Development of Palmer Long-Term Ecological Research Information Management*

Karen S. Baker

Scripps Institution of Oceanography
University of California at San Diego
La Jolla, CA 92093-0218

ABSTRACT

The Palmer Long-Term Ecological Research (LTER) project, established in the fall of 1990, is the seventeenth site in a network of research sites studying ecological processes over seasonal and decadal time scales. Research focuses on the pelagic marine ecosystem in the area West of the Antarctic Peninsula. Such long-term studies require site information to be recorded consistently, archived digitally, and accessed electronically. The Palmer LTER data management model includes a centralized archive and an emphasis on connectivity among participants who are widely distributed geographically. The model promotes a standard vocabulary and structure while accepting a variety of platforms and software.

1.0 THE PALMER SITE

LTER sites, encompassing diverse ecosystems from forests to lakes (Callahan, 1984), support data management to address both current research needs as well as the long-term aspects of these projects. Palmer LTER field work is conducted in Antarctica aboard oceanographic research vessels and at Palmer Station (64°40'S; 64°03'W), located in a protected harbor on the southwest side of Anvers Island midway down the Antarctic Peninsula. Palmer LTER participants with diverse scientific backgrounds joined together from five universities, permitting consideration of a significant portion of the Antarctic ecosystem including light, hydrography and ice, microbial and phytoplankton activity, krill and birds (Smith et al., 1995). Each component had developed previously an independent hardware and software organization specifically optimized for their own particular research questions. Accepting the reality of these diverse computer environments and tools (Table 1), the objective of Palmer LTER data management is to provide as powerful a connectivity as possible rather than promoting a standardization of hardware or software. Indeed, a variety of platforms enrichs options in terms of data analysis and display.

* Presented at Eco-Informa '96, Lake Buena Vista, Florida, 4-7 November 1996.

An emphasis on publicly available software is economical yet maintains flexibility and simplicity (Conley and Brunt, 1991), especially with current rapid developments.

Table 1. PALMER LTER Software Overview

Type	Platform	Name	DATAM	BIOPT	PHYTO	KRILL	BIRDS	MICRO	MODEL
WORDPROCESSOR	mac	word	DATAM	BIOPT	----	KRILL	----	----	----
	pc	word	----	----	PHYTO	----	BIRDS	MICRO	----
	pc	wordperfect	DATAM	----	----	----	----	----	----
	pc	wordstar	----	----	----	----	BIRDS	MICRO	----
	unix	latex	DATAM	BIOPT	----	----	----	----	MODEL
	unix	roff	DATAM	BIOPT	----	----	----	----	
	unix	tex	----	BIOPT	----	----	----	----	MODEL
BIBLIOGRAPHIC	mac	endnote	----	----	----	KRILL	----	----	----
	pc	papyrus	----	----	----	----	----	MICRO	----
	pc	procite	----	----	----	----	BIRDS	----	----
	pc	refman	----	----	PHYTO	----	----	----	----
	unix	bibix	DATAM	BIOPT	----	----	----	----	----
	unix	bibtex	----	BIOPT	----	----	----	----	----
SPREADSHEET	mac	excel	DATAM	BIOPT	----	KRILL	----	----	----
	pc	excel	----	----	PHYTO	----	BIRDS	MICRO	----
	pc	quattropro	----	----	----	----	BIRDS	----	----
STATISTICS	ibm	spss	----	----	----	----	BIRDS	----	----
	mac	matlab	DATAM	BIOPT	----	----	----	----	----
	mac	statview	----	BIOPT	----	KRILL	----	----	----
	mac	superanova	----	----	----	KRILL	----	----	----
	mac	systat	----	BIOPT	----	KRILL	----	----	----
	pc	excel	----	----	PHYTO	----	BIRDS	MICRO	----
	pc	matlab	DATAM	BIOPT	PHYTO	----	----	----	----
	pc	sigmaplot	----	----	----	----	----	MICRO	----
	unix	idl	----	BIOPT	----	----	----	----	MODEL
	unix	matlab	DATAM	BIOPT	----	----	----	MICRO	----
	unix	splus	----	BIOPT	----	----	----	----	----
GRAPHICS	mac	cricketgraph	----	BIOPT	PHYTO	KRILL	----	----	----
	mac	excel	DATAM	BIOPT	----	KRILL	----	----	----
	mac	deltagraph	----	BIOPT	----	KRILL	----	----	----
	mac	macdraw	DATAM	BIOPT	----	KRILL	----	----	----
	mac	matlab	DATAM	BIOPT	----	----	----	----	----
	mac	superpaint	----	BIOPT	----	KRILL	----	----	----
	pc	excel	----	----	PHYTO	----	BIRDS	MICRO	----
	pc	freelance	----	----	----	----	----	MICRO	----
	pc	matlab	DATAM	BIOPT	PHYTO	----	----	----	----
	pc	quattropro	----	----	----	----	BIRDS	----	----
	unix	gmt	DATAM	BIOPT	----	----	----	----	MODEL
	unix	idl	DATAM	BIOPT	----	KRILL	----	----	MODEL
	unix	matlab	DATAM	BIOPT	----	----	----	MICRO	----
	unix	ncargraphics	----	----	----	----	----	----	MODEL
DATABASE	pc	access	DATAM	----	PHYTO	----	----	----	----
	pc	dbase	----	----	----	----	BIRDS	----	----
	unix	ingress	----	----	----	----	----	MICRO	----
GIS	pc	arcinfo	----	----	----	----	BIRDS	----	----
	pc	arcview	DATAM	----	----	----	BIRDS	----	----
	unix	arcinfo	DATAM	----	----	----	----	----	----

