Volume 8
Number 2

ISSN 0730-9295

Information Technology and Libraries

June 1989

SPECIAL ISSUE: LOCALLY LOADED DATABASES IN ONLINE LIBRARY SYSTEMS

CONTENTS

ITLBDC 8(2) 97–232 (1989); Inf Technol Libr

ISSN 0730-9295

Information Technology and Libraries

Volume 8, Number 2: June 1989

SPECIAL ISSUE: LOCALLY LOADED DATABASES IN ONLINE LIBRARY SYSTEMS

CONTENTS

ITLBDC 8(2) 97–232 (1989); Inf Technol Libr

Information Technology and Libraries

Information Technology and Libraries is the official publication of the Library and Information Technology Association, a division of the American Library Association, 50 E. Huron St., Chicago, IL 60611; *Executive Director:* Linda J. Knutson. The journal is issued quarterly in March, June, September, and December.

Information Technology and Libraries publishes material related to all aspects of library and information technology. Some specific topics of interest are: Automated Bibliographic Control, AV Techniques, Communications Technology, Cable Systems, Computerized Information Processing, Data Management, Facsimile Applications, File Organization, Legal and Regulatory Matters, Library Networks, Storage and Retrieval Systems, Systems Analysis, and Video Technologies. *ITAL* welcomes unsolicited manuscripts. Submissions should follow the guidelines stated under "Instructions to Authors" on page 83 of the March 1989 issue.

Manuscripts of articles, communications, and news items should be addressed to the editor-designate: Thomas W. Leonhardt, *Information Technology and Libraries*, University of the Pacific, Stockton, CA 95211. Copies of books for review should be addressed to: Karin A. Trainer, *ITAL Book Reviews*, Yale University Library, 120 High St., P.O. Box 1603A, Yale Station, New Haven, CT 06520. Copies of software for review should be addressed to: David T. Buxton, *ITAL Software Reviews*, A349 Main Library, University of Arizona, Tucson, AZ 85721. Advertising arrangements should be made with Kurt Murphy, Hayden Library, Arizona State University, Tempe, AZ 85287.

Information Technology and Libraries is a perquisite of membership in the Library and Information Technology Association. Subscription price, $17.50, is included in membership dues. Nonmembers may subscribe for $35 per year in U.S., Canada, Mexico, Spain, and other PUAS countries; $40 in other foreign countries. Single copies, $9. Second-class postage paid at Chicago, Illinois, and at additional mailing offices. *Postmaster:* Send address changes to *Information Technology and Libraries*, 50 E. Huron St., Chicago, IL 60611.

Circulation and Production: Central Production Unit (Eileen Mahoney, Director of Production and Distribution Services; Chere Elliott, CPU Supervisor; Bruce Frausto, Debra M. Hall, Colleen King, and Debra J. Teague), American Library Association, 50 E. Huron St., Chicago, IL 60611.

Publication of material in *Information Technology and Libraries* does not constitute official endorsement by the Library and Information Technology Association or the American Library Association.

Abstracted in *Computer & Information Systems, Computing Reviews, Information Science Abstracts, Library & Information Science Abstracts, Referativnyi Zhurnal, Nauchnaya i Tekhnicheskaya Informatsiya, Otdyelnyi Vypusk,* and *Science Abstracts Publications.* Indexed in *CompuMath Citation Index, Computer Contents, Computer Literature Index, Current Contents/Health Services Administration, Current Contents/Social & Behavioral Sciences, Current Index to Journals in Education, Education, Library Literature, Magazine Index, NewSearch,* and *Social Sciences Citation Index.* Microfilm copies available to subscribers from University Microfilms, Ann Arbor, Michigan.

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984. ♾

Expanding the Online Catalog

William Gray Potter

At one time, a library's catalog was designed to index every intellectual work in the collection—not just books, but also articles in periodicals held by the library, pamphlets, maps, government publications, the whole range of materials acquired by a library.[1] By 1900 all but a few highly specialized libraries found this was an impossible task, and today most catalogs provide access only to the books and the set titles of serials. Readers interested in the other types of materials are compelled to consult indexes that are separate from the catalog such as printed indexes, CD-ROM databases, or commercial online services. While these tools are usually of high quality, they are scattered throughout a library. The reader has lost the unifying function of the catalog, the ability to locate any item from a single source.

Tyckoson estimates that most library catalogs, be they on cards or online, index only 2 percent of the works in a collection.[2] The remaining 98 percent of the works include journal articles, government publications, individual essays in collections, maps, pamphlets, etc. Many libraries have specialized collections with their own catalog or index distinct from the main catalog.

Technology offers a possible solution to return libraries to the ideal of yesterday by providing access to a greater variety of material through the online catalog. Some libraries have loaded indexes to journal articles along with their main catalog. Others have loaded indexes to specialized collections and to government publications.

Beyond a library's local collection lies a wealth of information held by other libraries or collected by affiliated academic and government departments. While this information may be of tremendous value to readers, access to it is difficult and murky at best. Again, online catalogs have been used to improve this situation.

The articles that follow in this issue of *Information Technology and Libraries* describe the experiences of several libraries who have used technology to provide greater access to the materials available in their collections and, in some cases, to information outside their libraries. In most cases, the work described here has been performed in conjunction with an online catalog so that the reader is provided with a single source, a common interface, a unified environment in which to retrieve information.

These articles represent the experience of the following institutions: the Georgia Institute of Technology, Carnegie Mellon, Dartmouth, Vanderbilt, University of Pennsylvania, the California Institute of Technology, Arizona State University, and the Colorado Alliance of Research Libraries (CARL). Figure 1 provides an overview of the various databases available at these institutions. There is also a general essay by Charles Bailey that suggests some future directions. It should be stressed that this set of articles is not meant to be exhaustive. There are other libraries and organizations working in this area, including the Division of Library Automation of the University of California, Clemson University, the University of Delaware, the University of Southern California, and Lehigh, among others.

The purpose of this essay is to identify some of the major trends or themes re-

William Gray Potter is Associate Dean of Libraries for Technical Services, Automation, and Systems at Arizona State University and Editor of *Information Technology and Libraries*.

flected by the articles that follow. There are three trends that will be discussed: the unification of local collections, providing access to outside resources, and the incorporation of reference works and full text.

THE UNIFICATION OF LOCAL COLLECTIONS

There is a wealth of material in all libraries that is more difficult to find than the books that are usually reflected in catalogs—journal articles, technical reports, essays in collections, songs in collections, government publications, etc. Somewhere in most libraries there are specialized indexes that provide access to this other material, but barriers of cost and effort stand between them and most readers. Online systems now in place in many libraries offer the possibility of providing more unified access to collections.

The indexing power of some online catalogs can unlock works that are buried in the main catalog. For example, the MARC record for many collections of essays contains a contents note that lists the author and title of each essay. The same is true for collections of short stories. Some libraries index this contents note in their online catalog and thus release these individual essays or stories for searching. The CARL software, also used by Arizona State University (ASU), does this, as does the BRS software used by Dartmouth for searching its main catalog. So, even without adding additional databases, some online catalogs can increase access to local collections.

The most common first step of the libraries represented here has been to load periodical indexes alongside the online catalog. The set of databases provided by Information Access Corporation (IAC) has been mounted by Georgia Tech and Carnegie Mellon. Recent years of MEDLINE have been loaded by Vanderbilt, Penn, and Dartmouth. Some of the H.W. Wilson indexes have been loaded by Georgia Tech, ASU, and Vanderbilt. Penn has loaded ABI/INFORM. Cal Tech has mounted a subset of SciSearch that reflects its holdings. Georgia Tech has included *INSPEC*. CARL has created a database called UnCover that consists of the table of contents of up to 10,000 journals received by the member libraries of CARL and that is available to other libraries as well.

In most cases, these databases are presented along with the main catalog so that from a single terminal, a reader is given the choice of searching the main catalog or searching one of these other databases. Even those libraries that do not offer this menu approach, such as Penn, have attempted to keep the search interface for these databases as consistent as possible with that of the main catalog. The intention is to provide as much information as possible from a single source and within a common environment.

Beyond access to articles in journals, some libraries have created their own indexes to specialized collections or to works that are not covered in the general catalog. Usually, these indexes already exist in some other format and are transferred to the online system. Georgia Tech has included a catalog of its collection of architectural slides. Penn has loaded an index to its student newspaper and an index to a special collection of television scripts. ASU has loaded an index to its map collection and to its solar energy collection and plans to include records reflecting the holdings of its Government Documents Collection. It has also created an index to individual songs that are published in collections.

There is a trend, then, to providing users with a common environment from which to learn an increasing amount about what is contained in the whole of a library's collection. This common environment centers, for now, around the online catalog. While we are a long way from providing convenient and unified access to the entire collection, the work represented here is certainly an improvement.

ACCESS TO OUTSIDE RESOURCES

Even if a library could index 100 percent of its collection in a unified system, there are still vast resources that exist outside the library. Collections of other libraries and files created by affiliated agencies can be valuable resources if access is provided.

Libraries that have loaded periodical indexes have, in most cases, already provided access to materials outside their collection. While Cal Tech and Dartmouth tailored these databases to reflect only their hold-

Fig. 1. Databases Available in Selected Local Online Library Systems.

	Periodical Indexes	Reference Works/Full Text	Local Collections	Other
Georgia Tech Hardware = IBM Software = BRS	Magazine Index Newspaper Index Management Contents Computer Index Trade & Industry Reports Applied Science and Technology Index	Grolier Encyclopedia Commerce Business Daily	Architectural Slides	Georgia State University Catalog
Carnegie Mellon Hardware = IBM Software = Stairs	Magazine Index Newspaper Index Management Contents Computer Index Trade & Industry Reports	Grolier Encyclopedia American Heritage Dict.	Architectural Illustrations	Reference Bibliographies
Vanderbilt Hardware = IBM Software = NOTIS/BRS	MEDLINE General Sciences Index* Humanities Index* Social Sciences Index*			
Penn Hardware = IBM Software = BRS	MEDLINE ABI/INFORM		Penn publications Television scripts	
Dartmouth Hardware = DEC Software = BRS	MEDLINE	Grolier Encyclopedia	Theatrical programs and information Popular sheet music	
Cal Tech Hardware = IBM Software = BRS	ISI Sci Search			

Magazine Index, Newspaper Index, Management Contents, Computer Index, and Trade & Industry Reports are products of Information Access Corporation

Applied Science and Technology Index, Business Periodicals Index, Education Index, General Sciences Index, Humanities Index, and Social Sciences Index are published by H. W. Wilson

ings, the other libraries have loaded them completely, including citations to titles that they do not own. Thus, readers are, albeit indirectly, being referred to other libraries where the title is owned. CARL, in its UnCover database, does inform the reader which of the members libraries actually own which titles. At other libraries, readers are referred to interlibrary loan.

More active access to other collections is provided by some libraries. CARL allows readers to search the catalog of any of its member libraries. Georgia Tech has loaded the catalog of Georgia State University into its online system. While not discussed in this set of articles, there are many other sys-

Fig. 1. Databases Available in Selected Local Online Library Systems Continued.

	Periodical Indexes	Reference Works/Full Text	Local Collections	Other
ASU Hardware = Tandem Software = CARL	Applied Science and Technology Index Business Periodicals Index General Sciences Index Humanities Index Education Index Social Sciences Index UnCover Article Access	Grolier Encyclopedia Songs in Collections	Maps Solar Energy	Career Services
CARL Hardware = Tandem Software = CARL Shared by fifteen member libraries	UnCover Article Acess	Grolier Encyclopedia Roget's Thesaurus* American Heritage Dict.* Denver Business Journal*	Government Pubs	MetroDenver Facts InfoColorado

*in process of loading March 1989

Magazine Index, Newspaper Index, Management Contents, Computer Index, and Trade & Industry Reports are products of Information Access Corporation

Applied Science and Technology Index, Business Periodicals Index, Education Index, General Sciences Index, Humanities Index, and Social Sciences Index are published by H. W. Wilson

tems that provide this type of access, notably the LCS network in Illinois, MELVYL among the University of California campuses, and, of course, OCLC and RLIN.

There is some indication that for interlibrary programs to work effectively, we must tear down the barriers between the reader and the information sought. In the Illinois LCS system, it was possible for years for a patron at one library to directly request a book from one of the other twenty or so participating libraries. However, a long, complicated command was required. When a user interface program, based on a microcomputer, was introduced, this long command was shortened to a simple question: "Do you want to borrow this book?" If the reader responded "Yes," a paging slip was generated at the holding library, and the book was pulled and sent to the reader's home library. The result was that use of interlibrary borrowing through LCS tripled.[3] Bibliographic access alone is not enough. Delivery systems are needed as well.

There is also a growing recognition that there are databases being built in affiliated agencies, such as academic departments, that could be incorporated with the library's online system. Dartmouth sees its library system as a node in the Dartmouth College Information System (DCIS), a campus wide network with a variety of information resources based on a variety of computers. The three components of DCIS are a powerful user workstation, a high-speed communications network, and various host systems that contain information resources. The library's system is just one of the hosts on DCIS. The aim at Dartmouth is to develop an interface program that will search any of the hosts, using a common protocol, and display the results in a common format. Thus, users could use the same workstation to search the library's catalog or to search a file of Dante commentaries maintained by the Romance Languages department or even a list of student employment opportunities, all using common formats for searching and displaying information. Carnegie Mellon is working toward a similar arrangement with its campus network.

