RESEARCH

Dan
Judy / Please report
as LTER &
RWU pub.
Fred

Evolution of the Forest Science Data Bank

A progress report on managing complex research data.

By Susan G. Stafford, Gody Spycher, and Mark W. Klopsch

Development of sound data management and data security is becoming a requirement for national and international research institutions and collaborative research projects. As computer technology rapidly advances, our ability to analyze complex data sets and to build sophisticated models is enhanced, our research becomes more numerically based, and the need for better data management increases. In the past, researchers have tacitly assumed that the data would somehow "take care of themselves"; they have neither given much forethought nor allocated resources to ensure timely documenting, cataloguing, and archiving of data.

Today, collaborative studies by many researchers working cooperatively at different locations across the nation and the globe are becoming more commonplace. Such studies require a carefully designed system for data collection, documentation, and management to ensure future compatibility for sharing and merging data and for constructing broadly based models.

In 1984, we reported on the creation of the Forest Science Data Bank (FSDB) serving the Department of Forest Science at Oregon State University (Stafford et al. 1984). This article updates our progress in managing research data with our research colleagues and the general forestry community.

The Forest Science Data Bank remains integrated in the research process from the initial stages of research planning to information retrieval. In the last several years, data from a variety of individual and interdepartmental research projects in genetics, silviculture, ecology, and physiology have been incorporated in the FSDB. These data and data from the H.J. Andrews Experimental Forest in Blue River, OR (formerly supported by the U.S.-International Biome Program and now a National Science Foundation (NSF) long-term ecological research site) comprise the information managed by the FSDB. We now have more than 1,800 data sets and about 100 megabytes of data. The diversity in data sets has forced us to develop a system to accommodate a wide variety of data types, experimental designs, and research fields. We also place a high priority on enabling users unfamiliar with the original study to retrieve, understand, and use the data in secondary analyses or in synthesis with newer data. Consequently, our approach has had to be pragmatic and efficient while maintaining a high degree of flexibility.

Our motivation has not changed—we still focus on thorough documentation to ensure the usability of data beyond the initial study. Our documentation consists of formats, codes, definitions, study designs, and other information essential to understanding and analyzing the data (Stafford et al. 1986a, 1986b). This emphasis on documentation has been recognized nationally (Freeman 1986, Holzworth and Seilkop 1986) and internationally (Dyer 1987) and has served as a basis for establishing protocols for data management and documentation at several sites of the Long-Term Ecological Research (LTER) network. A major aim of the LTER program is to provide the basis for long-term collaborative research, especially joint experiments and comparative analyses, at different sites within and between biomes. Consequently, it is paramount that sound protocols for data management be in place at each site from a project's inception. These documentation efforts are a vital part of our research data management system and are becoming increasingly important as research dollars are harder to secure, grantsmanship becomes more competitive, and new research endeavors are forced to synthesize new and existing data to broaden their scope of inference and range of applicability.

The overall structure of the FSDB also remains as described previously (Stafford et al. 1984), and the FSDB continues to be managed by the quantitative sciences group in the Department of Forest Science at Oregon State University. However, the composition of the group has changed to reflect our new needs and direction in technology. We continue to have a consulting statistician/forest biometrician, a databank manager, a data analyst/assistant statistician—all with research experience—and up to three student data entry personnel (depending upon demand). We rely more heavily on our microcomputer specialist, and we have added a network coordinator to reflect the new emphasis on microcomputers and the distributed computer environment in which we now work. The strong conceptual framework has permitted us to make changes in personnel, hardware,

Susan G. Stafford is associate professor of forest biometrics, Department of Forest Science, Oregon State University, Corvallis. Gody Spycher and Mark W. Klopsch are research assistants, Department of Forest Science, Oregon State University. This is Pap. FRL 2321-206, For. Res. Lab., Oreg. State Univ.

and software without jeopardizing data integrity.

New hardware and software have, however, permitted major changes in operation. We have significantly improved the mechanics of the overall system and more fully implemented the initial conceptual structure. We have expanded its capabilities and shifted emphasis from mainframe to personal computers (PCs). As PCs increasingly appeared on researchers' desks, we lacked the ability to link PCs and needed a local area network (LAN). A recent NSF grant provided a powerful microcomputer-based graphics workstation, a LAN (Ethnernet), and a large amount of hard-disk storage for our facility.