The documentation and data storage of the Palmer LTER are organized through an electronic hub at the Institute for Computational Earth System Science (ICESS) at the University of California at Santa Barbara. To take advantage of existing strengths of team members and to foster integration of data, the data management focused initially on several distinct concepts: 1) acceptance of a diversity of computer platforms and tools, 2) establishment of a distributed system of communication, and 3) development of an electronically available central database. Internet connections are accomplished in a variety of ways, including campus broadband connection, gatorbox tunnel across the internet, and dial-up modem. Tools such as browsers using http and gopher as well as file transfers using ftp are used routinely. Connectivity provides electronic mail and file transfer immediately, but also facilitates transfer of computer knowledge, tools and resources.

A centralized archive and an online browse system were developed. The browse system has been superseded by a local gopher server to provide online menu access and by a

web server (http://www.icess.ucsb.edu/lter). Web page access contains pointers to commonly accessed information as well as links with related groups. All Palmer LTER investigators have accounts on the ICESS server so are able to access the archive directly. The central data disk has two levels of access: PAL/LTER and public. Information in both sections is maintained in ascii files so is portable and easily available to all investigators with their current hardware and software. Discussions regarding relational data base use are ongoing. The central archive, itself a backup of each individual investigators data, is backed up on a regular schedule.

To organize documentation, a common vocabulary and structure was developed. The documentation format includes study and data forms where a study consists of a ship cruise or a season on station. Within each study, data sets are either part of the predefined core data taken periodically or part of the opportunistic or mechanistic data. Describing each study is a list of the measurements, a participant list, and an eventlog listing chronologically the type and location of each measurement made during the study. This eventlog has proven an important method for co-ordinating cross-component work. The database is structured so that documentation files are immediately visible online. They do not have to be overseen by a librarian fluent in html; the Palmer archive is automatically updated whenever an investigator adds or updates files. Efforts are ongoing to simplify documentation in order to insure completion even while in the field. In subsequent years, the benefits of protocol manuals become evident. Each component has developed or is assembling procedures so that the data set documentation need only contain changes to general protocols or study specific observations such as unusual sampling situations.

Data is initially stored by the investigator in a directory to which Palmer LTER members have access. When data are released for public access, they appear with the documentation. The data owner decides which data types are appropriate for the database: raw, processed, synthesized and/or derived data. History and decisions with respect to the data are included in the data set documentation. Quality control remains the responsibility of the individual investigator. Periodically the data managers work with the scientists to consider the value of a synthesized or averaged data set from an original larger data set as in the case of both weather and ice data in the Palmer archive.