The collection of any library contains only a small subset of the universe of knowledge. Connections between the system that indexes that collection and systems that index other knowledge sets are needed if we are to best serve our readers. The online systems represented here show that these connections are being made.

REFERENCE WORKS AND FULL TEXT

Another feature that is developing as libraries expand their online catalogs is the inclusion of ready reference works. The *Grolier Academic American Encyclopedia* has been incorporated into the systems of Georgia Tech, Carnegie Mellon, Dartmouth, CARL, and ASU. Carnegie Mellon has also loaded the *American Heritage Dictionary*. CARL is also loading this dictionary as well as Roget's *Thesaurus*.

These reference works provide support for users of the catalog because they put background information on topics at the readers' fingertips. For example, someone searching for books about fractal geometry might be helped by the brief article on the topic contained in an encyclopedia. Further, a dictionary might be useful to ensure that fractal is spelled correctly.

Beyond providing such support for searching, the reference tools can stand alone as sources of information. Indeed, several libraries report that the encyclopedia is the most popular database they have loaded. One user at Carnegie Mellon referred to it as "the greatest thing since the invention of bedsheets."[4]

Georgia Tech has also incorporated the full text of *Commerce Business Daily* into its Online Information System. They keep six months of this service online and, as the name implies, update it daily. Faculty can establish profiles so that they can be notified when a relevant article appears.

These full-text services mark a radical departure for online library systems. All the other services provided as part of these systems merely point the reader to where information can be located. The full text of reference works and of journal articles, however, actually delivers the information. The services in place today should serve as an important first step to the eventual provision of a wide range of full texts of reference works, journal articles, and possibly complete books.

ADDITIONAL OBSERVATIONS

There is a slight identity crisis reflected in these articles. The terms used to describe the online systems discussed in these articles include Online Catalog (ASU and Dartmouth), Library Information System (Carnegie Mellon), and Online Information System (Georgia Tech). Some may see the word "catalog" as being too restrictive, others may prefer it because it still connotes the primary function of these systems. Georgia Tech and ASU see these systems as a step in the evolution of "electronic libraries." Perhaps that is the name these systems are moving toward.

There are many databases available to load into an online library system, and the selection process for each library is very interesting. Most select databases that satisfy the greatest, most general need. Some look at which databases were used the most in searches of commercial search services. Everyone strives for a balance of service and economy.

All of the libraries report tremendous success with these expanded services. Their reception by faculty and students mirror the general acceptance of online catalogs. However, more than one library reports that successful implementation of these services may bring a spiraling demand for even more services. Readers reacted to the first online catalogs with a request for access to journal articles. They apparently react to the appearance of periodical indexes with requests for even more, usually for additional indexes or for the full text of the article. Apparently, the more libraries do with these systems, the more readers will demand that we do.

Further, success breeds demand on traditional library services. Georgia Tech reports an increase in the use of its periodical collection (measured by pick ups) of 54 percent from fall 1987 to fall 1988. Other libraries have seen similar increases in use.

Carnegie Mellon has noted an increase in traffic at the reference desk. Apparently, we are developing systems that are doing exactly what we want them to do, which is to increase use of the library.

CONCLUSION

There is a common theme among these papers that the services provided by expanding the online catalog (or online library information system) are part of a larger set of services. There are appropriate tools for different levels of expertise and needs. While several libraries have loaded indexes to periodical articles, no library has been able to abandon online searching of commercial databases, and each library is continuing to purchase CD-ROM-based indexes. Emily Fayen, writing about Penn, points out that they have found that they actually need MEDLINE in all three forms—online as part of their local system, on CD-ROM, and through commercial services. The local system only contains up to three years of data but is free to readers. The commercial, remote version of MEDLINE meets the need of the serious researcher who needs to cover the entire file and needs the assistance of an experienced searcher. The CD-ROM version is an excellent teaching tool.

Penn's experience with three tiers of access suggests a pattern for other libraries. There are some databases, usually the more general ones or the ones that match an institution's strengths, that should be incorporated into the local online system. The wide use these databases receive will justify the expense. Other databases may be used less often but frequently enough to justify their purchase on CD-ROM. Still other databases will be used infrequently and may be so difficult to use that mediated searching of an online commercial database is justified.

There is a fourth tier that should be addressed, and that is printed indexes. Some indexes are simply not available in machine readable form. Others cover only the past few years in an online or CD-ROM version. There are also readers who do not want to go near a computer. So, printed indexes should be with us for some time to come.

The next steps for expanding online library systems lie along three complementary paths. The first is simply more of the same—more indexes to more sets of collections and more reference works. The second is the inclusion of the full text of more articles and, possibly, books. For those libraries that have mounted indexes to journal articles as part of their online system, the next logical step is to provide the full text of these articles. While this is technically possible today, the economics of it are cumbersome. Storage costs are significant, but, more importantly, agreements with publishers need to be developed. The third path involves providing greater connectivity from online library systems to other systems, including other library systems, commercial services, bibliographic utilities, local networks, CD-ROM servers, and other information providers in the community.

The progress of automated library systems as reported in this issue of *Information Technology and Libraries* is very promising. We have come a long way in a short time and, by doing so, are returning to ideas we had to abandon years before, ideas of unified and universal access to information. Much of the local collection may still be unavailable, discovering the holdings of other libraries may still be difficult, and the timely delivery of documents may still be unrealized. However, we are moving closer, and the closer we get, the better libraries will become.

REFERENCES

1. David Tyckoson, "The 98% Solution: The Failure of the Catalog and the Role of Electronic Databases," *Technicalities* 9:2 (Feb. 1989), p.8–12.
2. Ibid., p.10.
3. William Gray Potter, "Creative Automation Boosts Interlibrary Loan Rates," *American Libraries*, 17(4):244–46 (Apr. 1986).
4. Nancy Evans and Thomas Michalak, "Delivering Reference Information Through a Campus Network: Carnegie Mellon's Library Information System," *Reference Services Review* 15:7–13 (Winter 1987).

The Online Information System at Georgia Institute of Technology

Miriam A. Drake

The Online Information System at the Georgia Tech library began functioning in May 1985, when the first database, the library's catalog, was installed. Shortly afterwards, one professor, a friend of the library, decided to test the system. He dialed into the campus network at 10:00 one evening, signed onto the library's online information system, found a title he wanted, and sent a message to the library using the message module of the system. The book he wanted was on his desk at 1:00 the next afternoon.

The story illustrates the broad goals of the Georgia Tech library and the role of the Online Information System in achieving them: the system does not exist for its own sake or merely for making bibliographic information accessible. It is part of a larger program aimed at (1) augmenting the amount of information available to students and faculty through the campus network; (2) delivering all forms of information (bibliographic, text, numeric, and graphic) to the PC or workstation; (3) increasing faculty productivity; (4) introducing personalized information systems; (5) creating a rich learning environment for students; and (6) ensuring that every Tech graduate will be information-literate.

CAMPUS COMPUTING

Installation and operation of the online information system is a collaborative project of the library, Information Systems Applications, and the Office of Computing Services. Information Systems Applications is responsible for applications programming; the Office of Computing Services for computer operations, systems programming, and networking; the library is responsible for the overall management of the program.

The campus network, GTNET, has been in place since 1979, and recently a fiber-optic network was installed. This network, which includes 128 buildings and supports approximately 2,500 ports, functions as a trilevel, hierarchical network with group, department/building, and campuswide levels. While the protocol of choice is TCP/IP, the network allows support of other protocol groups. Building baseband networks are bridged to the campus backbone. All devices on campus with network access also have access to the library's online information system. In the future, all PCs and workstations will have access to computers on other Georgia campuses through the statewide network PEACHNET.

The library has one departmental network and plans to install several more along with a building network. These local networks will facilitate work within specific departments and provide electronic mail and bulletin boards for the staff in the building that are currently run from a mainframe computer on campus. The building network will be bridged to the campus network for communications with other departments.

SELECTION OF A SEARCH SYSTEM

For several years the library had been investigating integrated library systems.

Miriam A. Drake is Dean/Director of Libraries, Georgia Institute of Technology, Atlanta.

While some of the modules of existing systems could do basic library housekeeping, none of the integrated systems had the searching capabilities needed by Tech students and faculty. The institution is an engineering-intensive campus where 65 percent of undergraduates and 60 percent of graduate students are enrolled in engineering programs. The campus also is computer-intensive: every student and faculty member has access to a terminal, PC, or workstation in the office, home, dormitory, or clusters on the campus. Many faculty members and students have their own PCs at home, and all devices have access to the campus network.

Given the campus environment and the library's goals, an effective retrieval system had higher priority than housekeeping. The idea of an integrated system was abandoned in favor of buying modules for each function. The library selected BRS/Search because it offered full text, full Boolean logic, modules to format data, messaging, and profiling. In addition, it contained accounting software and ability to handle IDs and passwords. BRS/Search is running on the campus IBM 4381.

SELECTION OF DATABASES

Between 1985 and 1988 the following databases were added: *Magazine Index*, *Newspaper Index*, *Management Contents*, *Computer Index*, *Trade and Industry Reports*, *INSPEC*, *Applied Science and Technology*, the Georgia State University Library catalog, and the full text of the *Commerce Business Daily*. The databases chosen provide both general and technical information. The addition of the Georgia State University Library was especially welcomed by the faculty and students, because Georgia Tech and Georgia State University have a reciprocal borrowing program whereby staff and students from each institution can borrow from either library. Georgia State University library has mounted the Georgia Tech database on their system. The collections of the two libraries are complementary.

All databases except *Commerce Business Daily* are updated weekly or monthly. It is anticipated that six months of *Commerce Business Daily*, which is updated on a daily basis, will be kept online. Library catalogs will be maintained in their entirety. Other databases will be limited to five years. Since most technical information requirements can be satisfied with five years of data, it is not cost effective to use valuable storage space for lesser used material. If a user needs to go back more than five years, the searches are done using commercial vendors.

SYSTEM OPERATION

When users sign onto the system they are presented with one menu containing a list of databases and the dates covered. Users search in command mode. The system is not menu driven. Since most students and faculty at Georgia Tech have a basic knowledge of mathematics and are comfortable with computers, the command level is appropriate. The command level gives the user greater flexibility and the ability to define a search more precisely. When the system was installed each student and faculty member was issued an ID, password, and user manual. New users and people using the system in the library can consult staff at a help desk in the library terminal cluster. Help desk and reference staff give individual instruction as needed. An online help manual is available on the system.

Library staff demonstrate and instruct people on how to use the system in workshops and classrooms. Plans are under way to offer formal courses in searching. There will be both beginning and advanced level searching for students and faculty. The courses will be offered several times during the first weeks of each academic quarter and on demand during the quarter. The library's classroom is equipped with PCs so that people can practice under supervision of the instructor.

The Online Information System is available twenty hours per day, seven days per week. There are users on the system when it goes down at 2 a.m. and users waiting when the system comes up at 6 a.m. During 1987–88 more than one million searches were performed by the system's 17,000 users. The number of searches performed during the first half of fiscal 1989 was 47 percent greater than for the same period in fiscal year 1988.

Approximately 30 percent of the searches are carried out at remote terminals or PCs outside the library. As more campus buildings expand their networking capabilities, the proportion of remote searching will increase. For many years the library has provided document delivery services for faculty and staff through the LENDS service. The LENDS truck stops at campus buildings twice a day to deliver and retrieve documents, books, and other communications. More recently, the library has been using Fax to deliver documents to Tech faculty and staff at off-campus sites, other libraries, and corporate clients around the world. Faculty may request LENDS service by electronic mail, telephone, or the message module in the Online Information System. During fiscal year 1988 LENDS delivered more than 12,000 documents to faculty and staff. Many Tech faculty are heavy users of the library but have not entered the building in years. As more buildings are connected to GTNET, the volume of LENDS requests is expected to increase.

INFORMATION AWARENESS

In 1985 the library received a grant from the W. M. Keck Foundation to design a program to increase information awareness and use. The grant was based on the premise that the success of a learning society in the information age will be determined by the ability of people to find, obtain, and integrate information and knowledge to enrich their personal and professional lives. Today's student will graduate into an increasingly complex, information-oriented world. The challenge is to prepare students for self-motivated, lifelong learning and education.

New three-hour credit courses have been developed for management and electrical engineering. The existing course in chemical literature was enhanced to provide greater access to online sources for chemical information. The management course stresses information for management and provides hands-on experience with internal and external bibliographic, full-text, and numeric databases. The course is designed so that students can use their information training and skills in other courses, such as marketing and finance. The course in electrical engineering includes modules on patents, specifications, standards, manufacturers' catalogs, legal resources, bibliographic databases, and other sources. Students use Tech Data and CASSIS online. Both courses include a module on copyright and information, such as airline fares and schedules, movie reviews, consumer information, etc., for personal use. The student response to the courses as reflected in course evaluations has been enthusiastic. Greater demand for more sections and for courses in other disciplines has been made by students.

