sharing in order to better describe taxa. Rather than a network dependent upon the energy and personal resources of a single individual, regional natural history collections have been assembled in many locations over decades or centuries. Traditionally systematists have been both users of and contributors to these collections. The same individual frequently both accesses specimens (through loans, visits, and exchanges) and contributes specimens to the collection. Traditional reciprocal relationships between scholars and institutions have facilitated the open publication and exchange of scientific findings.

Today, in addition to physical specimens, natural history collections include electronic databases derived from voucher specimens. These databases are available to, and in demand by, systematists and non-systematists for the documentation of patterns of biodiversity. As systematists are increasingly share data with individuals who are not contributing to the development and maintenance of collections, the system of reciprocity has eroded. As a result, there are concerns in the systematics community about the lack of support for the maintenance of biological collections. At the extreme, a systematics institution could find it is competing for funds with a group whose proposal was in part based on access to that institution's specimen-derived database.

Within the systematics community, the issue of how to deal with access to and use of electronic databases is not fully resolved (Hathway and Hoagland 1993, L. Kristalka, pers. comm.). Many believe that as biocollection data enters the electronic age, it should be freely available for all. Yet, once information becomes available on the Internet, it becomes almost impossible to prevent copying. This has motivated discussions in the systematics community that the economic value of information may need to be redefined, away from "content" (protected by copyright) and towards the providing of services (Dyson 1995).

The ASC (Association of Systematics Collections) feels strongly that individuals or organizations should be able to come to a collection and use the resource, even to make their own data records. At the same time, however, an institution may choose to limit secondary transfer of information, restrict access to certain data fields, require the presence of a curator or collections manager when the collections are being consulted, or charge reasonable user fees. The sale of data obtained under collegial reciprocity and the unauthorized alteration of data may be explicitly prohibited by an institution. However, maintaining (and updating) a specimen-derived databases, requires curatorial management of specimens. To recover some of the costs of maintenance and stewardship, data products from specimen-based databases may be sold (e.g., as CD-ROMs, on-line client-vendor gateway arrangements, or contracted subscriptions to access data and its interpretation) by some institutions.

Most of the information in natural history collections is not yet available electronically, although this situation is rapidly changing (Miller 1993). Examples of well-developed, searchable electronic databases on the Internet are: ERIN (see Volume II, Directory 3), MUSE (muse.bio.cornell.edu), the Missouri Botanical Gardens (http://www.mobot.org/MOBOT/database.html), and the herbarium at the University of Texas, Austin (http://www.utexas.edu/depts/prc/web/main/html). The ASC is promoting the establishment of a large-scale network of biological collections and will provide some coordination of this by creating a database of collection resources (funded August 1995 by the NBS). This database will be on the Internet with hotlinks to ASC member institutions.

Biological Field Stations. The Organization of Biological Field Stations (OBFS) and the Southern Association of Marine Laboratories (SAML) have a long-standing interest in preserving and managing ecological data. Workshops were conducted workshops in 1982 and 1990 to encourage and foster development of data management at field stations (Gorentz 1992). Field stations maintain irreplaceable long-term ecological data sets that could be available for secondary use (see Volume II, Directory 2). The OBFS supports--but limited funding for--the development of means to preserve long-term ecological data sets collected at field stations. The OBFS is interested in developing a data exchange network, and hopes that their efforts could be linked with those of other ecologists and field-oriented biologists.

Lessons Learned about Data Sharing

These examples of data sharing attempts by other biological disciplines reveal several common themes. Data banks as a mechanism for data exchange tended to be **successful** when their development was accompanied by:

a) A **motivation** to meet scientific and question-driven needs of the field.
b) **Leadership** from a few (1-2) individuals, with common intellectual goals and strong influence, who initially dedicated a tremendous amount of time, energy, and funding to establishing the database.
c) Development of **free and easy-to-use software** for data transfer.
d) **Community acknowledgment** of the importance of data sharing activities and recognition for contributions.

e) **Support from key journals** that encouraged (or required) that data be deposited in the data bank prior to publication.
f) **Sustained, external funding** for data management and maintenance of the project, often forthcoming after a demonstration pilot project.
g) Development of **mechanisms for quality control,** either through a gatekeeper committee or through *ad hoc* review of contributed data sets (often provided by publications based on the data set).

The successful data banks we investigated also shared some **concerns** and provided important lessons regarding defining protocols that address logistical concerns for development of data exchange networks. These concerns and lessons include:

a) **Commercialization of data** potentially poses a threat to free and open exchange, especially in geophysics and systematics.
b) The **need for recognition and incentives for data contribution** and data activities including: peer-reviewed data sets as publications; development of accession numbers and means to cite contributed data sets; journal requirements that manuscript publication be accompanied by submission of an appropriately documented data set to a data bank.
c) **Maintaining financial support** for a data exchange network and developing means to recover costs, including user fees and professional society dues.
d) **Ethical misconduct** - though this has proved to be less of a concern than the skeptics of each group initially feared. There have been occasional problems with data users failing to cite the author or accession number of public access data; no incidences of outright fraud were brought to our attention.
e) **Data transfer agreements** can take on a variety of forms. These varied from data being available by permission of investigators only to data transfer agreements that lay out specific such as: terms and conditions of data use (giving credit, third party use, data security, freedom of information law obligations, etc.); limitations and disclaimers of the data set; definitions of categories of users; fees or in-kind charges; procedures for amending the agreement and dispute resolution.

Data Sharing Among Ecologists

Historically data sharing has been primarily an *ad hoc* activity within the ecological community. For data sharing to become a more reliable and beneficial activity in ecology, mechanisms must be established by which data can be contributed, disseminated, verified, and attributed. Although attempts have been made to establish ecological data exchange systems within sub-disciplines of ecology (see below), these past efforts generally failed because of the lack of sufficient endorsement and/or participation by ecologists. Lessons can be learned from past--and current--efforts to establish effective data sharing networks for the ecological community. We have summarized several of these below.

The Past: International Biosphere Program (IBP) and the ACCESS Project

The U.S. IBP program was established in 1968 and focused on whole ecosystem research. The program thus focused on measures of primary productivity, trophic structure, energy flow pathways (food chains), limiting factors, biogeochemical cycling, species diversity, and other attributes which interact to regulate the structure and function of communities (Blair 1977). Their was a strong emphasis on quantification and comparison of these processes in different ecosystems, and that led to the development of methods for data standardization and comparability (Loucks 1986).

While not designed as long-term studies, the IBP data are historically valuable and important because of the scope and geographic scale over which they were collected. These data can provide a baseline to assess temporal changes in a number of important environmental variables (e.g., photosynthetic rates with increased acidification, CO_2 enrichment, and global warming). Several of the IBP sites had a well-established history of ecological research, and in some cases IBP-initiated studies have been maintained as part of the LTER network or other long-term projects (see Boxes 2 and 3).

A major objective of the U.S. IBP was development of a data bank and mechanism for data exchange, forming the foundation of ecological data management. Each biome was expected to establish similar procedures for compiling and storing numerical data and to establish provisions for documenting the data (e.g., investigator's name, location of research, parameters measured, key words, restrictions on dissemination of the data, and a brief description of the data set and experimental method).

The utility and value of the data banks created by the biome programs were, in the end, controversial (Loucks and McElreath 1975, National Research Council Committee to Evaluate the IBP 1975, Mitchell, et al 1976, O'Neill 1976). An NRC report fairly early in the program stated: "The biome programs have similar procedures for compiling numerical material and storing it in data banks. Use of a common pattern ensures that data stored by one program can be read and used by another." (NRC 1974, pp 21-23). However, other reports expressed contrasting opinions: "The objective of maintaining uniform data banks that could easily transfer data between biomes was, in fact, a failure.... IBP data are available from individual investigators and sometimes from data banks, but no program is designed to respond to any significant number of outside requests for data." (Hinckley and Haug 1979, p. 37). "Problems with incomplete data sets, a lack of uniform data format, and a poor central catalog of the available data all acted to reduce accessibility and utility." "It appears that technical problems of handling ecological data have to be addressed and solved before data banks can serve a useful function." (Battelle Columbus Laboratories 1975).

Data management varied among the IBP sites and across sub-disciplines within the program. The quality of data management at the time the work was done determines how useful and accessible it is today. For example, the Tundra Biome housed only a third of the biome data collected at these sites and purportedly did not specify standards for data format or sampling procedures (Hinckley and Haug 1979). The Coniferous Forest Biome site, on the other hand, pioneered efforts at data documentation and developed forms that have made it possible for the Forest Science Data Bank to preserve several IBP data sets (see Box 2). Some IBP data, such as the Lake Wingra Basin database (Prentki et al. 1977), primarily were available as internal reports, although these also existed on tape. Those IBP sites that later became part of the LTER or other agency programs, still have many IBP data sets in their archives. The status of these data sets varies due to the quality of the metadata and changes in technology that have hampered attempts to retrieve these data (see Section 3).

> **Box 2. The Forest Science Data Bank.** The Forest Science Data Bank (FSDB) evolved directly from the efforts of the IBP Coniferous Forest Biome data managers. Currently, the FSDB is supported in part by LTER and houses the H. J. Andrews LTER databases. Much of the IBP data still exist today in the FSDB, and all of the long-term data sets from IBP traditionally handled by a data manager (e.g., stream flow, climate, and water quality) are intact. Many of the short-term experimental studies handled by individual researchers have incomplete documentation, require substantial efforts to reconstruct, and have not been fully restored. Some of the long-term data sets have required extensive and costly efforts to perform the QA/QC checking and assure completeness of the metadata. FSDB's experience in the challenge of handling diverse environmental and ecological data was recently reviewed by the National Research Council (1995).

It is important to put this summary of IBP data banks in context. In the 1970s data management methodology was sparse and the IBP data managers pioneered efforts to develop tools for managing ecological data. Hardware was limited to remote, mainframe computers, and software tools were essentially limited to custom-written Fortran programs and line editors. Data sharing was primarily internal. Evolution of methodology and technology in the information management field has improved data management and made ecological data more available from these sites. For example, at the Forest Science Data Bank after a Local Area Networks became available in the 1980s, the old tape library was ported over to a Novell server. This made data sets available and accessible on-line to investigators. In addition, database tools were available for extensive QA/QC checking of these data, and thorough review of data and metadata were possible.

Box 3. Oak Ridge National Laboratory. The Oak Ridge National Laboratory (ORNL) was the data repository for the IBP Eastern Deciduous Biome and today its archives contain many IBP legacy data sets. For example, the Walker Branch Watershed was initially set up by the Atomic Energy Commission prior to IBP (1967) as a long-term study with the goal of quantifying land-water interactions in an undisturbed, forested ecosystem by monitoring forest growth, hydrology, and chemical/nutrient cycling. Between 1969 and 1978, it was largely to partially funded by IBP, after which the new DOE resumed funding most of its projects. Although multiple investigators were involved during IBP days and collaboration occurred between them, nothing was public domain. Some soil and plant tissue samples were archived, however. In general, these types of data receive little institutional support for backing up old computer tapes or data reconstruction. While other projects at Walker Branch continue to be funded, this 28 year monitoring venture will soon be terminated.

Although some investigators have made extensive use of various IBP-derived data, in general the raw data and the published results (e.g., US/IBP Synthesis Series 1977-1981) remain underutilized. Much of the data are probably available only in hard copy (including unpublished internal data reports), not digitally, due to changes in computer technology and loss of the old tapes and computer cards. The internal nature of IBP hard copy reports also means that current library searchable databases only index a portion of the sources of these data.

Concern about "... communicating nationally what is available, where it is, and how to inquire..." and "The means for effectively responding to requests forwarded to centers where data are available..." had not escaped attention from scientists involved in the IBP (Loucks and McElreath 1975). This motivated several individuals who had been involved in the IBP to develop a guide to the data and resulted in the "User's Guide to IBP Biome Information" (Hinckley and Haug 1977). In addition to providing a directory to researchers, sites, and published and unpublished bibliographic sources, this User's Guide also demonstrated the need for a more general approach to public access of environmental and ecological data bases. In 1978, DOE funded the ACCESS project to examine the feasibility of establishing regional or national access to diverse environmental databases available through federal, state, and local governments. The ACCESS report provides a model for a national project of data access and addresses many of the technical problems of data management and access to diverse users, some of which are as relevant today as they were then (Armentano and Loucks 1979). The plan would have required much effort on the part of participating agencies, and was never followed up. However, the cost of such a cooperative effort might be received in a different light today--with fewer federal dollars available to support new ecological and environmental research, the cost:benefit ratio of a cooperative effort may have changed.

