

EVOLUTION OF THE KONZA PRAIRIE LTER INFORMATION MANAGEMENT SYSTEM

John M. Briggs

Division of Biology, Ackert Hall, Kansas State University, Manhattan, KS 66506-4901

Abstract. The overall objectives of the Konza Prairie LTER Information Management System are to assure data integrity (correctness, at all times, of all items in the research database), provide security for the database (protection against any loss of data), and facilitate use of data by the original investigator(s) as well as by future investigators. This program has expanded considerably from serving only a localized research group (in its original version in 1981), to its present capabilities of responding to requests for data from investigators across the globe. During this time, protocols for the development of this database have evolved with experiences gained from the growth of the research program on Konza Prairie and knowledge gained from other multi-disciplinary research efforts (especially other LTER sites). It is vital for the continued growth of a research program that the information management system be responsive to growth and adapts accordingly. This is especially true with ever-changing computer technology, and as the scientific use of the data changes over time.

INTRODUCTION

Ecological research is maturing from small-scale studies involving one or a few investigators in a single discipline for a short time period, to multi-disciplinary investigators examining regional and global patterns and processes for possible decade or longer studies. It is essential that a parallel growth in proper scientific information management also occurs (Stafford et al. 1994). This is especially true with the rapid change that has occurred in computer and network technology. The purpose of this chapter is to examine how the information management system of a field research site (the Konza Prairie Research Natural Area (KPRNA)) has developed from serving a localized, small number of independent researchers to its present capabilities of responding to requests for data from investigators across the globe.

KONZA PRAIRIE RESEARCH NATURAL AREA HISTORY

Konza Prairie was established as a research facility in 1972, primarily as a result of the efforts of the late Dr. Lloyd C. Hulbert. Initially, KPRNA included only 371 ha, but additional purchases in the early 1980's expanded the site to its present size of 3,487 ha. It was primarily established to examine the importance of fire, grazing, and climate in maintaining tallgrass prairie. The area is owned by the Nature Conservancy and is leased to the Division of Biology at Kansas State University for long-term research purposes. A watershed-level (catchment unit) fire frequency experimental design that includes replicated long-term unburned (20 yr) and annual, two-, four-, and ten-year frequencies of prescribed spring, summer, fall, and winter fires. Overlaid on this design is a grazing experiment with blocks of watersheds designated as ungrazed, grazed by native ungulates (*Bos bison*) and watersheds grazed by domestic cattle (*Bos taurus*) (Knapp and Seastedt 1998).

The Konza Prairie LTER Program

The Long-Term Ecological Research (LTER) Program of the National Science Foundation began funding research in 1980 (Callahan 1984, Franklin et al. 1990). Konza Prairie was one of the six original LTER sites selected by NSF in 1981 and is now in its fourth funding cycle (<http://climate.konza.ksu.edu/general/lter4/lter4.html>). Since it was a relatively young site in terms of ecological research compared to other LTER sites when the LTER program started, it did not have a large number of old historical data sets as did the other LTER sites. There are only two biological data sets from KPRNA that date prior to 1981 and only seven publications from 1971 to 1980. Thus, at a very early stage of the KPRNA LTER program, it was possible to incorporate sound data ecological management practices. This was based upon the desire of LTER sites not to repeat the IBP's mistake of not having an adequate data management program in support of the research program and also due to the efforts of the PIs during this early stage of development. In addition, meetings of the LTER community and NSF stressed the importance of information management at each site.

The Konza Prairie LTER information management system

During the early 1980's, considerable effort was made by the Konza LTER staff to implement a base-level research data management plan. Its primary goal was to have all interested researchers locate, interpret, and utilize data. This plan was designed using guidelines established by Gorentz et al. (1983) and is documented in Gurtz (1986). The overall objectives of that plan were to: 1) assure data integrity (correctness, at all times, of all items in the database), 2) provide security (protection against loss of data), and 3) facilitate use of data by the original investigator(s) as well as by future researchers. These simple but ambitious objectives are still being followed today (Briggs and Su 1994) and even though computers and perhaps more importantly, network technology are rapidly changing the way ecologists use and share scientific data, these guidelines ensured that the research program at KPRNA and information management system matured together over time.