SUCCESSES AND FAILURES

The Online Information System is a great success but this success has been achieved slowly. Students and staff do not use the system until they have specific need. Most students and faculty are not aware of the variety of databases available on the system. Most people know that the catalog is on the system. Some people know that other databases are there. Despite announcements on online bulletin boards, sign on messages, brochures in the library, and publicity in campus publications, users remain unaware of the presence of a database until someone points out that it is there. Recently, faculty have been informing students in the classroom about the availability of databases.

The library has been inundated with requests from people in business and government for access to the system. These requests have been denied because of database contracts and the need to have ports available for Georgia Tech people. Many outsiders have indicated that they want access only to the Georgia Tech catalog. In order to satisfy this demand the library is investigating the use of commercial vendors for this service.

A small percentage of students and faculty and a large percentage of library visitors have requested an easier to use query method and more help online. The searching courses to be offered by the library are expected to take care of student and faculty needs for training. No plans have been made to train visitors. Courses will be supplemented by individual training at the Help and Reference desks.

LIBRARY GOALS

The overall goals established by the library have been partially achieved. There is more information available on the campus network with more to come in the future. Students are no longer denied access to databases such as *INSPEC* and *Applied Science and Technology Index*. Searches requiring many paper volumes and person-hours of looking and copying now can be done in five or ten minutes from an office, dormitory, or PC cluster. Results can be downloaded, saved for later use, and/or used in bibliographies.

Faculty who previously read the *Commerce Business Daily* in print every day now can use the online version actively or passively. A faculty member can sign on and search as often as needed or can request automatic searching and delivery. Faculty profiles are run against the database each day, and the results are sent to the faculty member through the electronic mail system, so that daily reading is no longer necessary.

Faculty and student productivity is difficult to measure. Time savings alone represent significant savings. Both groups have reported that online searching has saved time, avoided redundancy, and resulted in cost avoidance. Several faculty and staff have reported that using *Computer Index* has resulted in better buying decisions for computer hardware and software.

Users also have reported some unexpected results. For example, students preparing for company interviews and plant trips often know more about the company and the industry than the recruiters because they have access to online databases.

Information resource productivity has increased dramatically. During the year prior to the local installation of *INSPEC*, the library staff performed less than 500 searches at a cost of $45,000. During the first six months of the local *INSPEC* service, more than 43,000 searches were performed by students and faculty. More library materials are being used by more people. The increased usage is reflected in higher photocopy volume, greater external circulation of books, and greater in-house use of journals. During the fall quarter, 1988, external circulation was 36 percent greater than in fall 1987. Use of current periodicals, as measured by pickups, increased 54 percent from fall 1987 to fall 1988.

The creation of personalized information systems is under way and is being pilot tested with the *Commerce Business Daily*. Each faculty member will be asked to renew and/or modify the profile each quarter. The program will be evaluated during 1989 through interviews with faculty, renewal rates, and tracking of proposal writing activity. The program also will be extended to other databases during the year.

The library has not succeeded in delivering all forms of information to PCs and workstations; however, several pilot systems involving graphics and numeric databases are in the planning stages and will be implemented during the next year.

The Online Information System has created expectations that cannot be fulfilled in the near future. Faculty regularly request that the full text of journals in their discipline be available on the system at no cost to them. They fail to realize that, at this time, the cost is prohibitive in terms of the subscription, software, storage, and computer time.

ELECTRONIC LIBRARY OF THE FUTURE

While more and more information and data will be available electronically in the future, paper will not disappear. Books, journals, conference proceedings, and technical reports will continue to be important items in libraries. They offer the advantages of portability and browsability. Works intended to be read cover to cover will remain in paper. Readers are not likely to want history, philosophy, drama, or poetry from a screen. Paper journals will be scanned, read on airplanes, and carried in backpacks and briefcases. Libraries may not retain back issues of journals in paper because the cost of binding, shelf space, and maintenance are becoming increasingly expensive. Publishers may offer back issues on optical disks or supply copies on demand through projects such as ADONIS. Publishers also may be willing to experiment with electronic journals in disciplines where the latest information is important. Librarians and publishers have recognized that they have mutual interests in the dissemination and preservation of information. Librarians, publishers, authors, and editors will work together to achieve a delivery system that satisfies reader demand and provides a reasonable rate of return on investment.

Some publications will disappear in paper because paper is no longer the format of choice. Abstracting and indexing services, directories, statistical material, and other items containing numeric data or information that is dynamic in nature will be published electronically. Numeric information on paper is expensive to use because of the cost of rekeying and human error. Directory information requiring frequent updating is better published electronically: people often receive incorrect information from paper directories that are outdated before they are published.

Users will have new capabilities at the PC or workstation. Windowing software will enable them to watch cable television, access online information systems, work on a spreadsheet, and write a paper simultaneously. They will access a greater variety of information and be able to manipulate it electronically as well as to "cut and paste" words, graphics, and pictures.

Networks will permit access to a variety of databases and information sources in different locations. Local communications systems will provide gateways, easing access to remote computers. This expanded access will create the need for accounting and payment systems, personalized database management systems, and ways of dealing with information glut.

ROLE OF LIBRARIANS

Librarians will be critical to the success of the electronic library. The librarian's role in the future will become more oriented to problem solving rather than collection building. Librarians will become more involved in teaching people how to learn. The creation of an environment for self motivated information seeking and learning will be a challenge in the electronic environment. In order to become more effective problem solvers, librarians will have to know more about the client or user and the context of an information request or need. With greater involvement, librarians may be in a better position to see the results of their work and experience a new form of satisfaction from their work.

Another challenge for librarians will be tailoring information systems to individual user's needs and ways of working. Now online systems are intended for general use. Some systems accommodate different levels of knowledge by offering novice, intermediate, and command level searching. The user must adapt to the system rather than the system adapting to the user. Since librarians are familiar with how different people approach information systems and ways different disciplines use information and data, they are uniquely qualified to design expert or other systems which will match the user's methods and needs.

Academic library users in the future are likely to arrive at the door or the electronic system being computer-literate but not information-literate. While data processing experts know about computing cycles and hardware, they cannot help users find information. Librarians are likely to take on the role of consultants in directing users to appropriate online sources for all types of information and data. Building on their knowledge of client needs and working methods, librarians can add substantial value to information finding and use.

While new roles emerge, the traditional responsibilities of preserving the records of humanity's achievements, failures, culture, history, and knowledge will become more important. Librarians have unprecedented opportunities to make a difference in the lives of people. We cannot discard or disregard the past. While we are helping people solve twenty-first-century problems, we must be prepared to make it easy for people to study the past.

Technology is making our jobs simultaneously easier and more complex. It will not solve all our problems. The tools created by technology do not provide service—people do. In the future our libraries will not be judged by the size of our collections or the number of databases resident on the local computer. They will be judged by the quality of our services and the value we provide for our clients and users. ■■

Development of the Carnegie Mellon Library Information System

Nancy Evans

Carnegie Mellon University was among the first institutions to offer an online information system that encompasses more than standard library catalog information. Library Information System (LIS) development began in 1985, and the first system version was made public in March 1986. This project provides a test bed for ideas, concepts, and problems currently at the forefront of librarianship. We are not only testing the feasibility of providing local information systems more comprehensive than an online catalog. In the area of management, we are addressing the convergence of campus computing services with libraries. In systems design, we are exploring both ways to handle disparate types of information with a single interface and ways to make this interface transparent to the user. Questions about approaches to collection development and public services are being discussed and explored. The complexities of information vendor participation, including reasonable pricing for and protection of copyrighted information resources, are still being defined. The model for local information systems is itself evolutionary; there are neither definitive answers nor a definitive model.

Since the introduction of LIS at Carnegie Mellon, several other academic and public libraries have either introduced local information systems of their own or have started active planning for such systems. Each institution will address questions in the areas listed above. And each institution's answers will ultimately be based on local opportunities and constraints. An earlier paper has already described LIS and pointed out aspects of system development and acceptance.[1] What follows here focuses on issues of innovative institutional organization and support and some project outcomes, both expected and unforeseen.

ENVIRONMENT

The reasons for developing LIS sound deceptively simple, belying the complexity of the project. The minicomputer used for LS/2000, Carnegie Mellon's first online catalog system, is too small to allow access to users outside the library. Users wanted access from offices, dormitories, and off-campus housing. The campus computing system being developed in the mid-1980s would allow users much simpler access to more powerful computing. For example, the multiple-window system would allow the user to search a database, create a personal bibliography, and work on a paper simultaneously, without complicated loggings on and off. Carnegie Mellon's reputation as a leader in computer science had attracted a user community that expects sophisticated, "cutting-edge" computing services. Providing widespread access to information and integrating these information resources with daily teaching and research activities would provide the community with the services it expects.

Nancy Evans is Data and Information Services Manager, Public Services, University Libraries, Carnegie Mellon University, Pittsburgh.

The concept of integrating a campus information utility with a campus computing system was not novel.[2] By 1985 Carnegie Mellon was uniquely positioned to develop and test such a utility. To begin with, the university has long been demonstrably committed to innovation in all aspects of computing. Faculty, staff, and students are accustomed to extensive and intensive uses of computing. In 1982 Carnegie Mellon and International Business Machines Corporation (IBM) entered a partnership to develop Andrew, an extensive campuswide network of powerful, UNIX-based personal workstations; a networking environment that allows these machines to share information and programs; and advanced software to support computing applications for teaching and research. The commitment to the Andrew project demonstrated the commitment of the university administration, from the president down, to computing research. It also demonstrated that the administration wanted a whole system—the "scholar's workstation," much discussed in the literature of higher education—not just new hardware or operating systems or applications programs. Within a year after the agreement with IBM, the university had also established the Center for Design of Educational Computing, part of whose mission has been to develop software tools for the Andrew environment. The Andrew concept put computer users in the center of the computing universe. They would no longer be isolated users, logging in from remote terminals to use time-shared mainframe computing. Rather, Andrew users would perform much computing on their local workstations, calling on resources located on other workstations or on larger computers used only as file servers, as these resources were needed. Access to information, whether bibliographic, numeric or statistical, campus news, or full-text resources, was an obvious choice to develop for the new campus system.

The university libraries' administration and staff have been equally committed to innovative applications of computing in information services. Library administration has been largely stable throughout this decade, allowing continuity in planning and executing projects. Like other Carnegie Mellon personnel, the libraries' staff have grown accustomed to using campus computing systems, both mainframe and desktop computers, for communication and basic office automation. Library public service staff have offered mediated online database searching since 1975 and end-user search systems since 1984, increasing both their own skills with online systems and the skills and appetites of users for these systems. The first online catalog system, LS/2000, has been in use since 1984 as well. Administrative support, staff expertise, and user demand, combined with the university's commitment to Andrew development, created an appropriate environment to develop LIS.

In several other ways Carnegie Mellon provided a rich environment for LIS development. The partnership with IBM produced equipment grants, including the IBM mainframe and software on which LIS is based; the new campus backbone network; and IBM personal computers. Grant equipment supplied by Apple, IBM, and other manufacturers is available to students in public facilities, and favorable pricing from IBM and Apple has enabled many students, faculty, and staff to buy their own personal computers. This installed base of equipment—the delivery system for LIS—ensured a large user population.

Technical support and expertise from outside the libraries was a crucial element in system development and planning. IBM personnel on campus assisted in some software development. More significantly, the university organization itself has changed in order to serve better both information and computing needs and to permit easier cooperation between relevant staff members. In the summer of 1986 the Division of Academic Services was formed, including the university libraries, Academic Computing, Computing Services, Data Communications, and other departments whose chief functions are to support the teaching and research missions of the university. Although most LIS development work was accomplished by library staff, chiefly in the Automation and Planning department, critical pieces of development and ongoing support have come from other departments within this division. For example, Computing Services has supplied programming support and machine operation for the IBM mainframe.

Financial support for the LIS project has come primarily from internal university resources. In the past, the libraries have benefited from Mellon Foundation's automation grants, which partially funded acquisition of the LS/2000 system and related computing equipment. With the advent of Andrew, the libraries, along with the rest of the university, have been the beneficiaries of IBM-funded campus network installation and equipment donations as well as assistance from IBM personnel. The university's recognition of the importance of LIS to campus, however, and of academic information service development elsewhere has been consistently translated into funding for this project, without a loss of funding for other library operations.

PLANNING AND DEVELOPMENT DECISIONS

Little more than six months elapsed between beginning of work on LIS and its introduction to the Carnegie Mellon community; eight months after its debut, several new files were added. (See appendix A for a description of the system, including technical information and a contents list.) The rapid pace of development was possible in part because hardware and software already existed on campus, so there was no need to spend time comparing hardware and software possibilities. This pace was necessary in order to have a working system ready to demonstrate at EDUCOM 1986, scheduled at Pittsburgh in November. Even with these two major components selected, several other major decisions remained. Some, such as user interface design and the handling of textual as well as bibliographic information, were essentially technical.[3] Others, such as system contents and how to prepare for public introduction, touched on more traditional functions of the public service staff in collection development and library instruction. This timetable required that many decisions had to be made without extensive study or comparison of alternatives, leaving more thorough evaluation until we had some experience with the existing system.