2.0 PALMER DATA MANAGEMENT

A data manager develops and maintains the central structure in co-ordination with the individual investigators' data analysts who reside at home sites. The data manager was made a member of the Palmer LTER executive committee in 1993 in recognition of the significant role data management must play in the development of an LTER site. Building upon an established department computer/network group has played an important part in support of the Palmer LTER system. Computer system analysts at ICESS provide input on networking, software, hardware and database planning, while making available computer technology. Routine system functions such as data backups, large data set storage, gopher and web servers, color products and peripheral interfacing are available.

Planning that benefits data co-ordination began with the design of regional (900 km by 200 km) and nearshore (3.7 km) sampling grids centered near Palmer Station. The grids provide a common structure to field studies and geographical registration of diverse data sets (Waters and Smith, 1992). Further, electronic mail aliases, a group activity calendar,

periodic electronic agendas, meeting schedules, field documents and the Palmer annotated bibliography are maintained online in addition to all pertinent documents such as abstracts, proposals and meeting notes. A contribution number is issued for each publication directly supported by Palmer LTER funds and a list of such contributions is maintained online. In addition to published works, the bibliography contains works in preparation as well as submitted abstracts which serves to keep scientists informed of work in progress. Development and maintenance of a timeline listing milestones and changes in vision documents the historical perspective for the site.

In this remote environment where members may be in the field or aboard ship for months with internet available intermittently, a simple form of electronic notebook has been developed at the Palmer site. Also, with the inevitable change of personnel and computers in a long-term project, these files have evolved to fill the need for information in one location covering historical summaries, calibrations, and methods. These files with previous results are available when a scientist on board ship must decide on a sampling location. If an instrument is examined, the calibration history is available in these notebooks. File storage format facilitates access by personal computers.

The principal investigators and the data manager work together to create online documentation for each study. Annual reports along with maps are completed while a similar report for each season is under development. Attention is given to possible data synthesis and standardization. For example, overview tables for each cruise have become a standard, generating interest in similar tables for each season.

Data sets from past Antarctic projects within the Palmer area have been made available locally or referenced on the net. Data such as coastline and bottom topography for the area have been acquired. In addition, with climate variability of central importance in long-term ecological research, the data manager is active in working with weather records within the Palmer area (Baker and Stammerjohn, in press; Smith et al., in press). Two Automatic Weather Station (AWS) units, installed for the Palmer LTER in cooperation with the University of Wisconsin, are monitored and sensor failures identified rapidly.

3.0 NETWORK DATA MANAGEMENT

The LTER sites together form a network which itself interacts with other national and international networks. Whether scientists at an LTER site discuss local data management, LTER site data managers collaborate on an information system, or nations consider monitoring for the Antarctic Treaty, all must consider how best to pursue a unified plan given existing, independent needs of the participants. From general concepts on ecological data management (Michener, 1986; Michener et al., 1994) have developed a variety of site specific models (Stafford et al., 1988; Briggs and Su, 1994; Veen et al., 1994; Ingersoll et al., in press). In addition to learning from other community efforts (Thorley and Trathan, 1994; JGOFS, 1988; National Research Council, 1995), each new site benefits from the collective experience of the other LTER sites. Information is exchanged by site data managers through periodic electronic interactions and at yearly meetings when each participant provides a site summary report and works within committees. Results are sometimes summarized in reports such as those outlining minimum technology requirements, data management and data policy. Further advantages of being part of the LTER network include participation in cross-site surveys such as those summarizing electronic

equipment as well as in cross-site activities such as the ongoing climate and biodiversity studies. For instance, after the annual data manager meeting in 1992, a site capabilities survey was initiated to facilitate sharing of information across sites (Porter et. al., this volume). After considering other site policies (Porter and Callahan, 1994), the Palmer LTER developed a data policy to promote rapid internal data availability as well as final public online availability. In response to a biodiversity workshop, a species list for the Palmer site was developed by reviewing archives and building upon the National Oceanic Data Center's comprehensive taxonomic list.

Data management and computer networking are key elements of all LTER sites. The Network Office (http://lternet.edu) facilitates LTER connectivity. Through the LTER Network Office arise opportunities to co-ordinate on cross-site activities such as collaboration with National Atmospheric and Space Administration on atmospheric measurements and with global positioning system (GPS) projects. The Palmer LTER also co-ordinates closely with Antarctic Support Associates, the company responsible for logistics as well as equipment purchase and ownership for the Antarctic.

