

## ISSUES AND CONCEPTS OF DATA MANAGEMENT: THE H.J. ANDREWS FOREST SCIENCE DATA BANK AS A CASE STUDY

Susan G. Stafford

Department of Forest Science, Oregon State University, Corvallis, OR 97331-7501

**Abstract.** Managing scientific research information to promote ecological research and facilitate widespread availability to the broader scientific community can be an overwhelming task. The Quantitative Sciences Group in the Department of Forest Science at Oregon State University helps address this need within a context of integrating research information management into the research planning process. The history of the Forest Science Data Bank is described. Lessons learned and strategies for successful long-term ecological information management are shared.

### INTRODUCTION

The scientific community is in the midst of an information explosion coupled with a technology revolution (Stafford et al. 1994). The amount of data beaming down from satellites over shorter and shorter time scales, over larger and larger regions, has been likened to receiving a "Library of Congress" worth of data everyday. Data acquisition is clearly not the problem anymore; managing the data is the challenge! The WWW and Internet connectivity have fueled the scientific community's and funding agencies' expectations for ready-access to on-line data and metadata (i.e., documentation essential for understanding the who, what, when, where, how of the data). Complex issues (e.g., global change, sustainability, biodiversity, and emerging diseases) require interdisciplinary collaboration and synthesis at much broader spatial and longer temporal scales (Levin 1992, Kareiva and Anderson 1988). These issues are faced by individual scientists working alone or in teams associated with personnel at independent biological field stations or as part of the National Science Foundation-funded Long-Term Ecological Research Network. The administrative organization and affiliation are independent of the growing expectation for sound data management policies and procedures.

### THE FOREST SCIENCE DATABANK

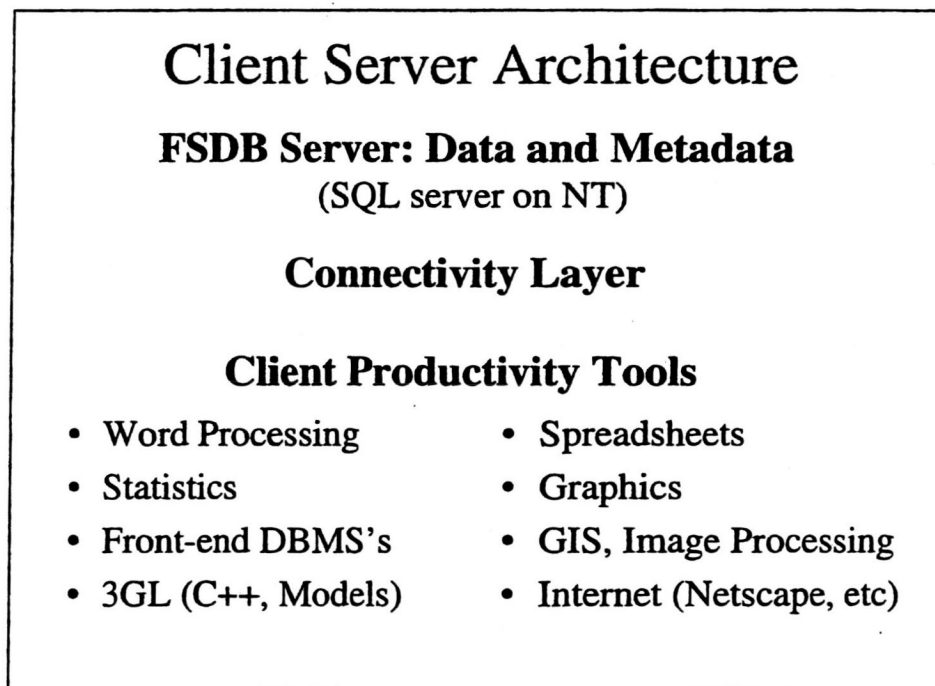
It is important to have a well-defined statement of purpose for managing long-term research information. The mission of the Quantitative Sciences Group (QSG), staffed by both Oregon State University and U.S. Forest Service Pacific Northwest Station personnel, is threefold: to enable success, solve problems, and promote scientific exploration. Our goals are to facilitate research, as well as anticipate future needs. To be successful, research information management must be integrated into research planning. The systematic approach we have used at the H.J. Andrews Long-Term Ecological Research site (LTER) (Franklin et al. 1990) is comprised of: study planning, data production, data analysis, and data interpretation and synthesis (Stafford 1993). This approach can be implemented easily at other research stations.