LIS development began when library administrators, particularly the director and the assistant director for automation, discovered that space was available on the IBM 3083 mainframe. Also, the university already had access to IBM's STAIRS retrieval software. As discussed above, library staff had been interested in making library catalog information more widely available on campus, particularly in providing an information resource to add to the Andrew system. The IBM 3083 could be used immediately as a stand-alone database storage and retrieval machine, and it had potential as a file server for the Andrew network. Although automation staff investigated at least one other retrieval system, the combination of an available machine and software and some campus technical support for both made the choice of the IBM and STAIRS a compelling one. The decision to use the IBM machine depended on the university's agreement to continue to support this machine and its operating system, even though computing systems staff had been trying to reduce the number of different machines and operating systems on campus. Without this agreement, work on the IBM could have been wasted effort.

Early in system development designers agreed on several general guidelines for the user interface. These included a commitment to use one interface for all files and to keep screens as similar as possible between files. In addition, the system was planned to be self-documenting, that is, whatever help the user might need in order to search would either be on the screen where the user took an action or received a system response, or in a one-screen long help file. These decisions were made based on reports in the user interface literature and on research and experience at Carnegie Mellon.

Selection of the IBM and STAIRS for system software placed additional constraints on system design. Some features available on commercial online systems, for example, the ability to store a search strategy for sequential use on several files, either were not technically possible or would have taken programming resources that were not available. In addition, extra programming work was necessary to get the IBM to work with the plethora of terminal and personal computer types on campus and to provide access through Andrew workstations and independent departmental networks of research machines. During the period of LIS development, several Carnegie Mellon departments installed their own networks which, while connected to the campus network, required special programming for access to the IBM machine. In some cases, particularly that of the Computer Science department, departmental programmers did the work.

Selection of system contents was based on the principle of providing a general information resource of broad interest to all the campus, with particular emphasis on serving the needs of the undergraduate population. A corollary principle has been to increase use of library collections. The library catalog was an obvious choice. The selection of five databases from Information Access Corporation (IAC) was based on several factors. IAC's end-user system, Search Helper, and later the optical disk-based Info Trac had been very popular resources for two years. The libraries also subscribe to IAC's microfilmed Business and Magazine Collections, and by making the IAC databases available we could improve document delivery. Two databases, *Magazine Index* and *National Newspaper Index*, fit the emphasis on general resources; the other three files served large but more specialized elements of the campus community. Moreover, IAC was the only database vendor at the time offering the sort of general databases we wanted. Grolier's *Academic American Encyclopedia* and, later, the *American Heritage Dictionary*, also fit the requirements of a basic reference collection. Several local files, particularly Arch-Pics (an index to architectural illustrations) and library-produced reference bibliographies, increase ease of access to library collections.

A set of major system planning and design decisions centered on how to ensure that users understood what they were using and how best to use it. LIS is intended to be used more often outside the library, and hence away from human assistance, than from a terminal near the reference desk. Like other end-user systems, LIS is supposed to put the user in control of his own information seeking. Public service staff at Carnegie Mellon share the reservations of their colleagues elsewhere about the quality of search results untrained users can obtain if left to their own resources. Moreover, many online system users are casual; that is, they do not use a system often enough to remember directions from one use till the next. The design goal was to try to eliminate the need for printed documentation or other help sources by making the system internally self-documenting as far as possible. We knew from earlier research at Carnegie Mellon that users tend to consult documentation only after they encounter problems.[4] We also believed that users' tolerance for reading long screens of text in online help files is severely limited. The option of an expert system, or at least a totally transparent interface, is not yet available, particularly since LIS is accessed through everything from sophisticated workstations to dumb terminals.

The challenge to designers, then, was to second-guess the kinds of assistance users might normally need at any step in the search process, from selecting a database to formulating a search strategy to interpreting results. Users need, in general, three sorts of help: assistance in mastering the mechanics of a system (for example, in knowing what character to use for truncation); assistance in understanding the process of searching; and knowledge of the contents of a given file. We attempted to put the right sort of help in the most likely spot in the search system, either directly on the search screen or in a one-screen-long help file. In addition, we tried to make each error message meaningful so users could understand and recover from mistakes. LIS is, by and large, a menu-driven system (although a Boolean search option is available). Prior to introducing LIS, we tested screens and help files informally, enlisting students to give some feedback. Two graduate students in the English department's technical writing program also tested the system, still in a limited but a more rigorous way. One of the greatest difficulties proved to be finding exactly the correct, unambiguous, meaning-laden terms and keeping use of these terms consistent throughout the system. The combination of menus, informative but not wordy screens and help files, and meaningful error messages does not answer all the questions and problems users encounter. Some more sophisticated users regret that LIS is not as flexible a system as Dialog or BRS. This attempt to bring the greatest good to the greatest number of users has on the whole satisfied the Carnegie Mellon community.

We did opt for some printed documentation, for those users who like to have something written to work with. The original help document included instructions for logging on from different types of equipment and some sample searches that demonstrated the power of the search system. More recently, reference staff have pared this document down to two pages describing the kinds of searches and how to do them. Separate one-page documents describe the contents of individual files.

IMPACT ON SYSTEM STAFF

Carnegie Mellon has been fortunate to have staff with the interest, ability, and time to devote to LIS development. The LIS development project, and ongoing system maintenance, have kept several library staff involved in varying degrees for over three years. Automation staff have all along had other duties, including maintaining the LS/2000 online system. No attempt has been made to measure the number of hours of staff time expended on LIS, or to compare these costs with the cost of buying an off-the-shelf retrieval system such as BRS Search. As with any system design project, predicting the time needed for system development would have been impossible. In playing the role of pioneer, we have encountered unanticipated problems that significantly increased workloads and eventually necessitated some staff shuffling.

One problem area concerns the leased bibliographic databases and full-text files included in LIS. Deciding how to handle disparate types of information while maintaining a single search interface and file format becomes increasingly difficult when many files from many sources must be made to fit the system. For example, different vendors have different fields in bibliographic records. IAC files are not in MARC format, so fields differ from the library catalog. Even though IAC maintains field names across its several databases, not all databases use all fields. Working from a list of all possible fields made design of the IAC files' displays more certain and more in keeping with the choice of fields for display in the catalog file. A sort of master range of fields, and limit and sort options, was developed by working with the catalog and IAC files so that subsequent additions can be worked into this standard format. The full-text files are handled with a somewhat different set of screen displays and search options. A full discussion of the technicalities of full-text display is, however, beyond the scope of this article.

A second technical problem arose with the vendor-supplied tape products themselves. Both of our original vendors, IAC and Grolier, were new to the business of supplying tapes directly to universities. Although both vendors were cooperative and lent some technical assistance, we encountered miscellaneous problems with both products. None of them were loaded into the system without some unexpected anomalies. For example, updating IAC files was difficult and time-consuming, particularly adding changes and corrections for data previously loaded.[5] From time to time tapes have been defective or records missing. These difficulties have increased the staff time necessary to keep the system updated and running.

IMPACT ON PUBLIC SERVICES

Although LIS has not been systematically evaluated since its introduction, library staff have formed observation-based opinions on the impact this system has for users and for staff. The campus community have become enthusiastic LIS users and, in some cases, active participants in further system development. At the same time, public service staff have discovered that some of their original reservations about the system are in fact quite legitimate concerns.

We have relied on informal methods to gauge the success of LIS design with the public. One method is frequency of use. Although we have a sort of transaction log, we can also tell that use has grown steadily, as indicated by random checks of the number of users logged on at different times. Two of the three Carnegie Mellon system libraries have both LIS and LS/2000 terminals publicly available, and reference staff have noted that users will wait for access to LIS even when LS/2000 terminals are available. (Part of this preference is due to the larger number of files available on LIS.)

The Carnegie Mellon community have not been shy about sharing their comments about the system. The computer mail address and campus extension of the Automation Office are given on the top-level search screen. Most comments have been favorable. Many users suggest improvements—additional files that would be useful, search system enhancements that would make searching easier. Others have independently developed simplified access protocols so that those using departmental rather than central computing can have access to LIS. Reasons for this level of acceptance are not subtle. Foremost is convenience: the system is available in an office or at home. However, the fact that LIS is easy to use without outside assistance must account for part of its continuing popularity. The online encyclopedia, at first a curiosity, has gained many dedicated users. Many users embrace the idea of a distributed information utility, even if they would prefer a different search system, user interface, or set of files.

It is hard to measure the impact LIS available has had, or will have, on the process of teaching, learning, and research. We suspect that the percentage of the community who use LIS and library resources, is larger than the percentage who used the card catalog of LS/2000. Indeed, comments from faculty indicate that those who previously would have sent an assistant to do their library work are now at least searching the library catalog on their own from their offices. Use of the IAC microfilm collections mentioned above has increased substantially, as measured by the large increase in microfilm copying. For example, between the fall 1987 semester and fall 1988, microfilm copying increased 43 percent.

The availability of LIS may have contributed to a decrease in use of some traditional online search services. Since users can now do their own searching with no charge, use of mediated search systems like Dialog has decreased. The libraries, as noted above, have offered end-user searching for over four years; use of some of these systems has decreased in the past year due, at least in part, to revised policy limiting the number of no-charge searches allowed. At the same time demand to increase the number and variety of LIS databases has increased.

Reference staff describe several effects that LIS has had for them. One effect is on workload. Since library use in general has increased, the public service desks have experienced a concomitant rise in demand for service. Part of this increase is attributed to an increase in size of both the faculty and the student body, but part seems attributable to the percentage increase in library users. The types of questions answered during a tour of duty at the reference desk has changed somewhat. One reference librarian noted that the workload at the desk is heavier in part because reference staff have also become equipment technicians, feeding paper into printers and trying to troubleshoot problems. Carnegie Mellon librarians share this particular class of problems with colleagues elsewhere; one problem seems to be that the quiet ink-jet printers preferred for public service locations break down more often than noisy impact printers. Users are asking fewer questions about the online systems than they did about printed indexes; LIS does indeed seem to be self-service.

The self-service aspect of LIS has proven the validity of some concerns reference staff voiced before the system was introduced, however. A major problem is that users do not understand what they are searching when they select, say, *Magazine Index*. The user typically assumes that, if she found a reference by using the computer, the library owns the source document. Moreover, many users do not know that all the bibliographic files except the catalog contain limited years of indexing. As one librarian noted, the "smart" users will ask how to get indexes to earlier years of a category of journals. Another assumption is that if a search strategy fails the item in question does not exist or, conversely, if a strategy succeeds, the searcher has found all possible information on her topic. These concerns are not limited to LIS and indeed are not that much different from concerns public service staff have held for years. Libraries are, after all, complicated systems to operate. The difference now is that, with a system that gives users success, or the illusion thereof, the user may be less likely to seek help when she needs it. System design has not yet solved this class of problem.

User education, particularly formal library instruction sessions, is the traditional way to attack the problem of lack of library knowledge. Reference staff now include an introduction to LIS in basic library instruction. In addition, special all-comers sessions on searching were developed, scheduled, and publicized when the system was introduced. Attendance at these special sessions was predictably low, reflecting similar experience earlier with special sessions on LS/2000, so these have been dropped. The preferred approach to teaching search skills is to teach concepts, a method that has received favorable attention in the literature and in practice. Local experience shows, however, that, in a forty-five-minute session, most users can learn only basic facts about system contents and some moving-around skills. An introduction to LIS is included in the basic computing course required of Carnegie Mellon freshmen, and this introduction is kept simple.

LIS, and the introduction of end-user search systems in general, has had an interesting effect on the perception users have of the reference librarian. One reference librarian characterized this perception as both a technician—the person who can fix the printer—and as an expert with sophisticated systems. Another librarian agrees that systems such as LIS give reference staff more credibility with the public as information experts. The mystique of computing may account for some of this image enhancement. At a school like Carnegie Mellon, sophisticated computing is taken as a given. Students and faculty in fields not normally considered to be computing-intensive are, in fact, involved, knowledgeable, and appreciative of the benefits they derive from state-of-the-art information tools. The reference librarian's role has not become deprofessionalized. Rather, the librarian is playing to a wider audience than before and with perhaps greater acceptance as a professional performing useful, specialized services.

ISSUES FOR THE FUTURE

No online system is static. Since its introduction LIS has retained the same basic form, with some features and files added. Some user-suggested improvements and new files, particularly the *American Heritage Dictionary*, have been added. In the next two or three years, we expect to continue development of the current system, primarily by adding additional LIS files. We are also developing a very different system, not based on the IBM mainframe, which will be better integrated with the capabilities of the Andrew system and its workstations. The IBM machine is not part of the university's longer-range computing plans, which include reducing the number of supported operating systems on campus.

