DATA AND INFORMATION MANAGEMENT IN THE ECOLOGICAL SCIENCES: SYNOPSIS FROM A FIELD STATION PERSPECTIVE

Hilary M. Swain

Archbold Biological Station, Lake Placid, FL 33862

William K. Michener

Joseph W. Jones Ecological Research Center, Route 2, Box 2324, Newton, GA 31770

Abstract. This paper represents a synopsis of the Data and Information Management in the Ecological Sciences (DIMES) workshop from the viewpoint of the Organization for Biological Field Stations (OBFS). In addition to presenting highlights of the workshop, we examine the effectiveness of the workshop for the member field stations associated with OBFS. This paper is based on closing remarks at the workshop (Swain), interspersed with post-workshop observations by one of the organizers (Michener).

INTRODUCTION

Objectives of the DIMES workshop were to: initiate technology transfer; facilitate interpersonal networking; communicate training opportunities; produce hard copy and digital versions of the DIMES proceedings to serve as a resource guide; and identify future data management needs at field stations and research sites. Implementing onsite data management and integrating data management among sites were described in the opening session as "two of the primary challenges facing field stations over the next decade." The workshop organizers challenged all participants to two proximate workshop objectives; meet ten new people and learn ten new things. Post-workshop evaluations ranged from comments such as "overwhelming," to "great -- right on target," to "not technical enough," and reflected the broad spectrum of backgrounds and interests of the attendees. However, the general consensus of participants was that the workshop largely exceeded expectations. The diverse speakers were extremely effective at conveying information to attendees about data management, and most participants left with the sense that attention to data management is "increasingly overdue" at many field stations, and that many tools and techniques are available to facilitate data management at field stations and other institutions.

PARTICIPANTS

The DIMES workshop attracted approximately 100 participants. A survey of 65 attendees showed that the workshop reached its target audience. Geographic and institutional (Table 1) representation was diverse. Almost half (48%) of the attendees described themselves as data managers.

Table 1. Affiliation of workshop attendees.

Organization	Percentage
Organization for Biological Field Stations	38%
Field Stations and Research Centers (non-OBFS affiliated)	45%
LTER sites or LTER Network Office	28%
University (Faculty and Student)	26%
National Laboratories (e.g., Oak Ridge National Laboratory, National Center for Ecological Analysis and Synthesis)	5%
State or Local Government Agency	5%
Federal Agencies	15%
Other (e.g., consultants)	5%

Table 2. Strengths and difficulties of field stations and field research sites from a data management perspective.

	Strengths	Difficulties
1.	The size, diversity, and longevity of the legacy data sets held at field stations, and the institutional site-based knowledge, are an invaluable and irreplaceable ecological resource.	The rate of entropy (loss of information content over time) of the legacy data sets presents a seemingly paralyzing data management backlog for many field stations.
2.	Many of the legacy data sets are fairly site-specific with fewer thematic components, and therefore potentially easier to integrate.	Most field stations and sites are now also tackling regional analyses and cross-site comparisons. The expansion of spatial, temporal, and thematic scales of ecological study requires scaling up to much more extensive data management.
3.	Field stations typically embody a depth of natural history knowledge that complements the quantitative ecological data sets. In many cases, this tradition has included retaining original data forms and field notes on site.	The natural history aspects of many ecological data, and the scattered documentation of such knowledge, means that data management at field stations must deal with extensive metadata requirements.
4.	Several field stations and sites, particularly the LTER sites, have ongoing data management protocols and institutional policies which can act as models for other field stations.	Ongoing data management problems at field stations are 80% cultural. Large numbers of skeptics among research scientists are yet to be convinced of the value of integrated data management. The proprietary aspects of data have not been resolved at many field stations.
5.	Field stations are entering the computer equipment market at a time when prices have come down considerably, and there is increased capacity to network existing computing facilities.	Chronic budget shortfalls and lack of institutional support for data management are common at many field stations. The costs of data management are high and include: personnel (which may exceed data collection efforts), long-term curation and maintenance, archival facilities and metadata consultation.

A FIELD STATION PERSPECTIVE

Strengths and difficulties of field stations for data management

Previous studies have summarized the inherent strengths and weaknesses of field stations and field research sites from a data management viewpoint (e.g., Gorentz 1992, Gross et al. 1995, Lohr et al. 1995). Speakers at this workshop did not dwell on these issues, but clearly understood that successful data management at field stations is based upon acknowledging existing strengths and accommodating intrinsic difficulties. Discussion of field station strengths (Table 2) was accompanied, in most cases, with an understanding of the offsetting difficulties. The extent to which most speakers recognized the varied field station contexts into which their recommendations have to be implemented was reassuring to field station personnel.

Opportunities and challenges facing data management at field stations

The DIMES Workshop provided an overarching summary of the opportunities to use current tools for data management. The consensus was that "the tools are there" for each step of the data management process. Authors that specifically addressed data management tools in their contributions to this volume are listed below:

- Infrastructure design including hardware (Chapal), communications (Nottrott), and software (Baker)
- Data entry (Briggs)
- QA/QC (Edwards)
- Database management system processing (Porter)
- Metadata (Michener)
- Archival (Olson)
- Scientific visualization (WWW (Benson) and San Diego Supercomputer Center (Helly))
- Data and information resources (e.g., World Wide Web (Benson and others))

Although the tools for data management are generally available, implementation at field stations and field research sites presents a series of challenges. Presenters were encouraged to include "tricks" of the trade that they use to overcome cultural barriers to effective data management. Successful data management is only achieved in social environments that are receptive because there are long-term benefits as well as incentives to participate. Components that presenters viewed as critical for implementation were: institutional incentives and recognition; effective software support; and initial marketing to participants. Successful completion of a *site needs assessment* is critical to facilitate data management design and implementation. Site needs assessments include: identification of data and site objectives; developing policies for data sharing and data ownership; and assessing the infrastructure, personnel, and budget. Workshop participants were interested in seeing real-world examples of cross-site comparisons or interdisciplinary studies where the results clearly demonstrate the scientific value of participating in shared data management, to help market the advantages. Specific challenges include demonstrating how data management has effectively: delayed "data entropy" (*sensu* Michener et al. 1997); supported the use/re-use of data by the data originator and data re-use by others; and facilitated expansion of spatial, temporal, and thematic scales of ecological study.