The Forest Science Data Bank (FSDB) (Stafford et al. 1984, 1986, 1988) was developed by QSG to house data generated by LTER and collaborating scientists. The FSDB has enjoyed a rich history beginning in 1948 when the Blue River Experimental Forest was established (renamed the H.J. Andrews Experimental Forest in 1953), through the decade of the 1970's and the International Biome Program. Our goal has always been to keep improving and changing, mirroring, to the best of our ability, ever-present changes in technology (Table 1). Figure 1 depicts the three layers comprising the current client server architecture of the FSDB: the FSDB server housing data and metadata, a connectivity layer, and client productivity tools.

Table 1. History of the FSDB.

1973-80	Data on mainframe tapes, paper documentation, early abstract and format forms
1980-84	Tape library with automated access facility, documentation in CP/M databases, formalized abstracts, formats & codes.
1984-88	Transition to stand-alone PCs, metadata ported to Xbase, converted mainframe applications to PCs.
1988-93	Tape library ported to Novell server, restructured and cleaned LTER database, development of generic maintenance tools.
1993-96	Refined QC procedures, establish presence on World Wide Web
1997-	Planning port of FSDB onto SQL-server, normalize and expand metadata database.

Figure 1. FSDB client server architecture.



The FSDB "enterprise" encompasses several components: data; documentation (metadata); hardware/software, connectivity tools, and personnel. Space limitations preclude delving deeply into all aspects. Relevant literature is included in the bibliography. Taking a holistic, enterprise-view toward data management allows for a more balanced approach, including effective strategies for dealing with the various critical components and interrelated issues that must be addressed for continuity and long-term success.

## Data

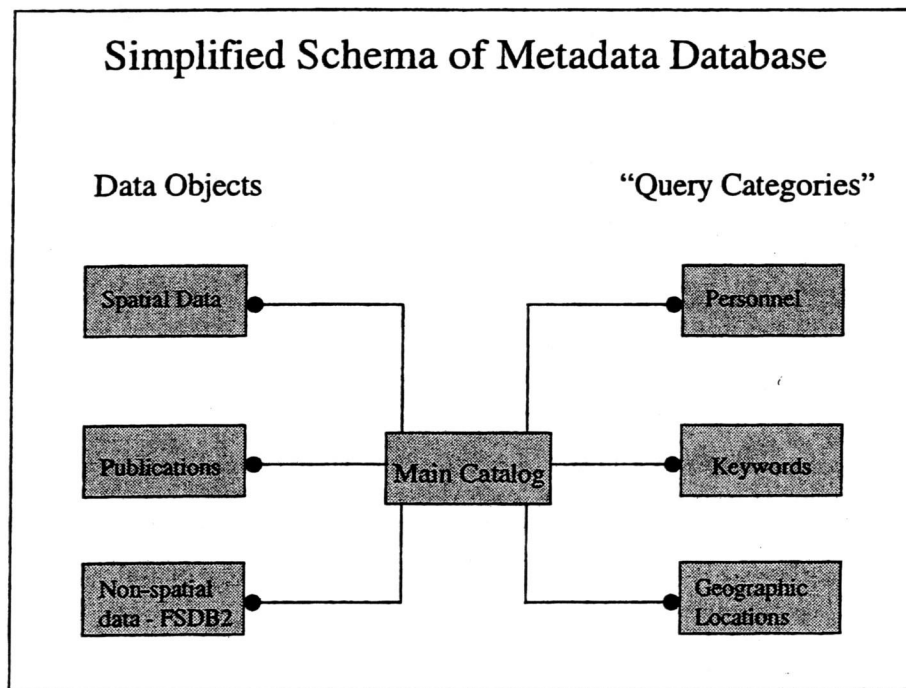
Recognizing that data and metadata are a “corporate asset” and need to be managed as such, the FSDB (<http://www.fsl.orst.edu/lterhome.html>, <http://www.fsl.orst.edu/fslhome.html>) currently houses over 2000 data sets from more than 250 studies. FSDB data include legacy data sets (e.g., IBP data sets), 500 Gb of spatial data (i.e., geographic information system (GIS) coverages and remotely sensed images), models, as well as text documents.