A major concern involves planning what information to deliver and how best to deliver this information. In traditional library terms, this would be collection development: what resources will take up what part of the collection budget. For electronic resources, the question may be even more complicated than for print resources. Many potential LIS files are also available on CD-ROM or through remote end-user systems like EasyNet or WilSearch. The challenge is to arrive at a reasonable and equitable selection of information sources available at no additional charge to users. Defining "reasonable" and "equitable" is no simple task. Since electronic resources in general, and LIS in particular, are highly visible, more users are aware of what resources are available for their areas of interest. A system that appears to cater more to one area than another raises equity issues that are difficult to resolve. For example, we could choose to focus on adding bibliographic data in the sciences, since a plurality of Carnegie Mellon's users are in engineering, science, and computer science. Faculty in the humanities may well feel ill-served, not just because their areas are not included, but also because they tend to have fewer funds to pay for mediated online searching. We also are considering questions of retention of print sources or duplicate purchase of print and online formats. A group of public service librarians have developed a collection policy that attempts to resolve some of these issues.

The prices of commercial information products are one constraint on the collection we hope to build. The issues of pricing of and protection for these sources are related. When vendors lease their products they are understandably concerned about the number of the potential users and how good system protection is. The vendor deserves protection both from unauthorized users and from users who attempt to resell the information. Often, however, the price structure for tape products puts them out of reach for everyone or greatly restricts the variety of tapes a library can afford. Or the vendor imposes conditions that are difficult or impossible to fulfill, such as monitoring the number of searches on the vendor's database and paying a per-search fee. Working out an equitable solution to pricing and security dilemmas is an important agenda item not only for Carnegie Mellon but also for other institutions who are leasing commercial products.

LIS has created a revolution of rising expectations among the Carnegie Mellon community. Easy access to a wide range of information has stimulated demand from all parts of campus, from steady library users to students and faculty who may never have used our information resources before. Now, many disciplines want local bibliographic access to their own, specialized literatures. Users suggest adding full-text reference sources, statistical data, literary texts—far beyond our current ability to purchase and store them. And, as more users see the potential for integrating information with the power of the Andrew system, more users are urging development of a more sophisticated interface and retrieval system. We cannot deliver all of this instantaneously. The interest and support of the user community, however, justify both the efforts to-date and system expansion and redesign in the next three years.

ACKNOWLEDGMENT

The author acknowledges her debt to many colleagues, without whom this paper could not be written. In particular, Mark Kibbey, director of Library Automation, and Gregory Diskin, systems designer, produced the system discussed here and have helped with the current manuscript. Dorothea Thompson and Susan Collins, reference librarians, shared their thoughts on public service effects. Several other authors of papers in this issue have provided insights and ideas. Joan Mitchell, director, Instructional Technology, read the manuscript and provided both information and comments. Thomas Michalak, director at Carnegie Mellon, made the whole enterprise possible. All errors of omission and commission are my own.

REFERENCES AND NOTES

1. Nancy Evans and Thomas J. Michalak, "Delivering Reference Information through a Campus Network: Carnegie Mellon's Library Information System," *Reference Services Review* 15:7-15 (Winter 1989). Readers who have not read the *RSR* article are referred to Appendix A for a description of LIS.
2. Earlier discussions include William Y. Arms, "Scholarly Information," *EDUCOM Bulletin* (Fall/Winter 1983), p.21–23.
3. For a more detailed description of interface design, see Nancy Evans, "Distributed Retrieval System for a University Library," *Proceedings of ASIS Mid-Year Meeting, 1988*, forthcoming; and Evans and Michalak.
4. Peggy Seiden and Pat Sullivan, "Learning an Online Catalog: Modeling the New User," paper presented at American Educational Research Associates Conference, New Orleans, April 1984.
5. Mark Kibbey and Nancy Evans, "The Library Information System: Carnegie Mellon University and Information-Access Corporation," *RQ*, 28:104–8 (Fall 1988).

APPENDIX A. SYSTEM DESCRIPTION

Technical: The Library Information System is based on a university-owned IBM 3083 mainframe computer, running under the VM/CMS operating system and using IBM's STAIRS retrieval software as the search engine. Users can access the system through most terminals in current use on campus and through microcomputers connected to the campus network. Users can also dial into the system from off-campus residences. Current access is not password-controlled, nor is it necessary for a user to have a Carnegie Mellon computer account.

Contents: When LIS was introduced in March 1986, the library catalog was the only file available. In November of that year, one database from Information Access Corporation (*Magazine Index*), Grolier Electronic Publishing's *Academic American Encyclopedia*, and several local files, including library bibliographies, were added. Four additional IAC files (*National Newspaper Index, Management Contents, Computer Database*, and *Trade and Industry Index*) have been added, followed by Houghton Mifflin's *American Heritage Dictionary*, the Carnegie Mellon libraries' serials holdings list, a locally-produced index to twentieth-century architectural illustrations, and a guide to statistical data tapes.

Files are updated on varying schedules. The library catalog is updated weekly. All IAC files are updated monthly except for *Computer Contents*, which has a biweekly update. Local files are updated as needed, generally once or twice a year.

How it works: The STAIRS retrieval system supports full Boolean search capabilities, sorting and limiting of search results, and searching specific record fields. In addition, the LIS indexes every word in a record, including numerical strings, such as call numbers, which are treated as "words." IBM 3083 response time is sufficiently fast to make full-text indexing feasible; response time varies depending on file size and system load.

System designers combined menus for file selection, limits, and sorts; single-character commands for moving around the system; and user-input search terms (see figure 1). After selecting a file, the user goes immediately to a top-level search screen, which allows him to enter up to four keywords from any field in a record (see figure 2). The system assumes the Boolean AND between these terms and performs the search. Search results are displayed in groups of ten short-title listings from which the user may display a full record (see figure 3), limit or sort search results by a number of file-specific parameters, or start a new search. Limits and sorts are chosen from a menu (see figure 4). At each step of the search process the user can read a one-screen help message describing options and processes in more detail (see figure 5). In addition to the default "and" search, the system supports more sophisticated searches incorporating the full range of Boolean operators and field limits (see figure 6). The user can choose this option at the top-level search screen. ■■

```
                    Carnegie Mellon Library Information System

FILE NUMBER OR COMMAND:                    1 Feb 1989    -- 13 Files

FILE  NAME               DESCRIPTION
--------------------------------------------------------------------------
1     Library Catalog    Current holdings of CMU Libraries

1A    Journal List       Journal titles owned by CMU Libraries

2     Magazine Index     Index to popular magazines, 1982-

3     Newspaper Index    Index to primary national newspapers, 1985-

4     Encyclopedia       Academic American Encyclopedia, full text

5     Computer Database  Index to computer magazines, 1986-

6     Announcements      CMU Library hours and information
--------------------------------------------------------------------------
COMMAND MENU:
? = HELP       ?n = HELPn, where n = file number       <RETURN> = MORE
Q = Quit
```

Fig. 1. Top-Level Menu.

```
                    CMU Library Catalog

TYPE in up to 4 search keywords OR a Command character.  Then press RETURN.

_



ADVICE:
1)  Use major keywords and omit short words like 'and', 'the', 'of',
    'his', 'for', etc.
2)  Use $ for truncation.  (school$ = school, schools, schooling ...)
3)  To enter a more complex search, choose option 3.
4)  Always press RETURN after entering data or a menu character.
5)  The DELETE key operates on the character UNDER the cursor.  To correct
    typos, move the cursor backwards with the back-arrow key and
    type over the error.
6)  Send comments and bug reports to mk72 tb or call x3692.
---------------------------------------------------------------------------
COMMAND MENU:
? = HELP               1 = Scan Index         4 = Author Search
Q = QUIT and logoff    2 = Change Database
                       3 = Command Search
```

Fig. 2. Top-Level Search Screen.

```
                    CMU Library Catalog

 38   Hits on  SOUTH AND CAROLINA


ENTER A RECORD NUMBER OR A COMMAND:

REC  DATE  AUTHOR                 TITLE                        NUMBER
  1  1952  TINDALL, GEORGE BROW   SOUTH CAROLINA NEGROES, 1    325.26 T58S C.1
  2  1986  FINCHER, JULIAN H.     DICTIONARY OF PHARMACY /[    RS51 .D55 C.1
  3  1985  CONFERENCE ON COATIN   COATINGS AND BIMETALLICS     671.73 C74C 1984  C
  4  1986  WORKSHOP ON MOTION--   PROCEEDINGS / WORKSHOP ON    621.367 W926P 1986
  5  1987  BELL, MALCOLM, 1913-   MAJOR BUTLER'S LEGACY : F    920.075 B43M  C.1
  6  1987  ADAMS, EDWARD C. L.    TALES OF THE CONGAREE / B    813 A21CA  C.1
  7  1986  U.S. NATIONAL CONFER   PROCEEDINGS OF THE THIRD     624.176 U583P 1986
  8  1973  CHANDLER, RAYMOND, 1   CHANDLER BEFORE MARLOWE;R    811 C45C  C.1
  9  1950  COTGRAVE, RANDLE, FL   DICTIONARIE OF THE FRENCH    443.2 C84D C.1
 10  1974  INTERNATIONAL BIOMAT   PROSTHESES AND TISSUE, TH    617.47 I61P V.1 C.1

--------------------------------------------------------------------------------
COMMAND MENU:
L = LIMIT this search      B = BACK up 10 Titles    S = SORT the results
R = RESTART a new SEARCH   ? = HELP                 RETURN = Next 10 Documents
Q = QUIT and logoff        U = Undo last Limit      H = Search History
```

Fig. 3. Short-Title Display.

```
LIMIT               CMU Library Catalog

298   Hits on  SOUTH AND CAROLINA

LIMIT NUMBER or COMMAND:



LIMITS:
    1 = YEAR of publication
    2 = MEDIA TYPE
    3 = LANGUAGE
    4 = LOCATION
    5 = KEYWORD


--------------------------------------------------------------------------------
COMMAND MENU:
? = HELP                R = Restart a new SEARCH          C = CANCEL
```

Fig. 4. Limit or Sort Screen.

```
                    CMU Information System Help
LIMITS
You can choose to LIMIT your search results by DATE of publication,
LANGUAGE, MEDIA, LOCATION in the CMU Libraries, or by adding a KEYWORD.
You should use these options to eliminate records from the retrieved
group and thereby reduce the results of your search to a manageable number.
Separate Help Screens describe each specific LIMIT search; choose the
type of limit you want and press ?.

NOTE:
1) If the system is slow performing your LIMIT search please be patient.
2) After you have LIMITed your search, you cannot go back to your original
   search unless you start over.
3) Sometimes the search system runs out of room to store a LIMITed set
   of records. If this happens you will see a message that you have
   exceeded the line limit. The best solution is to go back to the
   top search screen and add an additional keyword.
4) For more help you can consult a Reference Librarian. Call Hunt
   Reference at x2442 or Engineering and Science Reference at x2427.
----------------------------------------------------------------------------
 Press RETURN to continue
```

Fig. 5. Help Screen Attached to Limit or Sort Screen Chosen.

```
                         CMU Library Catalog
COMMAND SEARCH

..SEARCH _


ADVICE:
1) Enter keywords separated by operators. The operators are defined:
   OR    means EITHER of the keywords can be present.
   AND   means BOTH keywords must be in the same document, in any field.
   WITH  means BOTH keywords must be present in the same field.
   NEARx means the two words must be found in the same field within x
         words. x may have values between 1 and 7.
   ADJ   means two words must be ADJACENT to each other, in the order given.
   NOT   means the word following may NOT be present. NOT RECOMMENDED.

2) Common articles and prepositions may not be searched.
3) $ is used for truncation. (school$ = school, schools, schooling, etc.)
   Do not shorten a word too much or you will get poor results.

----------------------------------------------------------------------------
COMMAND MENU:
? = HELP            R = RESTART a new SEARCH      Q = QUIT and LOGOFF
```

Fig. 6. Command Search Screen.

Article-Level Access in the Online Catalog at Vanderbilt University

Flo Wilson

In February 1989, Vanderbilt University Library introduced a locally mounted MEDLINE file into its NOTIS-based online catalog. Planning and programming efforts are under way to add several H.W. Wilson databases by summer 1989. Considerations related to decisions about databases and search engine selection are described; key factors in the implementation process are presented.

The concept of a scholar's workstation is being seriously considered by libraries. A single microcomputer in an office, a dorm, or at home can serve as the mechanism for searching the libraries' catalogs, browsing databases (bibliographic, full-text, or numeric in nature), downloading and manipulating the data thus identified, and communicating with others about research efforts. This scholar's vision can only become a reality with the assistance of information professionals. Academic libraries are in an enviable position to continue the work on development of the information infrastructure. The creation of online public-access catalogs and the installation of intra- and inter-campus networks have made remote access to the library's traditional local bibliographic tool possible. The addition of local databases—different from the traditional catalog—and the implementation of gateways to external systems will continue to broaden access to the information landscape.

Over the last few years, libraries with the resources to explore locally mounted databases began some interesting and illuminating efforts in this area—the first steps toward broadening access beyond the catalog. As we explore the approaches taken, the databases selected, the difficulties encountered, and most importantly, the successes and failures of both systems and users, we will hopefully understand how users interact with some libraries' information systems.