The Present: LTER, LTREB, LMER, and ESA

The Long-Term Ecological Research (LTER) Network. The LTER Network is a network of 18 sites coordinated by a Network Office, currently located at the University of Washington (Risser and Lubchenco 1993). This network was established in 1981 by the NSF and today consists of approximately 700+ scientists, students, postdocs, technicians and staff who are located at various institutions associated with the 18 LTER sites. The LTER sites are located in the continental United States (15 sites), Puerto Rico (1), and Antarctica (2). For detailed descriptions of the LTER program, see Franklin et al. (1990), Van Cleve and Martin (1991), other publications of the Network Office, or contact the Network Office (LTER, University of Washington, College of Forest Resources, AR-10, Seattle WA 98195). The LTER Network also maintains a HomePage on the World Wide Web.

The LTER data management and exchange system is considered to be a model system and is being used to train and develop data exchange networks in other countries. Data management, in the broadest sense, has always been a high priority for the LTER program and for the individual sites--many of which had established long-term ecological research programs prior to the initiation of the LTER. For example, the Coweeta Hydrologic Lab and Research Site (Otto NC), has been a U.S. Forest Service site since 1933 and has extensive long-term (60+ years) data sets. Many other LTER sites have similar long-term histories. Data exchange and data management are carried out both at individual sites and within the network. At each site, these activities are coordinated by the site Data Manager; at the Network level, activities are coordinated by the Network Data Manager, Rudolf Nottrott (rnott@lternet.edu).

The LTER data managers meet annually and issue reports; they host workshops and sponsor symposia, to which other individuals and organizations are invited (Michener 1986, Michener et al. 1994). LTER data management activities cover a broad range of topics, including Information Systems, Metadata Standards, Spatial Data, On-line Access, Software tools, RTC (Recommended Technological Capabilities), and development of metadata catalogs (e.g. the Core Data Set Catalog, LTER Pub. No. 5, 1990).

Long-Term Research in Environmental Biology (LTREB). The LTREB program also is sponsored by NSF and supports long-term research conducted by individuals not associated with LTER sites. Although effective data management is an important criteria for evaluating LTREB proposals, these sites are not yet integrated into the data management and data exchange network of LTER. Data management--and protocols for data sharing--is the responsibility of each LTREB principal investigator and varies considerably among sites. There are currently no means for promoting data sharing or exchange among LTREB sites or other investigators.

Land and Margin Ecosystem Research (LMER). The LMER Program is also supported by the NSF and is designed to support integrated research into the structure and function of ecosystems at the margins of continents and of large river systems which drain from continents. The current four sites are located on the west coast (Columbia River), the east coast (Chesapeake Bay and Plum Island Sound, MA) and a river system (Georgia Rivers). Although the LMER is not a long-term research program in the same context as the LTER program, much of the research conducted at these sites is long-term in nature. The sites have a small coordination office, located at the Marine Biological Lab, Woods Hole, Massachusetts. Data management is an essential component of all four sites, but the activities are deliberately decentralized and data management is primarily a local, on-site activity. However, the LMER sites are currently implementing a network-wide information system for metadata. Information on the LMER sites and the data being collected at each sites is available on the LMER HomePage (see Volume II, Directory 3). A metadata catalog is currently being developed.

The Ecological Society of America. The ESA has an official archive (currently located at the University of Georgia) that contains records and correspondence of the Society, primarily in paper format. The ESA does not have an archive of ecological data, although the Publications Office has maintained a *de facto* data archive since 1981 when it established the EPS (ESA Supplementary Publication Service). The motivation for the EPS was conservation of journal space. Authors of papers accepted in the Society journals (*Ecology, Ecological Monographs, Ecological Applications*) were encouraged to submit extensive tabular material or software to EPS rather than publish it. The availability of this supplementary material was indicated in the paper.

Currently the EPS has 65 documents, 58% (38) of which are some sort of data tabulation (including species lists, pollen counts, data sources). Software (code and description) makes up 37% (24) of the remaining contributions. Initially, authors were required only to submit a paper copy of the document, which was then transferred and stored on microfiche. Now authors are required to submit the document in paper and diskette form. Requests for these documents are handled by the Publications Office, which does not track who requests information or inform the author of these requests. Lee Miller (Managing Editor until 1995) has indicated that there is a need to find a more permanent home for the EPS. Having this information available electronically would facilitate distribution and reduce some burden on the Publications Office.

The Future: Long-term Curation and Maintenance of Ecological Data

The task of devising a single system for preserving and enhancing the accessibility of ecological data is made challenging by the fact that ecological data sets include disparate types of data. Ecological data include measures of biological, chemical and physical variables acquired through direct measurements of earth, water, air, celestial objects and biota. Ecological data sets include information from experimental studies, *in situ* measures obtained from field surveys and monitoring programs, and remote sensing imagery. Remote sensing data is a special case because there are already substantial commercial and non-commercial distribution centers for these data and an industry and scientific community developed to support them.

Most ecological data is not commercially distributed and are collected by individual researchers at great expense and labor. They are expensive to process and to interpret due to the highly specialized nature of the measurements and the level of expertise required to make them correctly. In contrast to tree ring data or pollen data, for example, many types of ecological data are highly idiosyncratic to the research program that develops them. The measurements are typically sparse; often done infrequently over small spatial and temporal scales. While these characteristics make ecological data unique and quite valuable, they also make them a challenge to treat uniformly.

These challenges, however, are not insurmountable and should not preclude the establishment of a national archive for ecological data. The FLED committee supports the conclusion of the NRC report that "... all observational data that are nonredundant, useful, and documented well enough for most primary uses should be maintained." (NRC 1995, p. 40). There is a growing demand for ecological data--long-term and otherwise-- for synthetic and analytical research on basic and applied questions. Thus there is a need for a national repository for ecological data and the development of efficient means for accessing and contributing these data. The following discussion presents the perspective of the FLED committee on long-term curation and maintenance of ecological data sets, followed by our recommendations.

Curation vs. Archiving. When considering strategies for the establishment of long-term data collections it is important to recognize that this involves tow separate functions: curator and archivists. The curator has technical responsibility for, and expertise in, the content of a data collection (e.g., scientific knowledge of organismal characteristics), while the archivist has responsibility for and expertise in physical and logical organization, access methods, and integrity of the collection. These two functions may in fact be performed by one individual but they are, in principle, separable.

Many disciplines contribute to ecological data collections and the diversity of data is amplified by unique data acquisition methods within disciplines. Therefore, curators of ecological data must be expert in the data acquisition methods of diverse disciplines to understand the nature and limitations of the measurements made and how to interpret them. In addition, ecological data are not usually voluminous which means many data can be housed in a small number of computer systems maintained by archivists. As a result, there is potentially a need for many curators and few archivists, or else every curator must also be an archivist.

Centralization versus De-centralization of Data Storage. Centralization and de-centralization are relative terms. In the extreme, de-centralization means that every individual data owner (i.e., science researcher) is also the data curator and archivist. At present, the proliferation of computing resources in the hands of researchers argues strongly for the decentralization of data storage.

However, this is an impractical solution because it requires every researcher to have sufficient computing and staff resources with expertise in data management (the archivist function) as well as the science (the curatorial function). Given the current state of data management practice and methods, particularly in ecology, this approach would require extensive training of a large population of scientists in procedures for data exchange that are not yet developed for general use, nor widely agreed upon. On the other hand, centralization of data in a single location has failed due to the non-responsiveness of the central system to the diverse needs of the user community and the intrinsic inflexibility of a single system without competition and incentive for change and responsiveness.

The challenge is to achieve the appropriate balance between centralization and decentralization. Data must be brought together somewhere to enable integration in space and time and to facilitate the detection of resultant anomalies such as scale shifts due to different instruments or calibration methods. There also must be one or more catalogs or directories to ecological data where researchers can search for available data relevant to their needs.

A good example of an approach that achieves a successful balance between decentralization and centralization is found in the manner in which computer software and data are shared around the Internet community. While each piece of software or data is owned and curated by a given person or group, it is copied (i.e., mirrored) at more than one location to facilitate its accessibility and dissemination. This not only facilitates the activities of the end-user, but creates a community-of-interest which assists in finding problems with software, and provides the "creative tension" that encourages improvement of software.

For the long-term curation of ecological data, it might be possible to use an extant data center to perform the functions of mirroring and facilitating accessibility and dissemination. This would enable many users in the research community to access a data collection without draining the resources of an individual researcher's computer and communications system nor of the research staff. This is also beneficial in that the requirements for data integration require computer system resources and expertise rarely held by individual researchers.

The Costs of Curation. The establishment of a pilot project for curating and archiving ecological data as a means of exhibiting the utility and structure of a permanent ecological data archive and exchange network is not necessarily expensive. Existing resources, such as the National Science Foundation's Supercomputer Centers and similar centers found in some states, could be used as stepping stones to the development and proliferation of sound methods and facilities for ecological data storage and distribution. The NSF Database Activities (DBA) Program in the Division of Instrumentation and Resources at the NSF is specifically designed to provide funding for the establishment (but not the maintenance) of such activities. If an NSF Supercomputer Center plays a role in developing a centralized data archive, this role could become increasingly decentralized through technology-transfer once the methodologies are well established by the scientific community.

It is possible that commercial interests will eventually play a role in off-setting costs of maintaining an ecological data archive. Environmental data are already commercially valuable, primarily in the consulting and regulatory segments of commerce and government. We anticipate that the commercial value of existing long-term (and other) ecological data will be enhanced through data integration and this will provide further business opportunities.

Data Quality and Verification. Setting up an ecological data archive will require data quality and verification procedures at every step of the data management process. Data management includes at least: the acquisition of data; quality control and quality assurance (QA/QC) at two levels (source and integration); integration of data across disciplines to enable multivariate queries; report generation for quantitative and administrative needs; and, increasingly, the development of an information server for technical staff and public access to data resources via the World-Wide-Web. Below we describe options for these functions that could be used in establishing an ecological data archive.

The **data acquisition** process usually provides for the preparation of machine-readable data provided by data owners for integration into both a common file system (which acts as a master library and an archive) and a database management system (for interactive query and retrieval). Contextual, ancillary data maintained in the data system may include--but not be limited to--remote sensing, weather, and related data already in machine readable form.

An ecological data archive would require specified **Quality Assurance and Quality Control** (QA/QC) to assure that the data were accurate. Level i QA/QC uses statistical techniques to analyze data files to determine accuracy, consistency, and relevancy of data samples. It includes checking for outliers and valid data values within a file. This level of QA/QC is expected to be redundant with the QA/QC performed by the data author who is the originator of the data and who has ultimate responsibility for the initial and continuing integrity of the source data. The data curator would perform Level 1 QA/QC as a double-check on that original QA/QC in the spirit of best scientific practice and reports any anomalies to the data author for corrective action.

A separate and explicit step in the Level 1 QA/QC process must be the verification and authentification of a file. **Verification** is the determination that any given data file contains exactly what it is purported to contain by the data owner. Not only does this mean that any given copy of a file is identical to its source but that the content of the data file is what the owner states it to be. If, for example, the language is claimed to be English or the units are claimed to be in metric, this should be verified. These are not things which can be automated to any significant extent without a high degree of standardization during data acquisition.

Sometime in the future there may be standardization of methods and measures for many common ecological variables, but this does not exist at present and may not even be a justifiable given the individual nature of much ecological research. However, it would be reasonable to request that individual researchers utilize file verification methods built into most operating systems as invocable commands for comparing two files. These are generally reliable and easy to use for two files at a time. Other methods using cyclic redundancy checks (CRCs) and related methods commonly referred to as checksums can also be used as needed in situations requiring the processing of large numbers of files efficiently.

Authentication of a file is different from file verification. Authentication procedures address the question of whether or not a data owner would agree that a given file was created by them either directly or with their authority. This is the same problem faced by the financial industry in electronic funds transfer operations and consequently a great deal of work exists from which to draw upon. The most obviously relevant applications are digital signatures and public-key encryption methods (National Institute of Standards and Technology 1994). These techniques provide a means of attaching to a data file a unique digital number which can be compared to an independently published number provided by the data owner to ensure that the file is authentic.