Our new system offers many advantages. Data sets originally stored on mainframe tapes now reside on a file server accessible from any workstation or PC on the LAN. Documentation files, particularly formats and codes (useful in the old system only for viewing and printing), can now be accessed to generate data entry screens and data editing programs automatically—e.g., in SAS (Statistical Analysis System Institute, Inc. 1985)—or to write programs to export data sets into the major analysis systems. This eliminates manually reentering formats into analysis programs.

We are moving toward a paperless system and have increased our reliance on commercially available database management software. We currently use dBASE III+. Documentation is entered on any microcomputer attached to the LAN in a menu-driven system. Typically, users with a new study first obtain a new data set code (study code) in the appropriate study category. Users are then asked to update the catalog of studies by supplying a study title, names of the principal investigator and others involved, funding source, and other pertinent information. An abstract, data formats, and codes are entered into empty but structured templates. Information common to documentation files (keys) and catalogs, once available, is copied automatically from file to file. Quality control by the quantitative sciences group (QSG) ensures that documentation files are properly entered and that formats and data sets conform to each other. This is achieved through programs and visual checks. Most users work in temporary subdirectories and submit documentation and data sets to the QSG for permanent storage. Except for hard copies of files desired by the user, the system is paperless. We make backup copies of our FSDB files weekly, keeping three copies of all files. One of these copies is kept in a secure location off-site.

Our role within the Department of Forest Science and the College of Forestry continues to be anticipating future needs and facilitating the use of current technology—both hardware and software. Clearly, networking is here to stay and will become more important as networking capabilities increase within the public and private sectors. We are moving into landscape ecology and addressing issues of pattern and patch dynamics. These questions lend themselves to resolution by geographic information systems. Any credible data management system must interface smoothly with these applications and also be suitable for research with LANDSAT data and remote-sensing imagery (Waring et al. 1986). Data documentation protocols must accommodate the majority of data types, and databank personnel must be astute in matching new software to the research needs of their clients. The system will be abandoned quickly if it is perceived as too limited, clumsy, or unfriendly. We see increased reliance on graphics capabilities for depicting the complex interdependencies and the multivariate relationships among data. ■

Literature Cited

DYER, M.I. 1987. Definition and designation of biosphere observatories for studying global change. Scope-MAB workshop report. Biol. Int. 16 (special issue):1–62. Int. Union Biol. Sci., Paris, France.

FREEMAN, R.R. 1986. The national environmental data referral service: a publicly available on-line data catalog. P. 143–54 *in* Research data management in the ecological sciences, W.K. Michener, ed. Belle W. Baruch Libr. in Mar. Sci. 16, Univ. S.C. Press, Columbia.

HOLZWORTH, D.A., and S.K. SEILKOP. 1986. Maintaining data integrity in research databases. P. 181–91 *in* Research data management in the ecological sciences, W.K. Michener, ed. Belle W. Baruch Libr. in Mar. Sci. 16, Univ. S.C. Press, Columbia.

STATISTICAL ANALYSIS SYSTEM INSTITUTE, INC. 1985. SAS language guide for personal computers, version 6. SAS Institute, Inc., Cary, NC. 429 p.

STAFFORD, S.G., P.B. ALABACK, G.J. KOERPER, and M.W. KLOPSCH. 1984. Creation of a forest science data bank. J. For. 82(7):432–33.

STAFFORD, S.G., P.B. ALABACK, K.L. WADDELL, and R.L. SLAGLE. 1986a. Data management procedures in ecological research. P. 93–114 *in* Research data management in the ecological sciences, W.K. Michener, ed. Belle W. Baruch Libr. in Mar. Sci. 16, Univ. S.C. Press, Columbia.

STAFFORD, S.G., M. KLOPSCH, K.L. WADDELL, and R.L. SLAGLE. 1986b. Optimizing the computational environment for ecological research. P. 73–92 *in* Research data management in the ecological sciences, W.K. Michener, ed. Belle W. Baruch Libr. in Mar. Sci. 16, Univ. S.C. Press, Columbia.

WARING, R.H., J.D. ABER, J.M. MELILLO, and B. MOORE III. 1986. Precursors of change in terrestrial ecosystems. BioScience 36(7):433–38.