Legacy FSDB datasets include: aquatic/hydrology, geomorphology/vegetation, meteorology, terrestrial vegetation/litter decomposition, biodiversity, wildlife ecology, forest science/genetics/forest engineering, vegetation management and soils data. Specifically, we have over forty years of meteorological and hydrological records (see Henshaw, Bierlmaier, and Hammond, this volume), over eighty years of forest growth and mortality records from six western states, and over thirty years of continuous vegetation succession data.

The legacy data sets are a double-edged sword. Clearly, they add immensely to our wealth of long-term data and significantly increase our overall “portfolio” of data resources. They also pre-date, in many instances, computer technology, so data collection was done differently then and far more care is required to insure adequate documentation of field procedures and study objectives. In addition, the original researchers are frequently no longer on-site and some are deceased.

More recent acquisitions of the FSDB include: a multinational LTER woody debris decomposition project (Harmon 1991), large-scale bird and bird habitat surveys (McGarigal and McComb 1992), and a historical fish habitat database on the Columbia River Basin conducted by the U.S. Fish and Wildlife Service in the 1930's and 1940's (McIntosh et al. 1992).

Figure 2. Metadata database structure.



## *Metadata*

Metadata and data are equally important in the FSDB. The FSDB Metadata system includes: database catalogues, table definition files, domain tables, and tables containing database-specific rules records (for specific examples, see Henshaw, Bierlmaier, and Hammond, this volume; for a more generic discussion of data quality, see Edwards, this volume). We are in the process of further refining our metadata database structure (Figure 2). Examples of various metadata forms (abstract, variable format, variable definition, code definitions, etc.) have been published elsewhere (Stafford 1993).

We have used standardized metadata structures that are identical for every database. We use metadata for data presentation, guiding users in understanding database content, supporting global queries of data catalogs, generating data set documentation exports, and enabling generic access functions [e.g., web page creation, automatic import/export of flat files to relational database management systems (RDBMS) files]. We have used the metadata to develop project-specific "rules" for individual databases. For example, using metadata from the Andrews Reference Stand Monitoring Study, we can run quality assurance (QA) and quality control (QC) checks on newly entered data (for more on this topic, see Edwards, this volume). We can flag entries where trees changed species, shrank dramatically, grew dramatically, or came back to life after being dead for several years.

## *Hardware, software, connectivity, and personnel*

Hardware decisions need to be considered in conjunction with software when assessing connectivity and making personnel decisions. We support six operating systems (AIX, SunOS/Solaris, Data General, Macintosh, Windows-NT, DOS/Windows 3.1x) on 5 platforms (IBM, Sun, Data General, Macintosh, Intel). Personnel are critical to the success of the whole operation. This equates to over 1300 user units. We have been extremely fortunate to hire individuals who are both computer-savvy, as well as interested in science. This has been a winning combination. Disciplinary interests of personnel include soils, statistics, entomology, GIS, remote-sensing, and computer science.

As a point of reference, from 1994 to 1995, the Novell LAN expanded from 180 to 280 PCs and more than tripled the amount of disk space from 5.8 Gb to 18.2 Gb. The UNIX network expanded from eighteen to 27 Suns and from 9.4 Gb public disc space to 25 Gb. In 1997, 450 PCs and 45 Sun Workstations were supported. All decisions need to be considered with an eye toward growth and scalability (See Porter, this volume).