Licensing. The specifics of data transfer contracts will need to be carefully considered in establishing an ecological data exchange network. Several options are available and have been used by other groups that share data. One option is to license data to users. A license is a contract or agreement between a copyright holder and an authorized copy holder to use the data in a prescribed manner. Licensing data is a good idea for several reasons. First, it is a relatively simple method of determining who has an authorized copy of the data. Second, it ensures that the licensee has an accurate copy of the data. Third, it formalizes the obligations of both parties in explicit terms. Specific data transfer contracts such as licensing, will need to be explored in developing the system for long-term curation of ecological data.

Recommendations

The problems of ecological data exchange that the FLED committee is addressing are not new. They were referred to frequently during the IBP synthesis, both at the outset and afterwards. Other environmental agencies have been interested in data exchange, and from time to time readdress the issue. The NSF has repeatedly demonstrated support for synthesis and exchange of ecological data--it supported synthesis at many of the biome sites after the IBP ended, it supports the LTER network, and has recently funded (in collaboration with the state of California and UC-Santa Barbara) the establishment of the National Center for Ecological Analysis and Synthesis (NCEAS).
Today we have better tools, including technology with a longer shelf life (e.g., CD ROMs), and greater electronic distribution capabilities (e.g., the Internet) that can support the archiving and exchange of ecological data. An organizational

structure is required to identify and articulate the goals and priorities of an ecological data archive and to promote its establishment. To promote the long-term curation of ecological data sets and their use, we recommend that the ESA:

1. Support the establishment of a pilot or demonstration project for curating and archiving ecological data as a means of exhibiting the utility and structure of a permanent ecological data archive and exchange network. The Database Activities (DBA) Program in the Division of Instrumentation and Resources at the NSF is specifically designed to provide funding for the establishment, but not the maintenance, of such activities.

2. Encourage the Ethics Committee to review the ESA Code of Ethics for explicit concerns regarding data sharing and the protection of intellectual property rights in the electronic age.

3. Promote interactions and collaboration between ecologists and scientists of other disciplines for the development of tools for successful data exchange among and within these disciplines by:

- co-sponsoring workshops and symposia that highlight the intellectual links between other disciplines and ecology and develop mechanisms for more effective data exchange among these disciplines;

- encouraging the ESA editorial board to work with authors to assure that papers published in ESA journals cite relevant databases from data exchange networks.

- supporting efforts by other disciplines (particularly systematics) to increase funding for the maintenance of collections and the development of associated databases to assure that these valuable sources of 'incipient ecological information' are maintained and available to ecologists.

Sources Cited

ASC (Association of Systematics Collections). 1993. An Information Model for Biological Collections. Draft report of the Biological Collections Data Standards Workshop, Cornell University, Ithaca, N.Y., Aug 18-24, 1992. Association of Systematics Collections, University of Kansas, Lawrence, Kansas.

Anonymous. 1995. Pollen Database Manual. Illinois State Museum. Springfield Illinois. 75pp.

Anonymous. 1993. National Center for Synthesis in Ecology: A Design Study (The Herndon Report). 32pp.

Association of Ecological Research Centers. 1989. A National Center for Integrating Ecological Research, a report to the National Science Foundation on the results of a workshop organized by AERC.

Alyward, B.A. et al. 1993. The Economic value of Species Information and its Role in Biodiversity Conservation: Case studies of Costa Rica's National Biodiversity Institute and Pharmaceutical Prospecting. A report to the Swedish International Development Authority. London Environmental Economics Centre, London.

American Institute of Biological Sciences. 1995. AIBS review to National Biological Service: content standard for non-geospatial metadata workshop. American Institute of Biological Sciences, Reston, Virginia, USA.

Armentano, T.V. and O.L. Loucks. 1979. Ecological and Environmental Data as Under-Utilized National Resources: Results of the TIE/ACCESS Program. The Institute of Technology (TIE), Indianapolis.

Bailar, J. C., M. Angell, S. Boots, E.S. Myers, N. Palmer, M. Shipley, P. Woolf. 1990. Ethics and Policy in Scientific Publication, Council of Biology Editors, xiii+290.

Baker P.T., and M.A. Little. 1976. Man in the Andes: A Multidisciplinary Study of High-Altitude Quechua. US/IBP Synthesis Series 1. Dowden, Hutchinson, and Ross, Inc., Stroudsburg, Pennsylvannia.

Bard, G.E. 1952. Secondary succession on the piedmont of New Jersey. Ecological Monographs **22**:195-215.

Barnett, V., S. Landau and S.J. Welham. 1994. Measuring sustainability. Journal of Ecological and Environmental Statistics **1**: 21-36.

Barnett, V. Statistics and Long-term Experiments: Past Achievements and Future Challenges. Pages 165-183 *in* R.A. Leigh and A.E. Johnston Long-term Experiments in Agricultural and Ecological Sciences, CAB International, United Kingdom.

Barrett, Hope R. 1995. Catalog of Long-Term Research Conducted by the USDA Forest Service, Northeastern Forest Experiment Station. USDA Forest Service, Randor PA., General Technical Report NE-### (In press).

Barry, J.P., C.H. Baxter, R.D. Sagarin and S.E. Gilman. 1995. Climate-related, long-term faunal changes in a California rocky intertidal community. Science **267**:672-675.

Barton, G.S. 1995. Directory Interchange Format: a metadata tool for the NOAA Earth System Data Directory. Pages 19-23 *in* R. B. Melton, D. M. DeVaney, and J. C. French, editors. The role of metadata in managing large environmental science datasets. Pacific Northwest Laboratory, Richland, Washington, USA.

Battelle Columbus Laboratories. 1975. Final report on evaluation of the theoretical potential for energy conservation in seven basic industries; to Federal Energy Administration. Columbus OH, Battelle Columbus Laboratories, Springfield, Va.

Beatley, J.C. 1976. Vascular Plants of the Nevada Test Site and Central-Southern Nevada. TID-26881, NTES, United States Department of Commerce, Springfield, VA. 308pg.

Blair, F.W. 1977. Big Biology. US/IBP Synthesis Series 7. Dowden, Hutchinson, and Ross, Inc., Stroudsburg, Pennsylvannia.

Boden, T.A. 1995. Metadata compiled and distributed by the Carbon Dioxide Information Analysis Center for global climate change and greenhouse gas-related data bases. Pages 13-18 *in* R. B. Melton, D. M. DeVaney, and J.C. French, editors. The role of metadata in managing large environmental science datasets. Pacific Northwest Laboratory, Richland, Washington, USA.

Bohning-gaese K, M.L. Taper, J.H. Brown. 1994. Avian Community Dynamics Are Discordant In Space and Time. Oikos **70**:121-126.

Bokn, T.L., S.N. Murray, F. E. Moy and J.B. Magnusson. 1992. Changes in fucoid distributions and abundances in the innner Oslofjord, Norway: 1974-80 versus 1988-90. Acta Phytogoegr. Suec. **78**:117-124.

Bowser, C. J. 1986. Historic data sets: lessons from the past, lessons for the future. Pages 155-179 *in* W. K. Michener, editor. Research data management in the ecological sciences. University of South Carolina Press, Columbia, South Carolina, USA.

Bridges K.W., and H.L. Carson. 1981. Biological Organization in Silicted Hawiian Communities. US/IBP Synthesis Series 15. Dowden, Hutchinson, and Ross, Inc., Stroudsburg, Pennsylvannia.

Briggs, J. M., and H. Su. 1994. Development and refinement of the Konza Prairie LTER research information management program. Pages 87-100 *in* W. K. Michener, J. W. Brunt, and S. G. Stafford, editors. Environmental information management and analysis: ecosystem to global scales. Taylor and Francis, Ltd., London, England.

Brown, J. and S. R. Carpenter, editors. 1993. National Center for Ecological Synthesis: Scientiifc Objectives, Structure and Implementation. Report from a committee of the ESA and AERC. 32 pp.

Brown J., P.C. Miller, L.T. Tieszen, and F.L. Bunnell. 1980. An Arctic Ecosystem. US/IBP Synthesis Series 12. Dowden, Hutchinson, and Ross, Inc., Stroudsburg, Pennsylvannia.

Brown, J. H., and J. Roughgarden. 1990. Ecology for a changing earth. Bulletin of the Ecological Society of America **71**:173-188.

Brunt, J. W. 1994. Research data management in ecology: a practical approach for long-term projects. Pages 272-275 *in* J. C. French and H. Hinterberger, editors. Seventh international working conference on scientific and statistical database management. IEEE Computer Society Press, Washington, DC, USA.

Butler, D.R. and G.P. Malanson. 1994. Beaver Landforms. The Canadian Geographer. **38**:76-79.

Canham, C.D., G.G. Parker, and T.G. Siccama. 1992. Permanent Plots: A Directory of Long-term Studies of Vegetation. The Institute of Ecosystem Studies, Millbrook, New York.

Carpenter, S.R. and J.F. Kitchell. 1987. The tempral scale of variance in limnetic primary production. American Naturalist **192**:398-406.

Chrisman, N. R. 1994. Metadata required to determine the fitness of spatial data for use in environmental analysis. Pages 177-190 *in* W. K. Michener, J. W. Brunt, and S. G. Stafford, editors. Environmental information management and analysis: ecosystem to global scales. Taylor and Francis, Ltd., London, England.

Cloern, J.E. 1991. Tidal stirring and phytoplankton bloom dynamics in an estuary. Journal of Marine Research **49**:203-22

COHMAP Members. 1988. Climactic changes of the last 18,000: observations and model simulations. Science **241**:1043-1052.

Collins, S.L. and D.E. Adams. 1983. Succession in Grasslands: 32 years of change in a central Oklahoma tallgrass prarie. Vegetatio **51**:181-190.

Colwell, R. K. 1995. Ecological Society of America special committee on ESA communications in the electronic age. Bulletin of the Ecological Society of America **76**:120-131.

Committee on Earth and Environmental Sciences. 1992. The U.S. global change data and information management program plan. National Science Foundation, Washington, DC, USA.

Conley, W., and J. W. Brunt. 1991. An institute for theoretical ecology? - Part V: Practical data management for cross-site analysis and synthesis of ecological information. Coenoses **6**:173-180.

Dayton, P.K. M. J. Teigner, P.E. Parnell, and P.B. Edwards. 1992. Temporal and spatial patterns of distrubacne and recovery in a kelp forest community. Ecological Monographs **62**: 421-445.

Defense Mapping Agency. 1992. Vector Product Format, Military Standard 600006. Department of Defense, Washington, DC, USA.

Digital Geographic Information Working Group. 1991. DIGEST: A Digital Geographic Exchange Standard. Defense Mapping Agency, Washington, DC, USA.

Dyson, E. 1995. Intellectual Value. Wired:136-

Eagan, P. D., and S. J. Ventura. 1993. Enhancing value of environmental data: data lineage reporting. Journal of Environmental Engineering **119**:5-17.

Edmonds R.L. 1978. Aerobiology: The Ecological Systems Approach. US/IBP Synthesis Series 10. Dowden, Hutchinson, and Ross, Inc., Stroudsburg, Pennsylvannia.

Edmonds R.L. 1982. Analysis of Coniferous Forest Ecosystems in the Western United States. US/IBP Synthesis Series 14. Dowden, Hutchinson, and Ross, Inc., Stroudsburg, Pennsylvannia.

Edmondson, W.T. 1991. The uses of ecology: Lake Washington and beyond. University of Washington Press, Seattle.

Edmondson, W.T. 1994. Sixty years of Lake Washington: a curriculum vitae. Lake Reserve Management. **10**:75-84.

Egler, F.E. 1954. Vegetation science concepts. I. Initial floristic composition, a factor in old field vegetational development. Vegetatio **4**:412-417

Elser, J.J., C.J. Luecke, M.T. Brett, and C.R. Goldman. 1995. Effects of food-web compensation after manipulation of rainbow trout in an oligotrophic lake. Ecology **76**:52-69.

Emmingham, W. H., and G. A. Lundburg. 1977. Climatic and physiological data summaries for the H. J. Andrews reference stand network. Internal report #166, Coniferous forest biome, Ecosystem analysis studies, U.S. International Biological Program, University of Washington, Seattle, Washington, USA.

Evans D.D., and J.L. Thames. 1981. Water in Desert Ecosystems. US/IBP Synthesis Series 11. Dowden, Hutchinson, and Ross, Inc., Stroudsburg, Pennsylvannia.

Federal Geographic Data Committee. 1994. Content standards for digital geospatial metadata (June 8). Federal Geographic Data Committee, Washington, DC, USA.