Our focus here will be on the efforts undertaken by Vanderbilt University Library to extend the notion of the online catalog to encompass other databases. Over the last three years we explored a number of approaches; we just introduced a MEDLINE file for library users and expect to add several H.W. Wilson databases within the next few months. In this article I will identify our approaches to database selection and the choice of both hardware and software platforms. I will share our experiences through the development and planning process and anticipate the work left to be done in evaluating our successes to date.

VANDERBILT'S ENHANCED INFORMATION ACCESS PROJECT

Acorn is Vanderbilt's online public access catalog which serves the eight libraries of the Jean and Alexander Heard Library.

Flo Wilson is Assistant Director of the Library for Systems, Vanderbilt University Library, Nashville, Tennessee.

Based on the NOTIS software, **Acorn** was successfully introduced to the Vanderbilt community in September 1985. Some forty-five public terminals offered access to 550,000 titles (an estimated 60 percent of the collections) for approximately 8,600 students. In fall of 1985 plans were under way for the full implementation of all NOTIS functions and for completion of a retrospective conversion project; enhancements from NOTIS, such as keyword searching in the catalog, were anticipated.

The library's administration began anticipating future efforts and projects. Remote access was an obvious goal, and the university started planning for a campus-wide broadband network. Optical storage was at the time in the early stages of development for library databases. The notion that databases might be searched more cost-effectively and more successfully by end-users when mounted locally, rather than through the use of remote database services, seemed worth exploring. To experiment with our plans to enhance information access, we sought support from the Pew Charitable Trusts.

In early 1986 we were awarded a $750,000 grant from Pew; we received additional university support in the amount of $200,000 for the experimental project. The Enhanced Information Access Project, as it came to be called, set several different goals:

1. To extend the campus broadband network to several of the centrally located libraries;
2. To provide article-level bibliographic access electronically for library users in the library and across campus through the network;
3. To experiment with new and alternative technological developments related to article-level indices and full-text storage; and
4. To determine user success with, preferences for, and institutional costs of different formats, different user interfaces, and different coverages.

The project was originally funded for two years; the period was extended by nine months and reached funding completion in January 1989. An automation project librarian was hired in December 1986 to coordinate the grant efforts.

Early in the project, the network was extended to the libraries, accomplishing goal number one. It is goal number two, mounting databases locally on the mainframe, which will be discussed in this paper. We completed the introduction of the first database in February, and our evaluation is now under way. As a step toward goal number three, we made a significant commitment to CD-ROM databases, with over forty different products and twenty-four workstations in place. To date, users have expressed a strong preference for these electronic databases, a preference we expect to carry over to our mainframe-based indexes. The questions we hope to address in goal number four will await the results of the evaluations of both CD-ROM and mainframe technology.

DATABASE SELECTION QUESTIONS

After receiving the grant award and as a part of the initial planning process, a meeting was held with the library's division directors. A free-ranging discussion addressed the question of databases to be considered for acquisition. Consideration of availability was certainly a part of the discussion, but a number of databases were mentioned and added to the potential "wishlist" even if they were not known to be available in machine-readable form. This initial meeting permitted an overall view of the libraries' interests as we continued exploring feasible options. Since CD-ROM databases would be added as well as mainframe ones, it was important to identify all the possibilities.

Several criteria were established for selecting the databases that we would seek to mount locally on the mainframe for network as well as library access. These factors included availability, appropriateness, level of general interest and expected use, and the existence of the database in other formats. A particularly important element in the equation was the likelihood of ongoing funding support after grant support ceased. In general, high-use databases with interest and use beyond a single library would be appropriate for mainframe application; more specialized databases would be better candidates for CD-ROM implementation. In some cases we expected to make available both mainframe and CD-ROM based versions of the same database in order to test user preferences; this was seen as a legitimate duplication for research purposes.

MEDLINE

Our first choice seemed an obvious one. The single most heavily used database within the library system was MEDLINE, produced by the National Library of Medicine (NLM). In addition to use by library staff in mediated searching for library users, a large number of Medical Center staff used MEDLINE through their own accounts. MEDLINE was readily available as a tape subscription with already established distribution mechanisms. Other libraries' experiences with MEDLINE might be shared, since a number of medical libraries had already implemented mini-MEDLINE-type services within their online catalog environments. The final persuasive argument in favor of MEDLINE was the Medical Center's $100,000 contribution to the project.

Along with the selection of a database comes the decision relating to desired coverage. The MEDLINE database is huge, with close to 300,000 records added each year. We projected that we might be able to support two to three years of the MEDLINE file with abstracts. Our approach would be to have two years and the current year, rebuilding the file each year when the previous year's correction tapes were received. We discussed a number of alternatives with the staff of the Medical Center Library. If we were to exclude abstracts or exclude some journal titles, we would be able to support a longer time period. However, the librarians' knowledge of the use of the file led them to argue strongly that a fair test against the remote MEDLINE file required the full database. It was also felt that two to three years would address most immediate information needs.

The issue of coverage for MEDLINE arose again last fall when it became apparent that even with the additional disk storage which had been purchased we would be unable to support more than two years. The decision was to stay with the full file for a shorter period of time during the evaluation stage so that we might adequately assess users' preferences. Because of the redefinition of the current MEDLINE file by NLM, our present version contains all records with a 1988 publication date and all entry months from January 1989 to the present.

H.W. Wilson Databases

Even before testing the MEDLINE database, we looked ahead toward the next database to be added to our local system; the choice this time was not so obvious. Since MEDLINE covered a relatively narrow subject range, one of our goals in the second choice was to provide a database which would be of use to the rest of the libraries. To solicit broad input from the library staff, we distributed questionnaires to all reference and information services librarians. They were asked to identify databases they would like to see made available in the online catalog environment. Of the thirty-four responses, nineteen staff ranked *Readers' Guide to Periodical Literature* or another H.W. Wilson database as highest priority. Because of the variety of Wilson databases, we concluded that several could be made available and provide the broadest coverage. Wilson had not yet marketed its databases as tape subscriptions, but they were receptive to the idea. Thus, we began the discussion and negotiation process for acquiring *General Science Index, Social Sciences Index*, and *Humanities Index*.

Since we started working with the huge MEDLINE file, the Wilson indexes seemed small by comparison. We expect the full run of the three indexes to require far less space than our one- to two-year MEDLINE file. We decided to load whatever is in machine-readable form—1984 to present for *Humanities* and *General Science* and 1983 to present for *Social Sciences*. We posited that a combined file of these three indexes, covered by a single-search statement, might be most useful to searchers. Wilson agreed that we might try this approach if we would clearly label the individual records with the index name from which it was derived.

Choice of the Wilson indexes also provides a logical growth path for additional indexes to be added at a later date. Additional disciplinary coverage is possible with little incremental development work.

Databases Not Selected

Two other databases were given serious consideration in the decision-making process—*Psychological Abstracts* and the *Academic American Encyclopedia*. *Psychological Abstracts* was the most frequently mentioned choice by librarians, followed by the Wilson databases; it is an option which we will want to consider once we move beyond our immediate implementation plans.

One of the more interesting databases readily available was Grolier's *Academic American Encyclopedia*. Since this was also different in nature—full text rather than bibliographic—it seemed like an interesting experiment. We did not proceed with this possibility, however, because we could not generate any enthusiasm from library staff or university faculty for placing it online. Since library users were not accustomed to searching Acorn remotely and because there was little experience in using full-text electronic tools, the utility of this tool did not capture our imagination. In addition, our emphasis was on research tools, and a general encyclopedia did not adequately fit that criteria.

As we eventually move to provide more reference sources and as other libraries report on their successes with encyclopedias, we might add such a database.

Other databases which were discussed or suggested but which did not meet our criteria of availability, broad interest, heavy use, and/or supportability were *CIS Index*, *Current Contents*, *Magazine Index*, *PAIS*, *ERIC*, *The New York Times*, *Newspaper Index*, and *NTIS*.

SOFTWARE QUESTIONS

When we began planning for how our article-level databases would be presented to the public, we made an assumption that it would be desirable, from a user's perspective, to have the interface look as much like the online catalog as possible. From a single terminal using the same set of commands, a user could have easy access to Acorn and an easy method for changing to the article-level files. We envisioned a menu listing the database choices. A longer-range goal of incorporating holdings status in the article-level file would be simplified if the same hardware and software were used for both types of files.

The philosophy in implementing systems at Vanderbilt was to take advantage of development work done by others—that is, to be early implementers of innovative systems. Involvement in the specification, design, testing, and implementation of new products is consistent with this philosophy, but we are anxious for others with more expertise to assume the responsibility for the actual development and programming.

Consequently, in the summer of 1986, we began discussions with NOTIS Systems to elicit their interest in our project. Using the NOTIS software for our new project would ensure consistency in the user interface. Since Acorn had performed very well as a catalog and since keyword enhancements were scheduled for introduction within the year, all the basic search elements seemed to be present. Traditional catalog approaches of author, title, and subject searching could be provided as well as the more flexible, but slower, keyword approach. An existing patron file for circulation purposes would define eligible users and provide the foundation for accounting information. The development work required would be conversion of the databases, creation of a switching mechanism for moving from one database to another, and a sign-on function for users to provide security and accounting. NOTIS expressed great interest in this project and indicated that this was a development effort they would be willing to undertake.

For purposes of developing keyword search capability in the online catalog, NOTIS had selected the BRS/Search software as the underlying search engine. Fully integrated into the LUIS (NOTIS public catalog) interface, the search engine offered full keyword search capability with truncation, field qualification, and logical operators. Search results were presented in an index display in reverse chronological order (or reverse record order when the publication data was the same). While license restrictions with BRS did not permit use of their Search software with databases other than the catalog, NOTIS was able to arrange an exception for purposes of this joint Vanderbilt/NOTIS research project.

Vanderbilt contracted with NOTIS to provide a multiple database product for our research effort. This product would provide a database selection menu to library users, one step back from the traditional catalog. It would allow a variety of databases to be mounted locally. Each would have different "help" screens and record display formats; each could have a different authorization level in terms of which terminal and patron groups were allowed to search the database. It would provide for requiring patron sign-on. In order to collect data for evaluation purposes, the transaction logs and statistics available for LUIS would be expanded to include the multiple databases. MEDLINE would be the first database provided; additional databases would be selected at a future date. A general sense of direction was also discussed, with Vanderbilt suggesting the need for the incorporation of holdings information in the article-level database, patron sign-on requirements, an accounting function, and a gateway to other network or remote resources. All of these were seen as future NOTIS enhancements to the multi-database effort.

With this general agreement on the software to be used for the article-level files, many of our hardware decisions were also defined. An upgrade of our IBM4361 and operating system was completed, and significant disk storage was added. Network access was provided to support remote and dial-up use of the enhanced system.

Alternative Approaches

The other major software approach considered was the use of the BRS/Search software directly, rather than using the keyword capability through NOTIS. There was one major reason for not pursuing this alternative—the NOTIS option offered us integration with the library catalog at a level that would be impossible to achieve with BRS. With BRS, use of the article-level files would have had to occur on a different machine or under a different operating system, since a VSE version of BRS does not exist. In addition, even if the switching to a different operating system or machine was relatively transparent, the user interface would still be quite different from the online catalog.

Our decision to use NOTIS as the vehicle for providing our first article-level databases did not rule out the possibility of using other hardware and software to expand access to information resources on campus. We can foresee a time, in the near future, when there are services available on the university's VAX8800s (the current academic computing environment) which are appropriate for offering to library users. Examples include a recently cataloged titles list, library mailboxes on the university's electronic mail system, Hypercard stacks for a couple of Macintosh-based tools under development. Our goal will be to develop microcomputer-based software which can be distributed to users in the campus community, allowing them to negotiate the network transparently. Many of our current Acorn public terminals are microcomputers; we will continue to add microcomputers as terminals to allow use of future software capabilities. Many of our terminals also connect to Acorn through the university's broadband systems, supporting both DECNET and TCP/IP, making access to non-IBM based resources more readily available from a single workstation.

THE IMPLEMENTATION PROCESS

As other libraries consider adding article-level databases to their online catalogs, it may be useful to share the experiences of those who have made initial attempts. These are some of the steps in the implementation process and the conclusions we have reached to date.

Hardware Requirements

One of our first efforts under the grant was to upgrade the hardware used for the library system. We had purchased an IBM4361 Model 4 with 4 mg. of memory for our NOTIS implementation. Four IBM3370 disk drives, 730 mg. of storage each, supported the library's bibliographic database of 750,000 records. The operating system was SSX/VSE. Our grant-related upgrade brought us up to an IBM4361 Model 5 with 16 mg. of memory. We changed operating systems to VSE/SP. We added four additional IBM3370s disk drives.