## LESSONS LEARNED AND STRATEGIES FOR SUCCESS

- Strive to avoid creating an US versus THEM situation. Be creative with incentive programs to get buy-in from the research community served. Incentives can include user-friendly applications, "clean" data, shorter delays between collection and analysis, safe storage, and data back-ups.
- Data by themselves are of little value. Metadata must be maintained and recorded to insure the longevity and long-term utility of the data. We have found great variation between data managers' and researchers' expectations of what constitutes "adequate" metadata. Data managers must work with researchers to help educate and train the rest of the scientific

community on what expectations must be met. It is also important to keep track of data requests.

- As you face the future, plan for growth. It is predictable that your systems will grow and expand.
- It is important to start the tradition of managing research information. Use the technology to your advantage and remember, the genius of the future lies not in technology, but in your ability to manage it.

#### ACKNOWLEDGMENTS

I would like to acknowledge National Science Foundation support from DEB No. 90-11663 and 96-32921 to the H.J. Andrews LTER project.

#### LITERATURE CITED

- Franklin, J.F., C.S. Bledsoe, and J.T. Callahan. 1990. Contributions of the long-term ecological research program. *BioScience* 40(7):509-23.
- Harmon, M.E. 1991. The long-term intersite decomposition experiment team (LIDET). Soil Ecology Meeting, April 1991, Oregon State University, Corvallis, OR (Abstract).
- Harmon, M.E. 1992. Long-term experiments on log decomposition at the H.J. Andrews Experimental Forest. USDA Forest Service General Technical Report PNW-GTR-280.
- Kareiva, P. and M. Anderson. 1988. Spatial aspects of species interactions: the wedding of models and experiments. Pages 38-54 in A. Hastings, editor. *Community ecology*. Springer Verlag, New York, NY.
- Levin, S.A. 1992. The problem of pattern and scale in ecology. *Ecology* 73(6):1943-67.
- McGarigal, K. and W.C. McComb. 1992. Streamside versus upslope breeding bird communities in the central Oregon Coast Range. *Journal of Wildlife Management* 56:10-23.
- McIntosh, B. A., J. R. Sedell, and S. E. Clarke. 1992. Historical changes in anadromous fish habitat in the Upper Grande Ronde River Basin, Oregon, 1941-1990. Seventh Annual US Landscape Ecology Symposium, April 1992, Oregon State University, Corvallis, OR.
- Stafford, S. G. 1993. Data, data everywhere but not a byte to read: managing monitoring information. *Environmental Monitoring and Assessment* 26:125-141.
- Stafford, S. G., P. B. Alaback, G. J. Koerper, and M. W. Klopsch. 1984. Creation of a forest science data bank. *Journal of Forestry* 82(7):432-433.
- Stafford, S. G., P. B. Alaback, K. L. Waddell, and R. L. Slagle. 1986. Data management procedures in ecological research. Pages 93-113 in W. K. Michener, editor. *Research data management in the ecological sciences*. The Belle W. Baruch Library in Marine Science No. 16. University of South Carolina Press, Columbia, SC.
- Stafford, S. G., G. Spycher, and M. W. Klopsch. 1988. Evolution of the Forest Science Data Bank. *Journal of Forestry* 86(9):50-51.
- 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 & Francis, London, UK.





# **Data and Information Management in the Ecological Sciences**

## ***A Resource Guide***

*Edited by*

**William K. Michener, John H. Porter, and Susan G. Stafford**

*Published by*

**The Long Term Ecological Research Network Office**

**Department of Biology**

**University of New Mexico**

**Albuquerque**

*This publication should be cited as:*

Michener, W.K., J.H. Porter, and S.G. Stafford. 1998. Data and information management in the ecological sciences: a resource guide. LTER Network Office, University of New Mexico, Albuquerque NM.