Field Stations recognize there is a full spectrum of tools available for data management, but have low budgets and limited trained personnel. Workshop presenters provided advice on "low-end" and well as "high-end" solutions (Table 3). Further guidance is needed, however, as to "where to get on the ramp," depending on current circumstances and future needs. Specific topics of interest to field stations and research sites include: *technical interoperability* such as field station infrastructure (e.g., hardware, software, communications) and ecological data archives; *semantic interoperability* including standards (metadata, methods, syntax) and metadata tools (entry, search); *social interoperability* including data and information sharing and technology transfer (training, meetings); *funding* for the computational infrastructure and data recovery; and *reward systems* like recognition for data and metadata publications and other incentives.

Table 3. Synopsis of DIMES Workshop recommendations for low-, medium-, high-end technological solutions for various stages of the data management process.

Task	Low	Medium	High
Data entry	spreadsheet (e.g., EXCEL™)	full-screen data entry program with programmable QA/QC (e.g., EasyEntry™)	full-screen data entry program with QA/QC and database functions (e.g., SAS™ and relational DBMS)
Quality assurance/ quality control (QA/QC)	Manual	Range checks, field validation, etc. (e.g., EasyEntry™, SAS™)	Comprehensive graphical and statistical QA/QC (e.g., SAS™)
Database management system (DBMS)	non-DBMS with data management functions (e.g., merge, subset, Boolean operators, etc. (SAS™))	User-friendly PC-based DBMS (e.g., ACCESS™, PARADOX™)	Comprehensive PC- or UNIX-based DBMS (e.g., ORACLE™)
Archival	redundancy (i.e., disks and paper copies stored in two locations)	Tape, optical disk	off-site data archival facility (e.g., Oak Ridge National Laboratory DAAC)
Metadata	Paper	Word processor	DBMS
Hardware	PCs and printers	Workstation & color output	mixed PC & UNIX, multi-media
Software	WORD™ & EXCEL™	SAS, graphics	ARC/INFO™, ERDAS™
Network	Modem	Internal network (e.g., NOVELL)	Internet & WWW connectivity

FUTURE DIRECTIONS

Training and mentoring

The DIMES Workshop was a recognizable starting point for data management networking based on personal contacts developed at the workshop. Other ideas for training and mentoring included development and utilization of "hands-on" training centers (possibly in conjunction with the National Center for Ecosystem Analysis and Synthesis (NCEAS), Oak Ridge National Laboratory DAAC, or other established facilities). In addition, the concept of site visits by "Rapid Assessment Data Management Teams" was suggested. Such a team might include groups of 2-3 individuals drawn from a pool of experienced data managers who could "jump-start" the planning, design, and implementation processes.

Future meetings

Significant interest was generated at this Workshop for follow-up workshops and meetings. Possible venues include: NCEAS, other workshops sponsored by NSF-DBA, annual "Data Management" workshops/symposia at ESA or other Society-affiliated meetings, and a Journal/ Bulletin Board. It will be worth considering what other potential participants/groups were missing from the attendees at this workshop and how best to include them in future workshops and training efforts.

A closing note

The DIMES Workshop provided a superb compilation of the tools and techniques available to participants for implementing data management. Missing from the discussion, however, was the debate about a broad vision of collective success, in terms of ecological data management. How do various organizations integrate data management across multiple sites and regions? Clearly, LTER sites play a leadership role in this task, but what is the collective vision to tie together data management among the LTER Network, the Organization for Biological Field Stations, the Association of Ecosystem Research Centers, and members of societies such as the Ecological Society of America? Such a collective vision of success will lay the ecological and data management cornerstones upon which future generations can build.

ACKNOWLEDGMENTS

The many members of OBFS who participated in this workshop sincerely appreciated the encouragement and the organizing skills of all the conveners (Michener, Gosz, McKee, and Porter) and also acknowledge generous NSF support for field station participation at this meeting. Participants and Resource Guide users are encouraged to provide feedback to NSF and the workshop organizers (Michener, Gosz, McKee, and Porter).

LITERATURE CITED

Gorentz, J.B. 1982. Data management at biological field stations. Report of a workshop at Kellogg Biological Station, April 1990.

Gross, K.L., C.E. Pake, E. Allen, C. Bledsoe, R. Colwell, P. Dayton, M. Dethier, J. Helly, R. Holt, N. Morin, W. Michener, S. T. A. Pickett, and S. Stafford. 1995. Final report of the

Ecological Society of America Committee on the Future of Long-term Ecological Data (FLED). Volume I.
Lohr, S.A. P.G., Connors, J.A. Stanford and J.S. Clegg. 1995. A new horizon for biological field stations. Rocky Mountain Biological Laboratory. Miscellaneous Publication No. 3.
Michener, W.K., J.W. Brunt, J. Helly, T.B. Kirchner, and S.G. Stafford. 1997. Non-geospatial metadata for ecology. Ecological Applications 7(1): 330-342.