The Konza LTER investigators are committed to the documentation and archival of data collected at this site. It is considered one of the most important tasks that each investigator performs as part of their effort in the Konza LTER program. The LTER program at KPRNA (KNZ) is dedicated to having all long-term data sets and key short-term data on-line and available to the scientific community and general public as soon as possible. The ultimate goal is to have all data on-line within two years of collection, processing and the completion of appropriate quality control procedures. KNZ LTER researchers have an obligation to make available all LTER-funded data to the KNZ LTER database and to publish those data in a timely fashion. They also recognized that investigators must have a reasonable opportunity for first use of data they have collected. KNZ LTER data are defined and processed for on-line access according to the protocols outlined below.

Type I. Core, long-term data sets (with associated metadata) that address Konza LTER objectives and hypotheses as outlined in LTER proposals I-IV and that are supported primarily by LTER funds. These data will be available on-line two years after data are generated and quality control is completed. We recognized that some data sets would take longer to get on-line than others due to the time required for adequate quality control or due to the demand for certain data by others.

Type II. Short-term data sets supported primarily by LTER funds, key short- or long-term data sets supported by other funding, graduate student data sets, discontinued long-term data sets, one-

time surveys, etc. Data sets in the above categories that are supported primarily by LTER funds must be made available to the data manager/PIs and may be placed on-line within 2 years of completing quality control -- but only at the discretion of the data manager/PIs. Data not supported by LTER funds will be placed on-line only if mutually agreed upon by the investigator(s) and the data manager/PIs.

Our present KNZ information management system includes archived LTER data with an electronic data catalog that allows any person with internet access to browse, examine, or download any data set on-line without any restriction (<http://climate.konza.ksu.edu/toc.html>). The only requirement for individuals who use the data is to acknowledge the source of the data using the following simple format:

"Data for XXX was supported by the NSF Long-Term Ecological Research Program at Konza Prairie Research Natural Area"; where XXX is the list of data set(s) used in the publications, reports, or proposals. (Both the data access policy and the suggested formats for using KNZ data are at: <http://climate.konza.ksu.edu/intro.html>).

To reduce time and errors associated with data entry, and to maintain data integrity, over the past decade, specialized data entry programs and data checking protocols have been developed during the past decade (See Briggs and Su 1994, Briggs et al. this volume for additional information). The design of the current Konza Prairie LTER database is straightforward. All data sets are in ASCII format (with the exception of GIS coverages and satellite images; these are not on-line). Having data in only ASCII format and not in a relational database, does not permit complicated searches or subsetting of the data. Monitoring the use of data sets by researchers over the past ten years has revealed that most scientists simply want information about data sets (metadata) or simple access to the data sets. In the past, KNZ invested in developing a remote user access to our database using an ORACLE interface (Briggs and Su 1994). This was done prior to the huge explosion in WWW access. After six months of use and numerous complaints about structured query language, we simply installed a WWW server (<http://climate.konza.ksu.edu>; <http://climate.konza.ksu.edu>) and started using simple file structure as our database. Use of the site has grown from about two accesses a week using the ORACLE™ database, to over fifty accesses a day during the month of July 1997. As with other LTER sites (see other papers by Baker, Benson, Porter, this volume), we are not only using the WWW as a place to distribute data but as a tool to inform scientists, as well as the general public, about KNZ. However, based upon recent developments in ORACLE and the WWW (Henshaw et al., this volume; Benson, this volume), we are again exploring the possibility of using a relational database such as ORACLE, but this time with an easier interface. Thus, scientists could not only access the data over the web, but as Henshaw et al. (this volume) has demonstrated, the power and utility of a relational database can be fully utilized without scientists needing to know SQL.

It has only been possible to get the KNZ data on-line because of the past commitment that KNZ had to information management. Thus, any site that is interested in information management should begin with the task of identifying the objectives of their information management system and determine how their scientists want to access the data. At present, using the WWW as an outlet for their data sets can be a simple tool, that can, over time, develop as a powerful and vital research tool. However, if the basic premise for information management is not built into a site's system at the beginning, and if the senior scientists are not supportive of the information management system, even the most powerful system is doomed to fail (Strebel et al. 1994). An information management system must involve and be endorsed by the user community it was designed to serve (Stafford et al. 1994).

For long-term security, in addition to our WWW server, we store all archived files (data files that have been entered and verified as correct by the investigator(s)) on a variety of electronic media from 1/2" magnetic tapes, 8mm tapes, hard disks, to re-writeable optical disks. Our goal is to have at least three copies of our database stored in different physical places at all times to reduce the possibility of losing data due to hardware failures, changes in computer technology or disasters. Researchers should plan on computer technology to change and try to build systems that are hardware- and software-independent.