Frankiln, J.F. 1989. Importance and Justification of Long-Term Studies in Ecology. Pages 3-19 *in* Likens, G.E., editors. Long-Term Studies in Ecology: Approaches and Alternatives. Springer-Verlag, New York.

Franklin, J.F., C.S. Bledsoe and J.T. Callahan. 1990. Contributions of the Long-Term Ecological Research Program: An expanded network of scientists, sites and programs can provide crucial comparative analyses. BioScience **40**: 509-52

Garte, S.J. 1995. Guidelines for training in the ethical conduct of scientific research. Science and Engineering Ethics **1**:59-70.

Goldman, C.R. 1981. Lake Tahoe: Two decades of changein a nitrogen deficient oligotrophic lake. Proceedings of the International Association for Theoretical and Appliesd Limnology. **21**:45-70.

Goldman, C.R., A.D. Jassby, and T.M. Powell. 1989. Interannual fluctuations in primary production: Meteorological forcing at two subalpine lakes. Limnology and Oceanography. **34**:310-323.

Golley F. 1993. A History of the Ecosystem Concept in Ecology: More Than the Sum of the Parts. Yale University Press, New Haven.

Gordon, M. L., J. T. Barrett, R. A. DeAngelis. 1994. Realizing the Information Future. National Academy Press, Washington, D.C., Committee on Physical Sciences and Applications.

Gorentz, J.B. (Ed.). 1992. Data Management At Biological Field Stations and Coastal Marine Laboratories: Report of An Invitational Workshop. National Science Foundation.

Gosz, J. R. 1994. Sustainable Biosphere Initiative: data management challenges. Pages 27-39 *in* W. K. Michener, J. W. Brunt, and S. G. Stafford, editors. Environmental information management and analysis: Ecosystem to global scales. Taylor and Francis, Ltd., London, England.

Hairston, N.G.Jr. 1979. The adaptive significance of color-polymorphism in two species of Diaptonus (Copepoda). Limnol. Oceanogr. **24**:15-37.

Hardy, F.G., S.M. Evans and M.A. Tremayne. 1993. Long-term changes in the marine macroalgae of three polluted estuaries in north-east England. J. exp. mar. Biol. Ecol. **172**:81-92.

Heske, E.J., J.H. Brown, and Q. Guo. 1993. Effects of kangaroo rat exculsion on vegetation structure and plant species diversity in the Chihuahuan Desert. Oecologica **95**:520-524.

Hathway, E.C. and K.E. Hoagland, editors. 1993. ASC Guidelines for Institutional Database Policies. The Association of Systematics Collections, Washington, D.C.

Hinkley, A.D. and P.T. Haug. 1977. User's Guide to Biome Information From the International Biological Program (IBP). The Institute of Ecology, Millbrook, New York. 18pp.

Hobbie J.E. 1980. Limnology of Tundra Ponds: Barrow Alaska. US/IBP Synthesis Series 13. Dowden, Hutchinson, and Ross, Inc., Stroudsburg, Pennsylvannia.

Holmes, R. T., T. W. Sherry, and F. W. Sturges. 1986. Bird community dynamics in a temperate deciduous forest: long-term trends at Hubbard Brook, New Hampshire, USA. Ecological Monographs **56**:201-220.

Huntley, B. and H.J.B. Birks. 1983. An atlas of past and present pollen maps for Europe 0-13,000 years ago. Cambridge University Press, Cambridge, England.

Jamison P.L., S.L. Zegura, and F.A. Milan. 1978. Eskimos in Northwestern Alaska: A Biological Perspective. US/IBP Synthesis Series 8. Dowden, Hutchinson, and Ross, Inc., Stroudsburg, Pennsylvannia.

Johnston, A.E. 1991. Liebig and the Rothamsted experiments. Pages 37-64 *in* Judel, G.K. and M. Winnewisser, editors. Symposium "150 Jahre Agrikulturechemie", Justus Liebig Gesellschaft zu Giessen, Giessen.

Kareiva, P., and M. Anderson. 1988. Spatial aspects of species interactions: the wedding of models and experiments. Pages 35-50 *in* A. Hastings, editor. Community ecology. Springer Verlag, New York, New York, USA.

Kellogg Biological Station. 1982. Data management at biological field stations, report of a workshop held May 17-20, 1982. W. K. Kellogg Biological Station, Michigan State University, Hickory Corners, Michigan, USA.

Kirchner, T. B. 1994. Data management and simulation modelling. Pages 357-375 *in* W. K. Michener, J. W. Brunt, and S. G. Stafford, editors. Environmental information management and analysis: ecosystem to global scales. Taylor and Francis, Ltd., London, England.

Kirchner, T., H. Chinn, D. Henshaw, and J. Porter. 1995. Documentation standards for data exchange. Pages 5-8 *in* R. Ingersoll and J. Brunt, editors. Proceedings of the 1994 LTER data management workshop. Long-Term Ecological Research Network Office, University of Washington, Seattle, USA.

Lathrop, R.C. and S.R. Carpenter. 1992. Zooplankton and their relationship to phytoplankton. Pages 127-150 *in* J.F. Kitchell, editor. Food-web management: a case study of Lake Mendota. Springer-Verlag, New York.

Leigh, R. A. and A. E. Johnston, editors. 1994. Long-term Experiments in Agricultural and Ecological Sciences. CAB International, Oxford, United Kingdom. 428 pp.

Levin, S.A. 1992. The problem of pattern and scale in ecology. Ecology **73**:1943-1967.

Likens G.E. 1989. Long-term Studies in Ecology: Papers from the Second Cary Conference held in Millbrook, New York on May 13, 1987. Springer-Verlag, New York.

Likens, G.E. 1992. Excellence in ecology, 3: The ecosystem approach: its use and abuse. Ecology Institute, Oldendorf/Luhe, Germany.

Lively, C.M., P.T. Raimondi, L.F. Delph. 1993. Intertidal community structure: space-time interactions in the northern Gulf of California. Ecology **74**:162-173.

Louckes, O.L. 1986. The United States' IBP: An Ecosystems Perspective After Fifteen Years. *in* Nicholas Polunin, editor. Ecosystem Theory and Application. Hohn Wiley and Sons Ltd.

Louckes, O.L. and McElreath. 1975. A position paper on national ecological data resources. TIE Committee on ecosystem studies, Madison, WI.

Lubchenco, J., A. M. Olson. L. B. Brubaker, S. R. Carpenter, M.M. Holland, S.P. Hubbell, S.A. Levin, J. A. Mac Mahon, P.A. Matson, J. M. Melillo, H.A. Mooney,C. H. Pterson, H.R. Pulliam, L. A. Real, P. J. Regal, and P. G. Risser. 1991. The sustainable biosphere initiative: An ecological research agenda. Ecology **72**: 371-412.

Luecke, C. 1990. Changes in abundance and distribution of benthic macroinvertebrates after introduction of cutthroat troat into a previously fishless lake. Transactions of the American Fish Society. **119**:1010-1021.

Lundälv, T. and H. Christie. 1986. Comparative trends and ecological patterns of rocky subtidal communities in the Swedish and Norwegian Skagerrak area. Pages 71-80 *in* Heip, C., B.F. Keegan and J.R. Lewis, editors. Long-term Changes in Coastal Benthic Communities. Developments in Hydrobiology vol. 38. W. Junk, Dordrecht, The Netherlands. 340 pp.

Lundälv and Nichols, F.H., J. K. Thompson, L.E. Schemel. 1990. Remarkable invasion of San Francisco Bay (California, USA) by the Asian clam *Potamocarbula amurensis* II. Displacement of a former community . Marine Biology and Ecology **42**: 13-26.

Mahry T.J., J.H. Hunziker, D.R. DiFeo Jr. 1977. Creosote Bush: Biology and Chemistry of Larrea in New World Deserts. US/IBP Synthesis Series. Dowden, Hutchinson, and Ross, Inc., Stroudsburg, Pennsylvannia.

McGowan, J. A. 1990. Climate and change in oceanic ecosystems: the value of time-series data. Tree **5**:293-297.

Melton, R.B. 1995. Metadata in the atmospheric radiation measurement program. Pages 9-12 *in* R. B. Melton, D. M. DeVaney, and J. C. French, editors. The role of metadata in managing large environmental science datasets. Pacific Northwest Laboratory, Richland, Washington, USA.

Michener, W.K. 1986. Research Data Management in the Ecological Sciences. University of South Carolina Press, Columbia SC. Belle W. Baruch Library in Marine Science No. 16 426 pp.

Michener, W.K., J.W. Brunt, and S.G. Stafford. 1994. Environmental Information Management and Analysis: Ecosystem to Global Scales. Taylor and Francis/Burgess Science Press, London U.K. 555pp.

Michener, W. K., R. J. Feller, and D. G. Edwards. 1987. Development, management, and analysis of a long-term ecological research information base: example for marine macrobenthos. Pages 173-188 *in* T. P. Boyle, editor. New approaches to monitoring aquatic ecosystems. ASTM STP 940, American Society for Testing and Materials, Philadelphia, Pennsylvania, USA.

Michener, W.K., A.B. Miller, and R. Nottrott. 1990. Long-term ecological research core data set catalog. Belle W. Baruch Institute for Marine Biology and Coastal Research, University of South Carolina, Columbia, South Carolina, USA.

Miller, S.E. 1993. The information age and agricultural entomology. Bulletin of Entomological Research. **83**:471-474.

Milliken, G.A. 1984. Analysis of Messy Data. Lifetime Learning Publications, Belmont, CA.

Mitchell, R., R.A. Mayer, and J.Downhower. 1976. An evaluation of three biome programs. Science **192**:859-865.

Mittelbach, G.G., A.M. Turner, D.J.Hall, J.E. Rettig, and C.W. Osenburg. 1995. Perturbation and resilience: a long-term whole-lake study of predator extinction and reintroduction. Ecology. (in press).

Mooney H. 1977. Convergent Evolution in Chile and California. US/IBP Synthesis Series 5. Dowden, Hutchinson, and Ross, Inc., Stroudsburg, Pennsylvannia.

Murtaugh, P.A. 1981. Selective predation by *Neoysis mersedis* in Lake Washington. Limnology and Oceanography **26**:445-53.

Myster, R.W. and S.T.A. Pickett. 1990. Initial conditions, history and successional pathways in ten contrasting old fields. The American Midland Naturalist. **124**:231-238.

National Institute of Standards and Technology. 1992. Spatial Data Transfer Standard (Federal Information Processing Standard 173). National Institute of Standards and Technology, Gaithersburg, Maryland, USA.

National Institute of Standards and Technology. 1994. Digital Signature Standard (DSS). Computer Systems Laboratory, National Institute of Standards and Technology, Gaithersburg, Maryland.

National Research Council. 1974. U.S. Participation in the International Biological Program. Report No. 6 of the U.S. National Committee for the International Biological Program. National Academy of Sciences, Washington D.C.

National Research Council. 1991. Solving the global change puzzle: a U.S. strategy for managing data and information. National Academy Press, Washington, DC, USA.

National Research Council. 1993. A biological survey for the nation. National Academy Press, Washington, DC, USA.

National Research Council. 1995a. Finding the forest in the trees: The Challenge of Combining Diverse Environmental Data. National Academy Press, Washington, DC, USA.

National Research Council. 1995b. Preserving scientific data on our physical universe: a new strategy for archiving the nation's scientific information resources. National Academy Press, Washington, DC, USA.

National Science Foundation (NSF). 1994. Grant proposal guide NSF 94-2. National Science Foundation, Arlington, Virginia, USA.

Nichols, F.H., J.K. Thompson, and L.E. Schemel. 1990. Remarkable invasion of San Francisco Bay (California, USA) by the Asian clam Potamorcorbula amurensis. II. Displacement of a former community. Marine Ecology Progress Series **66**:95-101.

Office of Science and Technology Policy. 1991. Policy statements on data management for global change research. U.S. Global Change Research Program, National Science Foundation, Washington, DC, USA.

Ogden, J. C. and J. P. Ebersole. 1981. Scale and Community Structure of Coral Reef fishes: A long-term study of a large artificial reef. Marine Ecology Progress Series **4**:97-103.

O'Neill, R.V. 1976. Ecosystem persistance and heterotrophic regulation. Ecology **57**:1244-53.

Pace, M. L. 1993. Forecasting ecological responses to global change: the need for large-scale comparative studies. Pages 356-363 *in* P. M. Kareiva, J. G. Kingsolver, and R. B. Huey, editors. Biotic interactions and global change. Sinauer Associates Inc., Sunderland, Massachusetts, USA.