To provide for normal library database growth, increased storage requirements for catalog keyword indexing, and larger-than-anticipated MEDLINE file sizes, we found it necessary to purchase even more storage. Two IBM3380s with 2.5 GB storage each have been added, and now we should be able to support one to two years of MEDLINE and at least some of the Wilson indexes. Based on actual experience, one year's worth of MEDLINE has the following storage requirements:

Database (663,380 records)	770 MB
Author/title indexes	192 MB
Subject indexes	508 MB
Keyword indexes	859 MB
Total	2,329 MB

A related storage requirement previously under-anticipated was sort space. Creation of the keyword indexes for one year of MEDLINE requires a minimum of 3.8 GB. of sort work space. We hope to have similar figures for the Wilson files by summer 1989.

Another upgrade which we anticipate but cannot yet document is the need for a more powerful processor. We expect that the number of users on the system (once we're providing a number of services and supporting the type of processing that will be required for volumes of keyword searching) will require a larger machine than we currently provide. A faster machine would also permit faster indexing than is possible now. Generation of the entire set of indexes for the one year of MEDLINE requires forty hours of batch processing with our present hardware configuration.

A few terminals have been added to the Medical and Science libraries in anticipation of heavier use once MEDLINE is available; we now have sixty public catalog terminals/ports available. We are uncertain as to what the additional terminal requirements will be, but we do expect significant and increasing demand over the campus network, implying the need for more ports on the network rather than hard-wired terminals. We will use data generated from the use-statistics reports to assist in expanding access in appropriate locations.

Database Conversion

To expedite design work on the MEDLINE database, a decision was made to convert the MEDLINE records to a pseudo-MARC format. Since the public catalog software was designed to work with MARC records, many of the problems of indexing and display would be simplified if other databases could be approached in the same way. Vanderbilt library staff defined the mapping of NLM fields into MARC tags, taking into consideration the standard NOTIS/LUIS indexing and display programs; wherever possible, the mapping was consistent with appropriate MARC tag usage. A sample of a MEDLINE record converted into a NOTIS MARC record is shown in figure 1. A similar approach was taken with the records from the H.W. Wilson databases. Wilson provided detailed documentation on the use of the various fields in their records; we determined how to map those into our modified MARC format. NOTIS then prepared record conversion and load programs to handle the records as they were received from the vendors. Ongoing review and revision of the conversion programs will be required to handle changes made to the databases by the vendors.

Screen Design

A database selection screen was created as the first screen seen by the user. All possible choices of databases available from that terminal are listed on the selection screen using four-character database labels (figure 2). Entry into any of these labels will result in the display of the introductory screen for that particular database; figures 3 and 4 show examples of the first MEDLINE and Wilson screens. A NEWS screen and help screens specific to each database are provided.

Consistent with our decision to use the basic LUIS interface as the foundation for the multiple database project, the basic screen structure was already in place. The sets of screens to be used for MEDLINE and Wilson databases were rewritten to provide appropriate levels of help for databases quite different from the catalog. Additional screens were added to the MEDLINE keyword help-screen series to explain such concepts as check tags, registry numbers, and subheading abbreviations.

A significant design change was made to the bibliographic record display. Since many of the elements of the database citations were not expected to be as obvious as in the online catalog record, we opted for a labeled display for the article-level files. Figure 5 is the public display for the MEDLINE record that is shown in its MARC format in figure 1.

Schedule for Implementation

Implementation of an automated system always seems to take longer than originally projected; this development project is no exception. Our first substantive discussions with NOTIS were held in January of 1987. We expected delivery by fall of 1987 of MEDLINE conversion programs, keyword searching functions for the catalog and MEDLINE, and database selection. We were hoping to have MEDLINE in production by January 1988. We would add keyword searching to the catalog in March. Wilson databases would be added in summer 1988.

The NOTIS software products were delivered and installed by December 1987. It was then that we discovered the true difficulties in working with the large MEDLINE file and the implications for

```
 LTML MORE                                                AJJ3862
                                          NOTIS CATALOGING         A1C1
 ML# AJJ3862 FMT B RT a BL m DT 11/18/88 R/DT none    STAT mn E/LEV 1 DCF ?
 SRC d PLACE ??? LANG eng MOD   I/LEV   REPRO   D/CODE c DT/1 1988 DT/2 8812
 CONT t     ILLUS ???? GOVT ? BIOG ? FEST ? CONF ? FICT ? INDX ? ME/B 1

 035/1:  : ]a (NLM)88327240.
 072/1: 7: ]a M.
 100:10: ]a Stead W W.
 245:10: ]a Information management through integration of distributed resources.
 265:  : ]a Z1.107.567.875.
 500/1:  : ]a Duke University Medical Center, Durham, North Carolina 27710.
 520/2:  : ]a Duke University Medical Center conducted a strategic planning
process focused on information management needs beginning in 1983 and ending in
1985. That effort concluded that the institution was ready to establish an
Integrated Academic Information Management System (IAIMS). A model was proposed
in which information management was to be achieved through integrated
distributed resources. The elements of the IAIMS model are ongoing policy
development and planning; communications; an electronic library

 LTML DONE                                                AJJ3862
                                          NOTIS CATALOGING         A1C1

 520/3:  : ]a  or resource inventory; coordination of the development or
selection of the end-user function; user support; and ongoing evaluation. This
model is being tested to determine its effectiveness in meeting the
administrative, patient care, research, and educational needs of a basic
science department and a clinical science department at Duke University.
Author.
 650/1:12: ]a Academic Medical Centers ]x MJ.
 650/2:12: ]a Computer Communication Networks ]x organization & administration
(OG) ]x MJ.
 650/3:12: ]a Computer Systems ]x organization & administration (OG) ]x MJ.
 650/4:12: ]a Information Systems ]x organization & administration (OG) ]x MJ.
 650/5:22: ]a Libraries, Medical ]x organization & administration (OG).
 650/6:22: ]a Models, Theoretical.
 650/7:22: ]a North Carolina.
 650/8:27: ]a Support, U.S. Gov't, P.H.S.
 700/1:10: ]a Stead W W.
 773/1:0 : ]p Bull Med Libr Assoc ]w BOW ]x 0025-7338 ]g 1988 Jul; 76(3): 242-7.
```

Fig. 1. A MEDLINE Record in 'MARC' Format.

```
                HEARD LIBRARY INFORMATION SERVICE                      A1C4

  On this terminal, you may search the following databases:

          DATABASE NAME:                         DATABASE LABEL:

     ACORN (Materials held by Heard Library)     DPAC

     MEDLINE (1986 - present)                    DMED

     WILSON                                      DWIL
       General Science Index (5/84 - present)
       Humanities Index      (2/84 - present)
       Social Sciences Index (4/83 - present)

    Select the database you wish to search by entering the four-letter
    database label below.  You may switch from one database to another
    from any screen by typing the database label.

TO RETURN TO THIS SCREEN FROM ANY DATABASE, type d

TYPE A DATABASE LABEL AND PRESS ENTER==>
```

Fig. 2. Database Selection Screen.

```
        DMED: MEDLINE ON THE HEARD LIBRARY INFORMATION SERVICE            A1C4

    DMED is a subset (1988-present) of the MEDLINE database, an index to
international biomedical journal literature, which is produced by the National
Library of Medicine (NLM).  MEDLINE, updated monthly, contains all citations
published in INDEX MEDICUS and corresponds in part to the INDEX TO DENTAL
LITERATURE and to the INTERNATIONAL NURSING INDEX.

     TYPES OF SEARCHING:                        COMMANDS:
      - FOR USERS FAMILIAR WITH ACORN:           Type k=, t=, a=, or sm=
      (To start a search from any screen)        followed by a SEARCH TERM
                                                 (keyword, title, author,
                                                     or medical subject)

      - FOR INTRODUCTORY SCREENS FOR KEYWORD SEARCHES: Type k
                                           TITLE SEARCHES: Type t
                                          AUTHOR SEARCHES: Type a
                                 MEDICAL SUBJECT SEARCHES: Type sm

FOR LIBRARY SYSTEM NEWS, TYPE news

TO CORRECT A MISTAKE, type over the error.
TO CHANGE DATABASES, type d
TYPE COMMAND AND PRESS ENTER==>
```

Fig. 3. Introduction to the MEDLINE Database.

```
        DWIL:  WILSON DATABASES ON THE HEARD LIBRARY INFORMATION SERVICE     A1C4

    DWIL is a collection of databases produced by The H. W. Wilson Company, New
York.  Updated monthly, DWIL contains references to periodical articles as
indexed in Humanities Index, Social Sciences Index and General Science Index.
Material in this database is protected by copyright @ The H. W. Wilson Company,
All rights reserved.  The H. W. Wilson Company makes no warranties, express
or implied, including as to fitness for particular purposes, and will not be
liable for damages or other claims and demands arising out of use of its data.

       TYPES OF SEARCHING:                     COMMANDS
        - FOR USERS FAMILIAR WITH ACORN         Type k=, t=, a=, or s=
        (To start a search from any screen      followed by a SEARCH TERM
                                                (keyword, title, author,
                                                    or subject)

        - FOR INTRODUCTORY SCREEN FOR KEYWORD SEARCHES:  Type k
                                          TITLE SEARCHES:  Type t
                                         AUTHOR SEARCHES:  Type a
                                        SUBJECT SEARCHES:  Type s

FOR WILSON SYSTEM NEWS, TYPE news


TO CHANGE DATABASES, type d
TYPE COMMAND AND PRESS ENTER==>
```

Fig. 4. Introduction to the Wilson Databases.

```
MEDLINE SEARCH REQUEST:  K=LIBRAR$.SU. AND IAIMS
BIBLIOGRAPHIC RECORD -- NO. 2 OF 5 ENTRIES FOUND

UNIQUE IDENTIFIER:
    (NLM)88327240.
AUTHORS:
    Stead W W.
ADDRESS:
    Duke University Medical Center, Durham, North Carolina 27710.
TITLE:
    Information management through integration of distributed resources.
SOURCE:
    Bull Med Libr Assoc 1988 Jul; 76(3): 242-7.
ENTRY MONTH:
    8812.
ABSTRACT:
    Duke University Medical Center conducted a strategic planning process
    focused on information management needs beginning in 1983 and ending in

                    CONTINUED ON NEXT SCREEN:  press ENTER

TYPE m FOR NEXT RECORD.  TYPE i FOR INDEX.
TYPE r TO REVISE, h FOR HELP, e FOR MEDLINE INTRODUCTION.
TYPE COMMAND AND PRESS ENTER==>
```

Fig. 5. MEDLINE Record Display: First, Second and Third Screens.

```
MEDLINE SEARCH REQUEST:  K=LIBRAR$.SU. AND IAIMS
BIBLIOGRAPHIC RECORD -- NO. 2 OF 5 ENTRIES FOUND  (CONTINUED)

Stead W W. Information management through integration of distributed resources
 ... (CONTINUED)
ABSTRACT:  (CONTINUED)
    1985. That effort concluded that the institution was ready to establish an
    Integrated Academic Information Management System (IAIMS). A model was
    proposed in which information management was to be achieved through
    integrated distributed resources. The elements of the IAIMS model are
    ongoing policy development and planning; communications; an electronic
    library  or resource inventory; coordination of the development or
    selection of the end-user function; user support; and ongoing evaluation.
    This model is being tested to determine its effectiveness in meeting the
    administrative, patient care, research, and educational needs of a basic
    science department and a clinical science department at Duke University.
    Author.

                    CONTINUED ON NEXT SCREEN:  press ENTER

TYPE m FOR NEXT RECORD.  TYPE i FOR INDEX.
TYPE r TO REVISE, h FOR HELP, e FOR MEDLINE INTRODUCTION.
TYPE COMMAND AND PRESS ENTER==>
```

Fig. 5. Cont.

```
MEDLINE SEARCH REQUEST:  K=LIBRAR$.SU. AND IAIMS
BIBLIOGRAPHIC RECORD -- NO. 2 OF 5 ENTRIES FOUND  (CONTINUED)

Stead W W. Information management through integration of distributed resources
 ...  (CONTINUED)
MAJOR SUBJECT HEADINGS:
    Academic Medical Centers--MJ.
    Computer Communication Networks--organization & administration (OG)--MJ.
    Computer Systems--organization & administration (OG)--MJ.
    Information Systems--organization & administration (OG)--MJ.
MINOR SUBJECT HEADINGS:
    Libraries, Medical--organization & administration (OG).
    Models, Theoretical.
    North Carolina.
CHECK TAGS:
    Support, U.S. Gov't, P.H.S.


                TO SEE BEGINNING OF THIS RECORD, press ENTER

TYPE m FOR NEXT RECORD.  TYPE i FOR INDEX.
TYPE r TO REVISE, h FOR HELP, e FOR MEDLINE INTRODUCTION.
TYPE COMMAND AND PRESS ENTER==>
```

Fig. 5. Cont.

storage and batch-processing time. Following numerous attempts to work within the existing hardware limitations, we ordered the first of the IBM3380s. The storage arrived at the same time that the Vanderbilt programmer assigned to the project left and at the same time that significant programming staff turnover and shifting responsibilities occurred at NOTIS. In all, we estimated a minimum of a six- to nine-month delay because of these changes in programming support.