One of the more important and useful products of our information management system has been a "Methods Manual" that details procedures for ongoing and prior studies. KNZ staff has maintained this Methods Manual since 1981 that details how each LTER data set is collected. It includes items such as precise maps of the vegetation survey, sample data sheets, and very detailed procedures on instrument installation and use. The Manual provides the necessary details to interpret the more extensive data documentation files maintained for each data set. This document is updated yearly and a completely revised manual is produced every 5 years. We have found this document to be one of the most valuable items that our research group produces.

SHORTCOMINGS

One of the most important decisions a site has to make when developing an information management system is to decide which data sets are not going to be archived. KNZ has struggled with that decision and for many years tried to document everything (from short-term experiments, graduate student work to other funded projects on KPRNA). However, due to our limited resources, we have been forced to focus only on those data sets that are funded from the core LTER grant. (See <http://climate.konza.ksu.edu/general/lter4/lter4.html> for a complete list). While results from most of these non-documented data sets are published, the resulting publications typically do not include adequate detail to be considered properly documented data sets. Properly documenting and providing access to all data sets for the entire research community is beyond the scope of KNZ staff. We recognize, though, that short-term studies, if properly documented, could be examined in the future, and study sites possibly re-sampled to address new ecological questions. Thus, any short-term data set that is properly documented and archived in reality, becomes a long-term data set. Consequently, we encourage all scientists who work on Konza Prairie to properly document their studies.

SUMMARY

KNZ has learned many lessons in developing and refining their information management system over the past 15 years. Most of these lessons parallel Michener's (this volume) "rules of thumb" for metadata and other data management recommendations (Strebel et al. 1994), but warrant mentioning again.

- Incorporate interactions between scientists and data managers at the beginning of the project. Data managers and scientists need to work and, most importantly, talk to each other on a regular basis, not at the end of development of a project. If the senior scientists are not supportive of the information management system, it will fail.
- It is essential that the data manager(s) listen and respond to the user community. If users don't like the system the data manager is using, it should be changed.
- Plan on computer and network technology to change!

ACKNOWLEDGEMENTS

Konza Prairie Research Natural Area is a preserve of The Nature Conservancy and is managed by the Division of Biology, Kansas State University. This paper was supported by the NSF Long Term Ecological Research Program at Konza Prairie Research Natural Area.

LITERATURE CITED

- 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, Bristol, PA.
- Callahan, J.T. 1984. Long-term ecological research. *BioScience* 34:363-367.
- Franklin, J.E., C.S. Bledsoe and J.T. Callahan. 1990. Contributions of the Long-Term Ecological Research Program. *BioScience* 40:509-523.
- Gorentz, J., G. Koerper, M. Marozas, S. Weiss, P. Alaback, M. Farrell, M. Dyer and G.R. Marzolf. 1983. Data management at biological field stations. Report of a workshop at W. K. Kellogg Biological Station, Michigan State University, May 17-20, 1982. Prepared for the National Science Foundation.
- Gurtz, M.E. 1986. Development of a research data management system. Pages 23-38 in W.K. Michener, editor. Research data management in the ecological sciences. The Belle W. Baruch Library in Marine Science Number 16. University of South Carolina Press, Columbia, SC.
- Knapp A.K., and T.R. Seastedt. 1998. Grasslands, Konza Prairie and long-term ecological research. In A.K. Knapp, J.M. Briggs, D.C. Hartnett and S.L. Collins, editors. Grassland dynamics: Long-term ecological research in tallgrass prairie. Oxford University Press, New York, NY.
- Stafford, S.G., J.W. Brunt and W.K. Michener. 1994. Integration of scientific information management system and environmental research. Pages 3-20 in W.K. Michener, J.W. Brunt and S.G. Stafford, editors. Environmental information management and analysis: ecosystem to global scales. Taylor and Francis, Bristol, PA.
- Strebel, D.E., B.W. Meeson and A.K. Nelson. 1994. Scientific information systems: a conceptual framework. Pages 59-86 in W.K. Michener, J.W. Brunt and S.G. Stafford, editors. Environmental information management and analysis: ecosystem to global scales. Taylor and Francis, Bristol, PA.