Parmenter, R.R., J.W Brunt, D.I. Moore, and S.M. Ernest. 1993. The hantavirus epidemic in the Southwest: rodent population dynamics and the implications for transmission of Hantavirus-associated Adult Respiratory Distress Syndrome (HARDAS). *In* The Four Corners Region Pages 10 in Report to the Federal Centers for Dsease Control and Prevention, Atlanta, Georgia. na.

Peet, R.K. editor. 1985. Plant community ecology: Papers in honor of Robert H. Whittaker. Dr. W. Junk Publishers, Dordrecht, Netherlands.

Peterjohn P.G., Sauer J.R. Population Trends Of Woodlands Birds From The North American Breeding Bird Survey. Wildlife Society Bulletin, 1994 **22**:155-164.

Peterson, C. J., and E. R. Squiers. 1995. Competition and succession in an aspen--white-pine forest. Journal of Ecology **83**:449-457.

Pickett, S.T.A. 1989. Space for time substitution as an alternative to long-term studies. Pages 110-135 *in* G.E. Likens, editor. Long Term Studies in Ecology. Springer-Verlag, New York.

Pimm, S.L. 1991. The Balance in Nature?: Ecological Issues in the Conservation of Species and Communities. The Unversity of Chicago Press.

Pool, R. 1993. Beyond databases and e-mail. Science **261**:841-843.

Powell, T.M., J.E. Cloern, and L.M. Huzzey. 1989a. Spatial and temporal variability in South San Francisco Bay (USA). I. Horizontal kistuibutions of salonity, suspended sedimints, and phytoplankton biomass and productivity. Estuarine, Coastal and Shelf Science. **28**:583-597.

Powell, T.M., J.E. Cloern, and L.M. Huzzey. 1989b. Spatial and temporal variability in South San Francisco Bay (USA). II. Temporal changes in salinity, suspinded sediments, and phytoplankton biomass and productivity over tidal time scales. Estuarine, Coastal, and Shelf Science. **28**:599-613.

Prentki, R.T., D.S. Rogers, V.J. Watson, P.R. Weiter, and O.L. Loucks. 1977. Summary Tables of Lake Wingra Basin Data, (IES Report 1985). Institute of Environmental Studies, University of Wisconsin-Madison.

Reice, S.R. 1994. Nonequilibrium determinants of biological community structure. American Scientist **82**:424-435.

Risser, P.G., editor. 1991. Long-Term Ecological Research: An international perspective. Scope 47, Wiley, Chichester.

Risser, P.G. and J. Lubchenco. 1993. Ten Year Review of the National Science Foundation Long-Term Ecological Research (LTER) Program. National Science Foundation. 50pp.

Risser P.G., E.C. Birney, H.D. Blocker, S.W. May, W.J. Parton, and J.A. Wiens. 1981. The True Prarie Ecosystem. US/IBP Synthesis Series 16. Dowden, Hutchinson, and Ross, Inc., Stroudsburg, Pennsylvannia.

Robbins C.S., S. Droege, and J.R. Sauer. 1989. Monitoring Bird Populations With Breeding Bird Survey and Atlas Data. Annales Zoologici Fennici **26**:297-304.

Roemmich, D. and J. McGowan. 1995. Climatic warming and the decline of zooplankton in the California current. Science **267**:1324-1326.

Scheiber, H.N. 1990. California marine research and the founding of modern fisheries and oceanography: CalCOFI's early years, 1947-1964. CalCOFI Report **31**:63-83

Scholz, D.K., and T.B. Smith. 1995. The Global Land Information System: the use of metadata on three levels. Pages 25-27 *in* R. B. Melton, D. M. DeVaney, and J. C. French, editors. The role of metadata in managing large environmental science datasets. Pacific Northwest Laboratory, Richland, Washington, USA.

Schrader-Frechette, K.S. and E.D. McCoy. 1993. Method in Ecology: Strategies for Conservation. Cambridge University Press, New York. 328pp.

Sieber, J.E. and B.E. Trumbo. 1995. (Not) giving credit where credit is due: Citation of data sets. Science and Engineering Ethics **1**:1-20.

Silvertown, J., M.E. Dodd, K. McConway, J. Potts, and M. Crawley. 1994. Rainfall, biomass variation, and community composition in the Park Grass experiment. Ecology **75**:2430-2437.

Simpson B.B. 1977. Mesquite: Its Biology in Two Desert Scrub Ecosystems. US/IBP Synthesis Series 4. Dowden, Hutchinson, and Ross, Inc., Stroudsburg, Pennsylvannia.

Small, J.A., M.F. Buell, H.F. Buell, and T.G. Siccama. 1971. Old-field succession on the New Jersey Piedmont-The first year. William L. Hutcheson Memorial Forest Bulletin **2**:26-30

Southward, A.J. 1991. Forty years of changes in species composition and population density of barnacles on a rocky shore near Plymouth. J. Mar. Biol. Assoc. U.K. **71**:495-513.

Stafford, S.G., P. B. Alabach, K. L. Waddell, and R. L. Slagle. 1986. Data management procedures in ecological research. Pages 93-114 *in* W. K. Michener, editor. Research data management in the ecological sciences. University of South Carolina Press, Columbia, South Carolina, USA.

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 and Francis, Ltd., London, England.

Strayer, D. J.S. Glitzenstein, C.G. Jones, J. Kolasa, G. E. Likens, M. J. McDonnell, G. P. Parker, and S. T. A. Pickett. 1986. Long-term Ecological Studies: An illustrated account of their design, operation and importance to ecology. Occasional Publication of the Institute of Ecosystem Studies, Number 2. Mary Flagler Cary Arboretum, Millbrook, NY 28pp.

Strebel, D. E., B. W. Meeson, and A. K. Nelson. 1994. Scientific information systems: a conceptual framework. Pages 59- 85 *in* W. K. Michener, J. W. Brunt, and S. G. Stafford, editors. Environmental information management and analysis: ecosystem to global scales. Taylor and Francis, Ltd., London, England.

Swift, M. J. , P.D. Seward, P.G.H. Frost, J.N. Qureshi and F.N. Muchena. 1994. Long-term Experiments in Africa: Developing a Database for Sustainable Land Use under Global Change, Pages 229-252, *in* Leigh, R. A. and A. E. Johnston, editors. 1994. Long-term Experiments in Agricultural and Ecological Sciences. CAB International, Oxford, United Kingdom.

Taper M.L., K. Bohninggaese, J.H. Brown. 1995. Individualistic Responses of Bird Species to Environmental Change. Oecologia **101**:478-486.

Thrower N.W.J., and D.E. Bradbury. 1977. Chile-California Mediterranean Scrub Atlas: A Comparative Analysis. US/IBP Synthesis Series 2. Dowden, Hutchinson, and Ross, Inc., Stroudsburg, Pennsylvannia.

Thurston, J.M., E.D. Williams, and A.E. Johnson. 1976. Modern developments in an experiment on permanent grassland started in 1856: effects of fertilizers and lime on botanical composition and crop and soil analysis. Annales Agronomiques. **27**:1043-1082

Tilman, D., and J. A. Downing. 1994. Biodiversity and stability in grasslands. Nature **367**: 363-365.

United States Office of the Federal Register. 1973. United States Government Manual. Office of the Federal Register, Washington, D.C.

Valone, T.J. and J.H. Brown. 1995. Effects of competition, colonization, and extinction on rodent species diversity. Science **267**:880-883.

Van Cleve, K. And S. Martin. 1991. Long-Term Ecological Research in the United States, A Network of Research Sites. LTER Pub. No. 11, LTER Network Office, Univ.WA Seattle WA. 6th Edition, revised. 178 pp.

Webster, F. 1991. Solving the global change puzzle: a U.S.strategy for managing data and information. Report by the Committee on Geophysical Data Commission on Geosciences, Environment and Resources, National Research Council, National Academy Press, Washington, DC, USA.

West N.E., and J.J. Skujins. 1978. Nitrogen in Desert Ecosystems. US/IBP Synthesis Series 9. Dowden, Hutchinson, and Ross, Inc., Stroudsburg, Pennsylvannia.

Westoby, M. 1991. On long-term ecological research in Australia. Pages 191-209 *in* P.G. Risser, editor. Long-term ecological research. Wiley, New York.

Wiens, J. A. 1989. The ecology of bird communities. II. Processes and variations. Cambridge University Press, Cambridge.

Wilcove, D. S., and J. W. Terborgh. 1984. Patterns of population decline in birds. American Birds **38**:10-13.

Williams, E.D. 1978. Botanical composition of the Park Grass plots at Rothamsted 1856-1976. Rothamsted Experimental Station. Report for 1977. Part 2:31-36.

Appendix A: Abbreviations and Acronyms

ACeDB- A Caenorhabditis elegans Data Base
AERC-Association of Ecosystem Research Centers
ASC- Association of Systematics Collections
CalCOFI- California Cooperative Oceanic Fisheries Investigation
CD-ROM- Compact Disk- Read Only Memory
CERES- California Environmental Resources Evaluation System
CIESIN- The Consortium for International Earth Science Information Network
CODATA- Committee on Data
CGED- Committee on Geophysical and Environmental Data
COHMAP- Cooperative Holocene Mapping Project
CPR- Continuous Plankton Recorder
CRC- Cyclic Redundancy Checks
CSU- Colorado State University
DBMS- Database Management Systems
DDBJ- DNA Data Base of Japan
DIF- Directory Interchange Format
DNA- Deoxyribonucleic Acid
DOD- Department of Defense
DOE- Department of Energy
DOI- Department if the Interior
ECN- Environmental Change Network
EMAP- Environmental Monitoring and Assessment Program
EMBL- European Micro Biological Lab
EPA- Environmental Protection Agency
EPD- European Pollen Database
ESA- Ecological Society of America
ESPS- Ecological Society of America Supplementary Publications Services
EuroMAB- Europe Man and the Biosphere Program/ International Program
FGDC- Federal Geographic Data Center
FLED- Future of Long-term Ecological Data
FOIA- Freedom of Information Act
ftp- File Transfer Protocol
GIS- Geographic Information Systems
GSDB- Genome Sequences Databases
HISPID- Herbarium Information Standards and Protocols for Interchange of Data
HMFC- Hutcheson Memorial Forest Center
IBP-International Biosphere Program
IES- Institute of Ecosystem Studies
ITREB- International Tree Ring Database
LAPD- Latin America Pollen Database

LMER- Land Margin Ecosystems Research
LTER- Long-term Ecological Research
LTPEB- Long-Term Programs in Environmental Biology
LTREB- Long-term Research in Environmental Biology
LTSS- Long-term Studies Section
Mac- Macintosh Computer
NAPD- North American Pollen Database
NAS- National Academy of Sciences
NASA- National Aeronautics and Space Administration
NBIC- National Biological Information Center
NBII- National Biological Information Infrastructure
NBS- National Biological Service
NBS GAP- National Biological Service's Geographic Approach to Protection of Biological Diversity
NCBI- National Center for Biotechnology Information
NCEAS- National Center for Ecological Analysis and Synthesis
NCGR- National Center for Genome Research
netCDF- Network Common Data Format
NGDC- National Geophysical Data Center
NIH- National Institute of Health
NOAA- National Oceanic and Atmospheric Administration
NPP- Net Primary Production
NPS- National Park Service
NRC- National Research Council
NREL- Natural Resource Ecology Laboratory
NSF- National Science Foundation
OBFS- Organization of Biological Field Stations
ORNL- Oak Ridge National Laboratory
OTS- Organization of Tropical Studies
PDB- Pollen Database
QA/QC- Quality Assurance/Quality Control
RCSE- Research Collections in Systematics and Ecology
RTC- Recommended Technological Capabilities
SBI- Sustainable Biosphere Initiative
SDSC- San Diego Supercomputer Center
SMASCH- Specimen Management System for California Herbaria
SRER- Santa Rita Experimental Range
TIE- The Institute of Ecology
TNC- The Nature Conservancy
UCSD- University of California-San Diego
UK- United Kingdom
URL- Universal Resource Locator

US- United States
USDA- United States Department of Agriculture
USFS- United States Forest Service
USGS- United States Geological Survey
WAIS- Wide Area Information Servers
WDC- World Data Center
WWW- World Wide Web

Appendix B: FLED Committee Members, Meeting Participants, and Special Contributors

Committee on the Future of Long-term Ecological Data (FLED)