4.0 CONCLUSION

Information management which facilitates communication and synthesis builds upon key data management strategies. The primary goal for Palmer LTER data management has been to establish connectivity and to make data sets with complete documentation easily accessible as rapidly as possible since this is a requirement for effective, interdisciplinary, long-term research. Whether communicating to the hub, across institutions or to the field team, the development of the internet and the recent increase in reliable network software have played a critical role in the flow of data between sites and individuals. Maintaining as robust and powerful a system as possible while requiring a minimum of maintenance and support, remains a priority. The potential for a broader vision lends excitement to the challenge of addressing information management issues across the boundaries of the Antarctic, oceanographic, ecological and remote sensing communities.

The Palmer LTER is funded by the National Science Foundation Office of Polar Programs, Grant OPP90-11927. The original vision and continued support by the Palmer LTER lead PI's has been significant in data management development. Discussions with colleagues at the LTER sites, the network office and ICESS have been informative.

5.0 REFERENCES

K. S. Baker, and S. Stammerjohn, Palmer LTER: Palmer Station weather records, *Antarct. J. U. S.* in press.

J. M. Briggs, and H. Su, Development and Refinement of the Konza Prairie LTER Research Information Management Program, in *Environmental Information Management and Analysis: Ecosystem to Global Scales*, edited by W. Michener, J. Brunt, and S. Stafford, pp. 87-100, 1994.

J. T. Callahan, Long-term ecological research, *BioScience*, *34*(6), 363-367, 1984.

W. Conley, and J. W. Brunt, An Institute for Theoretical Ecology? Part V: Practical Data Management for Cross-site Analysis and Synthesis of Ecological Information, *COENOSES*, *6*(3), 173-180, 1991.

R. C. Ingersoll, T. R. Seastedt, and M. Hartman, A Model Information Management System for Ecological Research, *BioScience*, in press.

JGOFS, *Joint Global Ocean Flux Study: report of the JGOFS Working Group on Data Management*, Bedford Institute of Oceanography, 1988.

W. K. Michener, J. W. Brunt, and S. G. Stafford (Eds.), *Environmental Information Management and Analysis: Ecosystem to Global Scales*, Taylor and Francis, Bristol, PA, 1994.

W. K. Michener (Ed.), *Research Data Management in the Ecological Sciences*, Univ. of South Carolina Press, 1986.

NRC, *Finding the Forest in the Trees: The Challenge of Combining Diverse Environmental Data*, National Research Council, National Acadamy Press, Wash. D.C, 1995.

J. H. Porter, and J. T. Callahan, Circumventing a Dilemma: Historical Approaches to Data Sharing in Ecological Research, in *Environmental Information Management and Analysis: Ecosystem to Global Scales*, edited by W. Michener, J. Brunt, and S. Stafford, pp. 193-202, 1994.

R. C. Smith, K. Baker, W. Fraser, E. Hofmann, D. Karl, J. Klinck, L. Quetin, B. Prezelin, R. Ross, W. Trivelpiece, and M. Vernet, The Palmer LTER: A Long-Term Ecological Research Program at Palmer Station, Antarctica, *Oceanography*, *8*(3), 77-86, 1995.

R. C. Smith, S. Stammerjohn, and K. S. Baker, Surface Air Temperature Variations in the Western Antarctic Peninsula Region, in *Foundations for Ecological Research West of the Antarctic Peninsula*, edited by R. Ross, E. Hofmann, and L. Quetin, in press.

S. G. Stafford, G. Spycher, and M. W. Klopsch, Evolution of the Forest Science Data Bank, *Journal of Forestry*, *86*(9), 50-51, 1988.

M. R. Thorley, and P. N. Trathan, *Biomass Data Set Documentation*, British Antarctic Survey, Natural Environment Research Council, Cambridge, United Kingdom, 1994.

C. Veen, C. Federer, D. Buso, and T. Siccama, Structure and Function of the Hubbard Brook Data Management System, *Bulletin of Ecological Society of America*, *75*, 45-48, 1994.

K. J. Waters, and R. C. Smith, Palmer LTER: A Sampling Grid for the Palmer LTER Program, *Antarct. J. U. S.*, *27*(5), 236-239, 1992.