Edith Allen
Dept of Botany & Plant Sciences
University of California-Riverside
Riverside, CA 92521-0124

Caroline Bledsoe
Dept of Land, Air & Water Resources;
Hoagland Hall
University of California-Davis
Davis, California 95616

Robert Colwell
Dept of Ecology & Evolutionary Biology, U-42
University of Connecticut
Storrs, CT 06269-3042

Paul Dayton
Scripps Institute of Oceanography
LaJolla, CA 92093

Megan Dethier
Friday Harbor Laboratories
Friday Harbor, WA 98250

Katherine Gross, Chair
W.K. Kellogg Biological Station
Hickory Corners, MI 49060

John J. Helly*
San Diego Supercomputer Center
San Diego, CA 92186

Robert Holt
Museum of Natural History
University of Kansas
Lawrence, KS 66045

William Michener
Joseph W. Jones Ecological
Research Center
Newton GA 31770

Nancy Morin
Missouri Botanical Garden
St Louis, MO 63166

Steward T.A. Pickett
Institute of Ecosystem Studies
New York Botanical Garden
Milbrook, NY 12545

Susan Stafford
Forest Science Department
Oregon State University
Corvallis, OR 97331-1393

Catherine Pake, Research Associate
Dept of Ecology & Evolutionary Biology
University of Arizona
Tucson, AZ 85721

---

* Replaced Daniel Sulzbach, Executive Director SDSC, when he took a postion with Genentech, San Fransisco, CA. 94080

Meeting Participants and Workshop Contributors

1. FLED Committee Meeting, November 4-6, 1993, Washington, DC

Attending Committee Members:

Edith Allen, Caroline Bledsoe, Robert Colwell, Paul Dayton, Megan Dethier, Katherine Gross, (Chair), Robert Holt, Beryl Leach Project Manager, William Michener, Nancy Morin, Steward T.A. Pickett, Susan Stafford, and Daniel Sulzbach

Invited Participants:

Michael Allen, Program Officer
Division of Environmental Biology
National Science Foundation
Arlington, VA 22230

James K. Andreasen, Biologist
EMAP, Office of Research and Development, RD 680
Environmental Protection Agency
Washington, DC 20460

Peter Arzberger, Program Director
Division of Biological Instrumentation and Resources
National Science Foundation
Arlington, VA 22230

David Blockstein,
Executive Director
National Institute for the Environment
Washington, DC 20001-4521

Thomas Callahan, Program Officer
Long-Term Projects in Environmental Biology
Division of Environmental Biology
National Science Foundation
Arlington, VA 22230

James Gosz, Director
Division of Environmental Biology
National Science Foundation
Arlington, VA 22230

Anthony Janetos
NASA HQ, Code SE
Washington, DC 20546

Michael Ruggiero,Chief
Inventory and Monitoring Division
National Biological Survey
Washington, DC 20240

Susan G. Schram
Food and Agriculture Program Coordinator,
Consortium for International Earth Science Information Network
(CIESIN)
Washington, DC 20006

Anthony Socci
Division of Environmental Biology
National Science Foundation
Arlington, VA 22230

Paul Uhlir
Associate Executive Director
Commission on Physical Sciences, Mathematics, and Applications (CPSMA)
National Academy of Sciences
Washington, DC 20418

Bruce Umminger
Senior Advisor on Biodiversity
The Smithsonian Institution
National Museum of Natural History
Washington, DC 20560

Robert Unnasch
Director of Biological Management
Stewardship Division
The Nature Conservancy
Arlington, VA 22209

Donald Wilson
Director, Biodiversity Program
The Smithsonian Institution
National Museum of Natural History
Washington, DC 20560

2. FLED Work Group Leader Meeting, March 4-6 1994, SDSC, San Diego CA

Attending Committee Members:

Edith Allen, Caroline Bledsoe, Katherine Gross (Chair), Robert Holt, William Michener, Nancy Morin, Steward Pickett, Dan Sulzbach

Invited Participants:

Ted Case
University Of California-San Diego
Department Of Biology
La Jolla, CA 92093-0116

Michael Gilpin
Department of Biology
University of California San Diego
La Jolla, CA 29093-0116

Michael Mullin (0218)
Marine Life Science Research Group
Scripps Institute of Oceanography
La Jolla, CA 92093

Rich Minnich
Department of Earth Science
University of California-Riverside
Riverside, CA 92521

John Rotenberry
UCR Natural Reserve System
University of California-Riverside
Riverside, CA 92521

John Helly
San Diego Super Computer Center
San Diego, CA 92186

3. FLED Work Group Meeting, April 11-12, 1994, W.K. Kellogg Biological Station, Hickory Corners MI

Attending Committee Members:

Edith Allen, Caroline Bledsoe, Robert Colwell, Megan Dethier, John J. Helly, Katherine L. Gross (Chair), William Michener, Nancy Morin, Steward T.A. Pickett

Invited Participants:

Hal Collins
W.K. Kellogg Biological Station
Hickory Corners, MI 49060

Stuart Gage
Department of Entomology
Michigan State University
East Lansing MI, 48824

4. FLED Work Group Meeting: Sources of Data Long-Term, January 9 -10, 1995, San Diego, CA

Attending Committee Members:

Edith Allen, Paul Dayton, Megan Dethier, John Helly, Katherine Gross (Chair), Cathie Pake (Research Associate)

Invited Participant:

Harvey Chinn
University of California-Davis
Davis, California 95616

5. FLED Work Group Meeting: Metadata, March 27-29 1994, Newton GA

Attending Committee Members:

John Helly and William Michener

Invited Participants:

James Brunt
University of New Mexico
Department of Biology
Albuquerque, NM 87131-1091

Thomas Kirchner
Colorado State University
Natural Resources Ecology Laboratory &
Dept. of Range Science
Fort Collins, CO 80523

6. FLED Work Group Meeting: Data Sharing, April 1-5 1995, Tucson AZ

Attending Committee Members:

Edith Allen, Katherine Gross (Chair), Robert Holt, Catherine Pake (Research Associate)

Invited Participants:

James Brunt, Sevilleta LTER
Dept Biology
University of New Mexico
Albuquerque, NM 87131

Chris Fields
National Center for Genome Resources
Sante Fe, NM 87505

Elaine Hoagland
ASC 730 11th. St. NW
Second Floor
Washington, DC 20001

Malcom Hughes
University of Arizona
Tree Ring Lab
Tucson, AZ 85721

Lucinda McDade
Department of Ecology and Evolutionary
Biology University of Arizona
Tucson, AZ 85721

David Mount
Department of Molecular and Cellular
Biology University of Arizona
Tucson, AZ 86721

Elizabeth Pierson
USGS Desert Lab
University of Arizona
Tucson, AZ 85721

Robert Webb
USGS Desert Lab
University of Arizona
Tucson, AZ 85721

7. FLED Committee Meeting, April 27- 29, 1995, San Diego, CA

Attending Committee Members:

Edith Allen, Caroline Bledsoe, Robert Colwell, Paul Dayton, Megan Dethier, John Helly, Katherine Gross (Chair), William Michener, Steward Pickett and Cathie Pake (Research Associate).

Special Contributors

Scott Collins
National Science Foundation
Arlington, VA 22230

Julie Davis
Oak Ridge National Lab
Oak Ridge, TN 37830

Nancy Ferguson
Ecology and Evolutionary Biology
University of Arizona
Tucson, AZ 85721

John Heuer
Savannah River Ecology Lab
Aikens, SC 29802

Don Henshaw
Andrews LTER Data Manager
Pacific NW Station
Corvallis, OR 97331

Judy Meyer
Institute of Ecology
University of Georgia
Athens, GA 30602

Michael A. Huston
Oak Ridge National Lab
Oak Ridge, TN 37830

Orie Loucks
Department of Zoology
Miami University
Oxford, OH 45056

Dick Olson
Oak Ridge National Lab
Oak Ridge, TN 37830

Robert Robbins
Fred Hutchinson Cancer Research Center
Seattle, WA 98104

Fred Swanson
Forestry Sciences Lab
Corvallis, OR 97331

Dave Strayer
Institute of Ecosystem
Millbrook, NY 12545

Appendix C: Summary of Volume II: Directories to Sources of Long-term Ecological Data

The FLED Committee approached its charge to locate, describe, and determine the curatorial status of long-term ecological data sets from three angles. This resulted in the compilation of three different directories which appear in Volume II. United States Agencies and Organizations that collect and maintain long-term ecological data (Directory 1), Long-term Ecological Data Sets (often maintained by individuals) (Directory 2), and an Annotated Bibliography to Existing Catalogs and Directories of Long-term Ecological Data (Directory 3). Here we present a brief discussion of what these Directories include, as well as how the information was obtained.

Directory 1: United States Agencies and Organizations

This Directory provides an overview of the activities and efforts of over 25 United States agencies and organizations that collect and maintain long-term ecological data. A list of these agencies and organizations are given in Section 1 (Table 1). These descriptions are intended to provide a general introduction to the agency or organization and the types of ecological information they collect as part of their mandate and associated research activities. The summaries are admittedly incomplete as the activities of these agencies and the individuals who work in them are diverse. Our intent with these summaries is to provide a general introduction to assist individuals in locating the appropriate agencies or organizations which would have information of interest.

Among these agencies, we have highlighted specific programs or activities that we believe are particularly relevant to ecologists. This list of programs is not intended to be exclusionary, but rather reflects programs that were brought to our attention by individuals who work in or with these agencies and organizations.

Many of these agencies are currently in the process of providing on-line access to data. To reduce redundancy and inaccuracies in these summaries we have included in the description of each agency and organization an Internet address (when available). Because information is constantly being updated, we encourage individuals to access the latest information available on the Internet.

Directory 2: Long-term Ecological Data Sets.

The Directory of Long-term Ecological Data Sets organizes information on over

150 data sets. Our purpose in compiling this was to determine the scope and nature of the existing long-term ecological data sets, their custodial and curatorial status, and whether they are at serious risk of abandonment. In addition, we hope that this Directory will facilitate communication among scientists who are interested in rescuing legacy data sets and among individuals whose collaborative efforts can tackle questions that could not have been addressed otherwise.

This Directory is divided into two parts: Contributions by Individuals (about 130 data sets) and Contributions by Field Stations. Individual contributors were frequently principal investigators. They completed a Metadata Questionnaire (follows), generally on one or two data sets. In contrast, field station managers usually responded to an abbreviated questionnaire (follows) about several data sets archived at their station.

A list of project titles from individual contributors or participating field stations is given at the end of this section. The lists are arranged alphabetically by the name of the data caretaker or field station. More detailed information on the projects (the metadata) and contact information is given in Volume II (Directory 2). The information in this Directory eventually will be available on-line, through the ESA HomePage.

Contributions from Individuals

Over 100 long-term studies were contributed by 68 individuals to this Directory. None of the datasets discribed in this Directory have been included in previous compilations of Long-term studies (see Volume II, Directory 3). These data sets span a breadth of life forms, from hydrocorals to blue footed boobies, from dinoflagellates to saguaro cacti (Fig. 1). Plants make up the largest portion of the data sets (34%), and most of these focus on trees, shrubs, or cacti. The remaining studies of plants focus on terrestrial herbs (8%) or marine and freshwater plants (3%). Studies of animals are also well-represented, vertebrates comprise 21% and invertebrates and plankton 17% of the data sets. In addition, biogeochemical, streamflow, and climate data relevant to ecologists

40% for 20 years or longer (Fig. 3). The longest studies were conducted for over 90 years.

In about 10 % of the cases, no plan exists for continuation of the study. These represent historically valuable data sets, however, that can provide a snapshot in time that could be compared to similar data collected more recently. A large portion (23%) are at imminent risk of being abandoned due to lack of funds or interest in the project. We believe this to be the tip of the iceberg, however, because our search strategy included scientists associated with LTER and the LTREB program of NSF.

The vast majority (88.6%) of these data sets are entered, or in the process of being entered, into an electronically archivable format. However, most (73.4 %) are merely being archived on personal computers, on a medium that has a relatively short (5-6 years) shelf life. Most of these data sets are either freely available (11.4%) or available with permission from the primary investigator (84.1%). The remainder are not available generally because they are in active use or the data require significant restoration efforts before they could be of value to others (e.g., attachment of metadata).

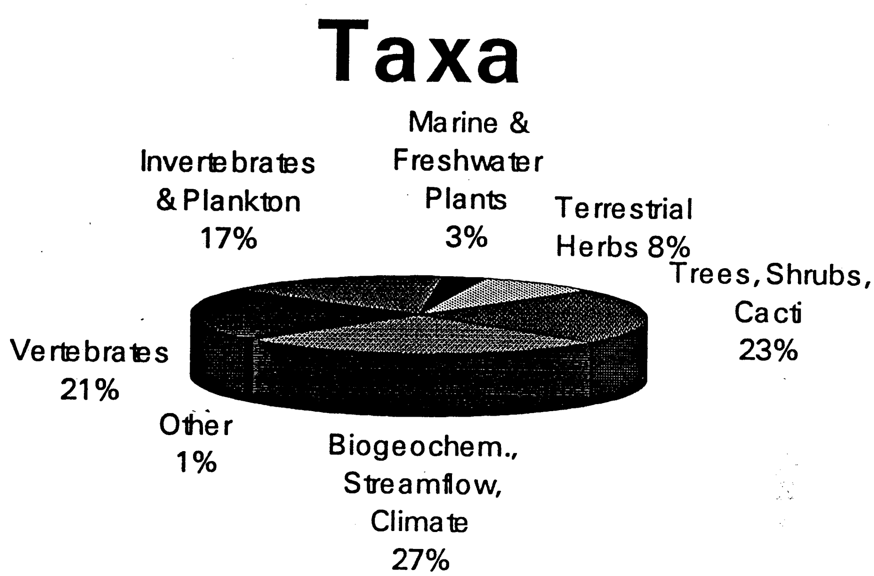


Figure 1. Taxonomic or focal interest of long-term data sets contributed by individuals.

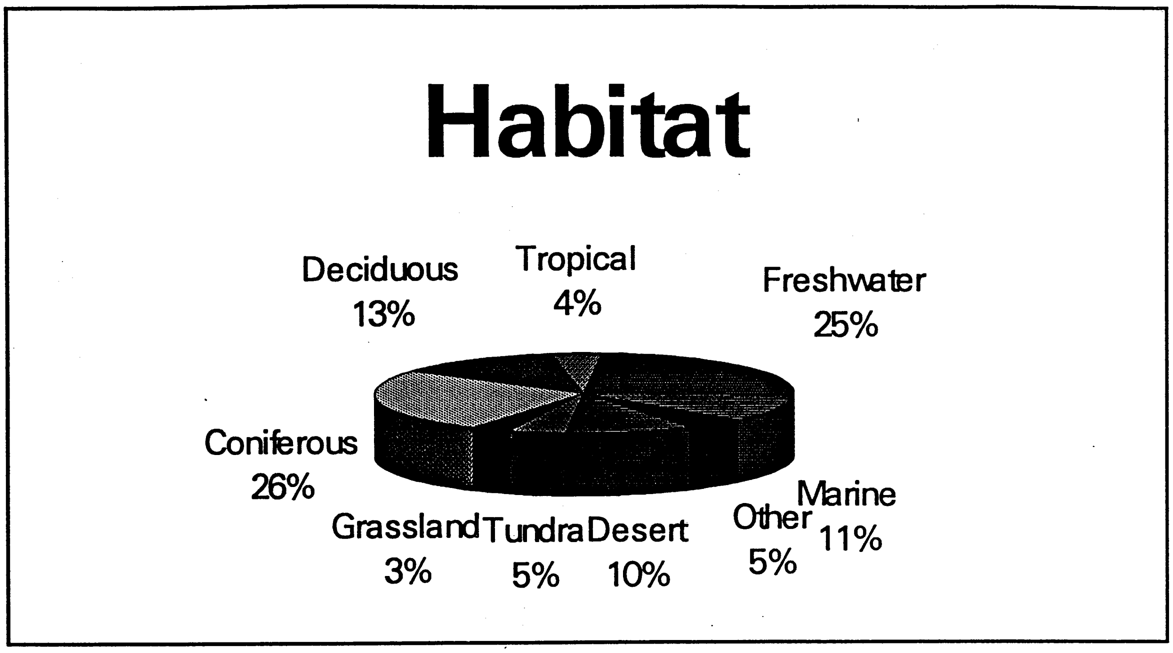


Figure 2. Habitat focuses of data sets contributed by individuals to FLED Directory 2.

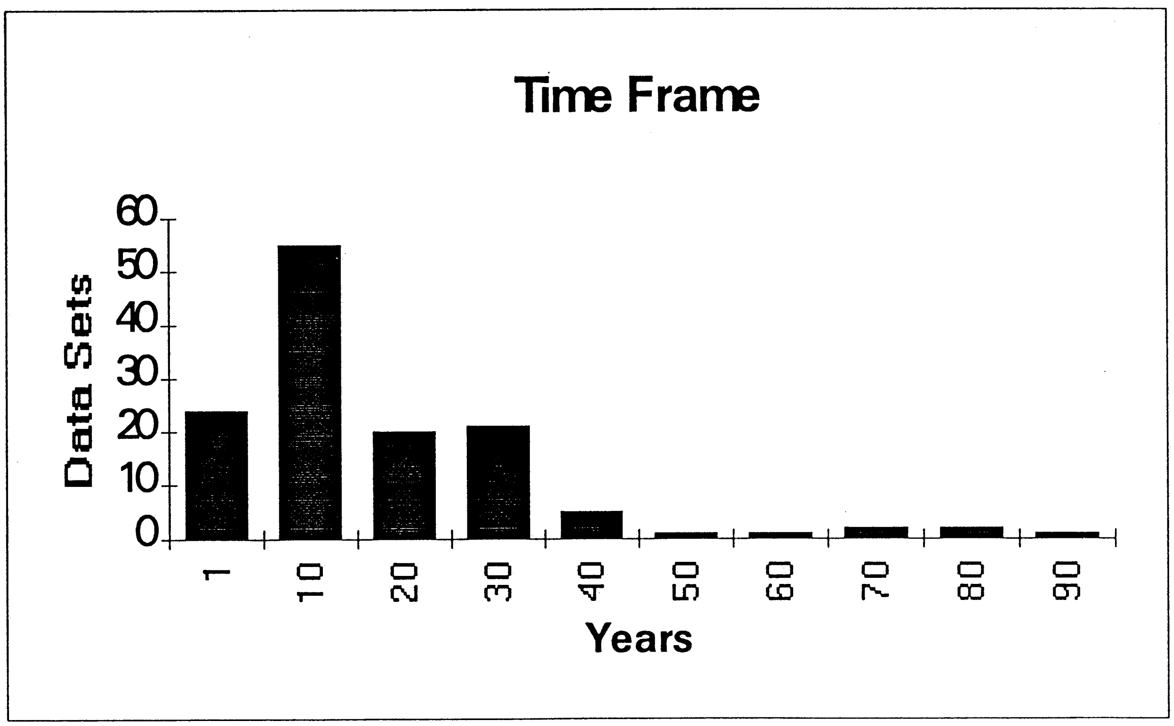


Figure 3. Frequency distribution of time frame (= length of time data collected) data sets contributed by individuals to FLED Directory 2.

Metadata Questionnaire for Individuals.

Recognizing the value of long-term data and the expense involved in acquiring and maintaining it, The Ecological Society of America has established a committee (FLED) to develop recommendations for mechanisms of preserving and archiving long-term data sets. In order to highlight the diversity and scope of long-term ecological data sets collected by individuals, the ESA would like to gather information on specific data sets, their current form, and whether the data are being maintained or are at risk of being abandoned. In this questionnaire, we are interested in information that we could use in aggregate form to inform ourselves before making recommendations. Secondly, with permission, the information will be published in a directory of classic long-term ecological data sets, to be presented to the ESA membership. This directory is likely to be made available both in hard copy and electronically on the Internet.

Information you provide on this form will be summarized and returned to you so that you can verify that the information we have recorded is correct.

1. Project Title:

2. Name of caretaker/contact person. Is he/she data manager, PI, or what?

3. Current institutional address:

4. Phone: FAX:

Email address:

5. Study site--location (county, state, latitude, longitude, etc.) and description:

6. Objectives for data collection. Include organisms, habitats, processes.

7. Methods.

7a. Note sampling methods, variables, types of data collected, and whether treatments were imposed:

7b. Do any of the following data accompany?

___climate ___biogeochemistry
___soils ___other:__________

7c. Relocatability of individuals or plots:
___exactly ___approximately ___no

8. Time Frame:

Start date:

End date:

Censused at what time interval(s)?

Still in progress?

Any gaps in continuity?____

9. Data format?

A. Paper: ___ field notes ___ tables ___ other

B. Electronic: ___ database ___ text ___ other

C. ___Photographs or ___ videotape

D. ___Maps (size, how many):

10. Curatorial status:

(a) Does a plan currently exist for continuation of data collection? ___

(b) Are these data (already collected) at risk of being abandoned?____

© Where are the data physically stored?

___on a personal computer ___box under someone's desk

___on-line data exchange network (which?):__________________

11. Publications using this data? (Up to 3 key complete citations).

12. What funding sources have been used to support this data?

personal funds NSF USFS NPS USGS university______

other: ____

13. Are you willing to be contacted by others interested in this data?______.
These data would be available to other investigators:

___with permission from primary investigators

___without any formal restrictions

Individual Contributors and Project Titles

1. Warren G. Abrahamson, PI - Vegetation changes in burned and unburned Florida scrub, scrubby flatwoods, sandhill, and flatwoods.

2. Steve Acker, PI - Tree Permanent Plots of the Pacific Northwest

3. David J. Anderson, PI - Cost of Reproduction and Evolution of Seabird Reproductive Life Histories

4. Jill Baron, PI - Loch Vale Watershed Long-Term Research Project

5. Barbara Benson, data manager - Historical Water Chemistry Data from the Wisconsin Northern Highland Lake District (Birge and Juday data)

6. Thomas B. Bragg, PI and data manager - Long-term effects of burning on reestablished tallgrass prairie.

7. James Brunt, Bob Parmenter, and Sandra Brantley, data managers - Monitoring Surface-active Arthropod Populations

8. Charles Canham, Gary Levitt, Clyde Jones, Rick Ostfeldt co-PI's - Relationship of forest ecosystem response variables to animal and plant population changes.

9. Phyllis Coley or Thomas Kursar, co-PI's - Long-term studies of herbivore and pathogen damage to tropical trees

10. Richard Condit, PI - Forest Dynamics Plot, Panama

11. Justin Congdon, PI - Life Histories of Long-lived Organisms (Snapping, Banding's, and Painted Turtles)

12. Virginia H. Dale, PI - Plant Reestablishment on the Debris Avalanche at Mount St. Helens.

13. Roger Del Moral, PI - Primary Succession on Mt. St. Helens

14. Art Dunham, PI - Populations in Fluctuating Environments: Population Ecology of Two Species of Arid Adapted Lizards

15. W. T. Edmondson, PI - Changes in salinity of Lake Washington and population dynamics and predation of *Diaptomus*, *Chaetoceros elmorei*, *Daphnia*, and cutthroat trout.

16. W. T. Edmondson, PI - Eutrophication and recovery of Lake Washington: diversion of sewage, effects of land development, and changes in population dynamics, community structure, and pH. Species studied are *Diaptomus*, *Oscillatoria rubescens*, *Epischura-Bosmina*, *Bosmina*, *Daphnia*, *Neomysis*.

17. W. T. Edmondson, PI - Consequences of meromixis loss and re-establishment.

18. Amatzia Genin PI and data manager - Long-term monitoring of the northern Gulf of Eilat

19. Charles R. Goldman, PI - Lake Tahoe Basin Environmental Research Project

20. Charles R. Goldman, PI - LTREB - Interannual Variability, Food-Web Interactions, and Climatic Forcing: A Program for Continued Long-Term Research at Castle Lake

21. Deborah M Gordon, PI - Behavioral Ecology of Harvester Ants

22. Charlie Halpern, PI or Gody Spycher, data manager - Plant Biomass Dynamics Following Logging and Burning in the HJ Andrews Watersheds 1 and 3

23. Mark Harmon, PI - U.S. Forest Service Growth and Field Studies

24. Larry Harris, PI - Benthic Community Structure in the Isles of Shoals

25. Larry Harris, PI - Fouling Panel Study

26. Larry Harris, PI - Kelp Bed Study

27. Paul F. Hendrix, David C. Coleman, D.A.Crossley, Jr., co-PI's - Organic matter dynamics and nutrient cycling in agroecosystems of the Southern Piedmont, Georgia

28. Don Henshaw, data manager - HJ Andrews Network of Meteorologic Stations

29. Don Henshaw, data manager - HJ Andrews Watershed Streamflow Summaries

30. Don Henshaw, data manager - Andrews Proportional Samples: Long Term Stream Chemistry Patterns

31. Don Henshaw, data manager - HJ Andrews Suspended Sediment Grab Samples

32. Shirley Hoh, PI - Monitoring Population of European Rabbit at San Juan Island National Historical Park

33. Michael A. Huston, PI - Tropical Tree Growth of Planted Trees

34. David Inouye, PI - Demography of Frasera speciosa (Gentianaceae)

35. David Inouye, PI - Flowering phenology of Rocky Mountain wildflowers

36. James R. Karr, PI - Ecology of Tropical Forest Birds

37. James R. Karr, PI - Ecology of Fishes in Small Agricultural Streams

38. Patricia Kennedy, PI - Predator-Prey Relationships in forested communities

39. James W. LaBaugh, PI - Hydrological Biogeochemical Interactions

40. Dick Lathrop, PI - Net Plankton of Lake Mendota

41. James N. Layne, PI - Small Mammal populations in major habitats in the southern Lake Wales Ridge region of Florida

42. Micheal R. Pelton and Peter McLean, co-PI's - Population Dynamics of Block Bears of the Smoky Mountains

43. Guy McPherson, PI - Interactions between Perennial Bunch Grass in a Semi-arid Savannah

44. Gary Mittelbach, PI - Fish Abundances and Growth Rates Over a Series of Lakes

45. Jon Moen, Researcher - Herbivory and plant community structure in two contrasting subarctic plant communities.

46. James T. Morris, PI - Long-Term Studies of Salt Marsh Primary Production

47. Vincent Nabholz, PI - Small Mammal Population data on 3 Water sheds at Coweeta Hydrology Lab (USFS)

48. Stephen C. Nodvin, PI - Long-term Inventory and Monitoring of Water Quality and Watershed Processes at the Great Smoky Mountains National Park

49. Dick (R.J.) Olson, PI and data manager - Geoecology: A County-Level Environmental Database for the Coterminous United States

50. Gordon Orians, PI - Demography and Behavioral Ecology of Red-winged Blackbirds

51. Tom Orum, Jeanne Mihail and Stan Alcorn, co-PI's - Annual Saguaro Census and Mortality Survey

52. Bob Paine, PI - Studies on the Structures & Organization of a Rocky Intertidal Community

53. Robert Peet, PI - Forest Succession in the Duke Forest

54. Robert Peet, PI - North Carolina Vegetation Survey

55. Mike Peters, PI - William Proctor Mount Desert Island Biological Survey

56. Elizabeth Pierson, PI and data manager - Demography Trends of Saguaro Populations in the Sonoran Desert.

57. Peter W. Price, PI - Population Data on Stem Galling Sawflies in Flagstaff, Arizona

58. Paul Reeberg, PI - National Park Service Fire Monitoring Program

59. Peter J. Richerson and Thomas H. Suchanek, co-PI's - The Applied Limnology of Clear Lake, California

60. Robert F. Rockwell, PI - Hudson Bay Goose Grazing Project

61. Mark D. Schwartz, PI - Eastern North American Phenology Network

62. Mark D. Schwartz, PI - Wisconsin Phenological Society

63. Rebecca Sharitz, data manager - Tree Population Dynamics in Seven South Carolina Mixed Species Forests

64. Arthur M. Shapiro; PI - Spatial and Temporal Pattern in Butterfly Faunas: Long-Term Studies on a California Transect

65. John Smiley PI or Nathan Rank - Willow-herbivore-predator interactions in Eastern Sierra Nevada Mountains, California

66. Bradley G. Smith, co PI and data manager - Region 6 Ecology Database USDA Forest Service, Region 6 (OR&WA)

67. Una Smith or John Terborgh, PI Spatial Distribution of *Astrocarynum macrocalyx* in a long-term plot, Manu National Park, Peru

68. Thomas J. Stohlgren, and Roger Pielke, co-PI's - Colorado Rockies Global Change Research Program

69. Shiro Tsuyuzaki, PI - Vegetation Recovery After Volcanic Eruptions

70. Charles Vaughn, co-PI - Herbaceous Productivity in Northern California Annual Grasslands

71. D. Lawrence Venable, PI - Demography of Sonoran Desert Winter Annual Plants

72. Robert H. Webb, data manager - Janice Beatly's Ecological Monitoring of the Nevada Test Site 1963-1974

73. Tara Williams, PI - Long Term Monitoring Program for the Southeast Utah Group, National Park Service

74. Jon Witman, PI - Rocky Subtidal Communities in the Gulf of Maine

75. Jon Witman, PI - Long-Term Patch Dynamics of Coral Reefs in St. John, U.S. Virgin Islands

76. Joe Wlosinski, andLinda Leake, data managers - Long Term Resource Monitoring Program for the Upper Mississippi River System

77. Jerry O Wolff, P.I. and data manager - Oak Mast as a Keystone Resource in Forest Community Dynamics

78. John E. Zapotsky, Program Manager - ELF Communications Systems Ecological Monitoring Programs

79. Joy Zedler, PI - Southern California Estuarine Monitoring

Contributions from Field Stations

We received responses from ten field stations that provided us with brief descriptions of 34 long-term data sets being archived. Biological Field Stations and Coastal Marine Laboratories have a history of collecting long-term data relevant to their region and an interest in making them available for secondary use (See Volume II: Directory 3 for on-line address of Organization of Biological Field Stations and Volume I: Part 4 for data sharing concerns of the Organization of Biological Field Stations).

The field stations that have responded to FLED thus far span the continent. Field stations tend to monitor ecologically relevant data on physical characteristics of their region: 13 (38.2%) of the data sets reported consist of biogeochemical, streamflow, or climate data. Nine (26.5%), data sets monitor terrestrial vertebrates and invertebrates, mostly birds and small mammals. Three (8.8%) of the data sets monitor marine organisms (snails, crabs, and plankton). Four (11.8%) data sets deal primarily with plants. Another three data sets (8.8%) are combined species lists or student projects on all taxa and two (5.9%) are photographic monitoring of the region.

Metadata Questionnaire for Biological Field Stations

The Ecological Society of America's FLED Committee, concerned with the Future of Long-term Ecological Data (Chaired by Kay Gross) is investigating many issues surrounding the preservation, curation, restoration, and continuation of long-term ecological studies. One aspect of our efforts is directed at locating long-term data sets. This is motivated by the need for the committee to be informed when making recommendations (e.g., How many data sets are precariously funded? are in need of archiving?). In addition, the information about data sets can serve to promote communication among scientists.

Could you provide us with more information about your Field Station by filling out this brief questionnaire?

(1)

Field Station Name:
Contact:
Phone:
FAX:
Email:
Address:

(2) Long-Term Data Sets archived (maintained) by the Field Station. Feel free to elaborate as much as you like. (Please include whether or not the continued data collection is at risk of abandonment and whether the data archives are at risk of abandonment due to insufficient funds, lack of interest, etc. Also, please note the format of the data (raw, summarized, or published) and whether access to data would require special arrangements, for example permission from particular individuals or investigators.)

2a. Project Title/Organism:

Time Frame: Start date:___ End date:___
Plans for continuation?___ Still in progress?___
Gaps in continuity?___
Collection at Risk ? ___Y ____N
Archives at Risk ? ___Y ____N
Data Availability: __Raw __Summarized __Published
Special Arrangements?

2b. Project Title/Organism

Time Frame: Start date:___ End date:___
Plans for continuation?___ Still in progress?___
Gaps in continuity?___
Collection at Risk? ___Y ___N
Archives at Risk? ___Y ___N
Data Availability: ___raw ___summarized ___published
Special Arrangements:

2c. Project Title/Organism:

Time Frame: Start date:___ End date:___
Plans for continuation?___ Still in progress?___
Gaps in continuity?___
Collection at Risk ? ___Y ____N
Archives at Risk ? ___Y ____N
Data Availability: __Raw __Summarized __Published
Special Arrangments?

3. Long-Term Data Sets In the hands of individual researchers.
Please supply us with the NAMES of researchers who may be interested in participating in our project, their EMAIL ADDRESSES or TELEPHONE NUMBER.

Are any of the projects you have referred to above already summarized in a directory (or example in "Permanent Plots: A Directory of Long-term Studies in Vegetation" --Canham, Parker, Siccama 1992) or already available on-line through an ecological data exchange network? If so, which ones?

Information or questions regarding this Questionnaire should be directed to :

Chair, The Long-Term Studies Section (LTSS)
c/o The Ecological Society of America
2010 Massachusetts Avenue, NW
Suite 430
Washington, D.C. 20036 USA

Project Titles and Contributing Field Stations

1. Archbold Biological Station, FL

First project title: Fire History of Archbold Biological Station.

Second project title: Climatological records of Archbold Biological Station. (Temp., precipitation, relative humidity, evaporation potential

Third project title: Bird nest records from ABS

Fourth project title: Ground water monitoring at ABS 30 seasonal ponds, 3 lakes and 3 wells.

Fifth project title: Limnological Monitoring of Lake Annie, ABS

Sixth project title: Small mammal population monitoring on 5 permanent grids and several transects at ABS.

Seventh project title: Annual mast census (acorns, palmetto berries, hickory nuts) on 4 permanent grids and several transects

2. Bodega Marine Laboratory and Reserve, UC Davis, CA

First project title: Wintering shorebird abundance in Bodega Harbor

Second project title: Ground based photomonitoring of Reserve

Third project title: MOMS meterological and oceanograghic monitoring

Forth project title: Aerial photo monitoring of the Reserve

Fifth project title: Monitoring of Littorina populations on rocky shores

3. F.T. Stone Laboratory, Ohio State University

First project title: Water temperature and transparency

Second project title: Diatoms and Plankton

Third project title: Satellite and Physical data

4. Long Marine Lab, UC Santa Cruz, CA

First project title: Meterological station: wind, air temp., Relative humidity, solar irradiance, barometric pressure, rainfall, more recently include sea temp., ocean surface currents.

Second project title: Monterey Bay oceanographic Data.

5. Manomet Observatory for Conservation Sciences, Massachusettes

First project title: Landbird population trend data

Second project title: Shorebird population trend data

Third project title: Colonial waterbird population/reproductive biology data

6. Mohonk Preserve, Daniel Smiley Research Center, New York

First project title: Mohonk Lake Cooperative Weather Station (NOAA).

Second project title: Shawangunk Mountains Groundwater and Precipitation (pH monitoring)

Third project title: Shawangunk Mountains Species Occurrence Data and Collections - including plants, insects, birds, herps., and mammals

7. Powdermill Biological Station, Carnegie Museum of Natural History, PA

First project title: Small Mammal Demographic Data. (12 species of small mammals;)

Second project title: Bird-Banding Project - Powdermill ~ 163 species.

8. Shannon Point Marine Center, Western Washington University

First project title: Long-term water quality database (seawater)

Second project title: Weather database for high and low air temperature, rainfall, estimated cloud cover:

Third project title: Species list, with sample locations and times of reproduction

9. Nantucket Field Station, University of Massachusetts

First project title: SE Mass./Nantucket herbarium collection

Second project title: Student research papers

Third project title: Hermit crab data

10. St. Croix Watershed Research Station, Minnesota

First project title: Weather data

Second project title: *Andrena asteris:* Long-term colony monitoring.

Third project title: Long-term response of fen communities to siltation.

Directory 3: Annotated Bibliography of Existing Catalogs, Directories and HomePages with Long-Term Ecological Data.

This Directory provides a summary of existing compilations of sources of long-term ecological data the have been compiled by carious groups. We have not included here data or information that is currently maintained by U.S. governmental agencies. Information regarding these agencies and the types of data they provide are given in Directory 1. Only some of these sources currently provide direct access to data sets - though a number plan to have data available (on-line) in the near future. Many of these sources provide expert contact from whom data networking to locate useful sources of long-term ecological data. This listing of sources is by no means a conclusive nor final guide to the vast amount of long-term ecological information available. It should however, along with the information and sources given in Directory 1 provide a useful guide to a variety of established compilations of long-term ecological data. Most of the resources described in this Directory are from U.S. organizations, though some international sources are listed (e.g. Australia, Canada, Brazil) and several of the U.S. sources includes information on international programs (e.g. Smithsonian, Man and the Biosphere). The content of these sources, but many focus on biodiversity, systematic collections, weather and related meteorological data, and environmental monitoring.