Again, it became apparent that storage was still a problem; another 3,380 was added in October 1988. NOTIS reinstalled a clean version of all of the contracted software in early November; we were finally able to test all of the software elements, and fixes were supplied in December. Keyword indexes were created for the online catalog; keyword searching was introduced to the public in December. MEDLINE indexing was under way, and the software was once again tested with the one-year file; MEDLINE was put into the production system on January 19, 1989, and then introduced to the public on February 1—one year later than the original planned introduction date. We now estimate delivery of the Wilson conversion programs and the additional multiple database software in May. June 1989 is an early target date for introduction of the Wilson files.

Training

Because of the similarity to the online catalog, we expect the learning process of the article-level files to be relatively straightforward for users of the catalog. Our individual libraries may offer training specifically for the catalog, article-level files, and keyword searching, but such training sessions have been poorly attended in the past. We expect that training will be incorporated into regular bibliographic instruction programs and one-on-one user assistance.

To assist in our evaluation of MEDLINE, we have enlisted the assistance of volunteers from the Medical Center. They have been invited to attend MEDLINE introductory sessions along with library staff. Participation has been good to date. As a result of their attendance, however, we have discovered that many of them have never used the online catalog. What we thought would be transferable techniques from Acorn, the online catalog, are not, and more acclimatization time may be necessary to use MEDLINE. We may discover that we are attracting new library users through our extended offerings.

In general there are two primary issues which need to be addressed if library users are to receive the assistance they need. The first is a generalized need to understand effective search techniques using keyword and Boolean logic. This need arises for use of the catalog and also for use of all of the CD-ROM tools that are now becoming available. The other issue is one of clearly identifying to users the nature of each of the available searching tools, electronic or not. The fear here is that making a variety of tools available from a single workstation will discourage users from exploring other, possibly more relevant, sources. These issues are not new; they may just be more apparent in the electronic environment.

Costs

Our costs to date for the project have been substantial. The hardware upgrade costs approximately $265,000. Our share of the NOTIS development cost for the multidatabase project and database conversion has been $105,000. Annual database subscription costs for MEDLINE are $12,500; Wilson databases, $17,400. Wilson charges varying subscription rates based on the number of terminals. Vanderbilt staff support has been .75 FTE librarian and .75 FTE systems analyst/programmer for each of the past two years. We have yet to identify all the continuing costs for software maintenance and staff support; we can foresee the need for an additional operator to handle the batch processing and updating requirements.

Evaluation

Goal number four of the Enhanced Information Access Project—to determine user success with, preferences for, and institutional costs of different formats, different user interfaces, and different coverages—is our next agenda item. We will approach this evaluation on three fronts:

• Identify the costs of our mainframe databases and measure the levels of use. We will use summary-search statistics generated from transaction data; we will also use the transaction logs to try to identify and quantify search sessions for comparison with other alternative sources.

• Review transaction logs for identification of user successes and failures; survey users on completion of search sessions to elicit their perceptions of success and the adequacy of coverage.

• Conduct a user survey of Medical Center staff for comparison with results from one conducted in October 1988 as a precursor to the introduction of MEDLINE on Acorn.

PRELIMINARY CONCLUSIONS

We expect that library users will find our multiple database resources attractive. In addition to the article-level files discussed above, we will add the Center for Research Libraries database and our Television News Archives Index and Abstracts. We will seek other logical additions as well.

The most attractive feature of our recently added MEDLINE file is its availability over the campus network; while our locally networked SilverPlatter CD-ROM MEDLINE provides greater coverage, the university-wide access is not yet available. When holdings status can be added to the mainframe file, this will represent another feature not easily duplicated on CD-ROM. Ultimately, we will need to make a decision about which approach to take in committing scarce resources.

We anticipate great success with the Wilson databases: relatively lower costs for ongoing maintenance and a broad base of users. A permanent commitment to these article-level files seems likely.

Costs will remain relatively high in relation to the print products; the level of service, however, is significantly greater. We foresee the need to measure and account for use for appropriate charging back of search costs.

We now have the technology to provide significant access to a variety of information resources from a single workstation, inside the library or out. Increasingly, this will be the expectation of library users; we must strive to continue these development efforts. ■■

Loading Local Machine-Readable Data Files: Issues, Problems, and Answers

Emily Gallup Fayen

As we reach the mid-point of 1989, the last year of a decade of drastic change for libraries, little doubt remains that the online catalog has been very successful with users. Furthermore, librarians who have lived through the experience of automating their libraries are for the most part enthusiastic about the change. Both library users and staff have slowly come to recognize the advantages of electronic access to information. Users are pleased with the power and flexibility of online catalogs, for the convenience of dial-up access, and for the ability in many cases to search union catalogs that include far more than the holdings of their local libraries.

For over a decade, libraries have had access to the huge databases of OCLC and the Research Libraries Group as resources for interlibrary loan, not to mention as sources of cataloging data. These great national resources have changed forever the way libraries process their materials and handle user requests. While it is difficult to make the case that maintaining these resources actually saves money, they do allow libraries to meet their own and user's needs in ways that would be impossible without them.

Hardly an online catalog has been installed and opened to the public without someone saying (usually on the first or second day that the catalog is available), "This is really great! But now how can I use it to search journal articles and newspapers?" For the most part, perhaps this question has caught us unprepared. By giving users access to the online catalog we have solved some of their information retrieval problems (as well as solving some of our own) but in so doing have opened their eyes to what could be. Their level of consciousness has been raised. Many will never be satisfied with the old style access again. There are countless stories of times when the online catalog is unavailable for one reason or another. Users could find what they need from the old card catalog, but they almost universally prefer to go away and come back when the online catalog is working again.

So, it is quite natural to expect that users should have the same type of easy online access to journal articles, newspapers, and many other types of information. And, in fact, this is what has been happening: for years libraries have been routinely providing mediated or end-user searching of remote online databases in an effort to satisfy their patron's needs for access to these materials. This, too, is not enough. Perhaps as a result of their hands-on experience with the online catalog and with their more limited but generally happy experiences with mediated searching, library users have come to realize how useful, easy, and intuitive online searching can be. Furthermore, they want to do it themselves. They have jumped at the chance to use CD-ROM products, and in Dick Kollin's words, their reaction to unlimited use in a free, month-long test of the Answer Machine system in the library was "like throwing meat to lions." These users quickly showed us (if we

Emily Gallup Fayen is Director of Information Systems, University of Pennsylvania, Philadelphia.

had any doubts), that if online searching is available, cheap, and easy, everyone will use it.

The major drawback of the traditional online remote search services is that someone (the user, library, or company, or all three) must pay for the service, either per unit time, unit retrieved, unit printed or using some other measure, guaranteeing reinforcement of the idea that "there is no such thing as a free lunch," that is, information is a commodity to be parceled out to those who can pay. Consequently, in the academic environment at any rate, researchers with big grants and faculty from wealthy schools can afford to have lots of online searching done for them. At the other end of the spectrum, poverty-stricken graduate students may be unable to afford to do much searching, even if the results might shorten their work by several months or years.

Academic libraries try to overcome these inequities by subsidizing online searching. One approach is to require all users to pay a flat fee for each search. The library pays the difference between the fee charged and the actual cost of doing the search. This arrangement works quite well. Users who need quick searches of databases that are in the public domain pay the same amount as those who need a very extensive search of a costly file. Thus students are not penalized for having an interest in an expensive field. Of course, the library picks up the final tab.

Unfortunately, mediated searching is very expensive. Users who need it frequently want exhaustive searches conducted by an expert, so that database use can be extensive. Further, mediated searching is almost always performed during the usual workday hours when few discounts or other cost-savings arrangements from the utilities or the database vendors are in effect. Aside from the high costs involved, the other major disadvantages of this arrangement to the library are the need for elaborate record-keeping and the fact that the library's budget for online searching cannot be predicted with any accuracy at all at the start of the fiscal year.

Libraries have been able to reduce the costs somewhat by offering end-user searching through services such as BRS/After Dark, Knowledge Index, and Dow Jones. These special arrangements with the utilities help to reduce the overall bite, but the problems of charging fees for the service, the record-keeping, and the unpredictability still remain.

The CD-ROM has done much to change this situation. With these systems, once the library or some other sponsoring body pays for the necessary hardware and leases the files, they can be made available free (or almost free) to users. A few CD-ROM vendors keep track of usage for which they wish to be paid a supplement in addition to the basic license fee, but most do not. Further, the CD-ROM vendors have been clever enough to offer their products as subscriptions, so most libraries can pay for them out of their acquisitions monies. The big advantage of using CDs is that the cost is known ahead of time. Providing funds are available to acquire the basic system and datafiles, unlimited access to files is available to all users.

A disadvantage of CDs has been that until quite recently, these units were for the most part limited to use by one person at a time. Infotrac offered a three-user system, and some of the CD-ROM vendors are now bringing out systems that can access several databases simultaneously and can be used by more than one person at a time. The big disadvantage of even these new systems is that the workstations must be linked via a local area network, so true remote access is not possible. This severely limits the usefulness of these new products, at least in the academic environment, where one of the goals is to provide students, faculty, and researchers with access to information from wherever they may be on campus, either in libraries, classrooms, offices, or laboratories; or in homes or dormitories.

SELECTING THE DATABASE

Making data files available locally is another approach to providing the desired level of access while controlling costs. Although it is very expensive to mount a large database locally, if the anticipated use is high enough, the cost per search can become quite reasonable in the end. The critical issue is selection of the databases. Ideally, a database selected must be of vital interest to a large segment of the user community and preferably of interest to more than one such segment. For example, MEDLINE is useful not only to those working strictly in medicine but also to those working in health care administration, biochemistry, and other related disciplines. There should also be no license or other restrictions on use of the data file (other than a blanket license fee) that would curtail or restrict its use, at least by members of the licensing institution.

The database owners and providers all have an interest in the value of their information, and it is within their rights to insist on some restrictions on its use. For example, although the University of Pennsylvania wants to provide free and unlimited access to databases for members of the Penn community, we will not extend this privilege to the city of Philadelphia. This protects the vendors' interests as well as our own. Database providers are still working out many of the details of these agreements and in most cases seem open to negotiation. If library administrators ask for reasonable provisions, stick to their principles and refuse to lease databases from providers not meeting the provisions, the vendors will slowly come around. Those who don't will be out of business soon, so it won't matter anyway.

SELECTING THE SOFTWARE

The choice of software for mounting local databases may be determined entirely by the existing computer hardware environment and possibly by the software environment as well. There are several software packages available that may be used. BRS/Search, BASIS, DocuMaster, IBM/STAIRS, and STAR are some well-known products.

An approach taken by some libraries is to make access to other datafiles as much like their online catalog as possible. This is the path taken by Vanderbilt University. They chose to convert MEDLINE to a MARC-like format and load it into NOTIS as a separate database. This gives their users the advantage of having a similar interface and command structure for both the online catalog and other datafiles. The disadvantage is that it dilutes some of the power of the search software because certain features were not implemented by NOTIS in folding BRS/Search into their own software.

The University of Pennsylvania Library selected BRS/Search as the underlying software for its online databases. The package is available in several versions: BRS/Search runs under MVS for those users with access to a large IBM or IBM-compatible mainframe, either shared or dedicated. A slightly different version of the software runs on IBM or IBM compatibles under VM/CMS. BRS/Search also exists for the DEC environment and runs under VAX/VMS or one of several versions of UNIX; there is also a PC-based version of the system. BRS/Search may be licensed from the company for a fee of approximately $100,000, depending on the hardware platform and intended use. In addition to the initial charge, there is an annual maintenance fee.

The University of Pennsylvania selected the newly available VM/CMS version of BRS/Search because it would run on the library's mainframe computer in parallel with NOTIS. The library has an IBM 4381 Model Group 1 with 16 MB of main memory and over 20 GB of online disk storage. The library uses this equipment to provide support for NOTIS and also to provide online access for patrons to several machine-readable databases. The online catalog database now has over one million records. The system supports nearly 200 terminals with remote access for up to about twenty simultaneous users. So far, there has not been any degradation in response time as a result of this additional load on the computer.

Penn chose not to follow the Vanderbilt approach but instead acquired a separate license for BRS/Search that would allow the library to load any desired databases, either those acquired from commercial sources or those created locally without the need to convert to MARC format. We were also unwilling to give up some of the BRS/Search features for the sake of conformity with NOTIS. One of the major advantages of BRS/Search is that it was selected by NOTIS as its internal support for keyword and Boolean access to the online catalog. It is important for users to learn only one basic mode of access to information, whether for searching the catalog or searching other data files. Ideally, there should be one common command language that could be employed to search the catalog or other data files and would run on a wide range of equipment from PCs to mainframes.

In 1987 BRS introduced a new product called BRS/Onsite. This product is a stripped down version of the full BRS/Search software. Licensees receive the search software without the database loader. In addition, licensees may subscribe through BRS for pre-loaded copies of popular databases such as MEDLINE, ABI/INFORM, and others. Arranging to receive the data by this method saves countless hours of support time on the part of the library systems staff or other computer support group. This product is very attractive to libraries who want to experiment with loading local machine-readable files but which do not have extensive local systems support resources. BRS/Onsite relieves the library of the need to write conversion programs for the databases, to design file structures for them, and to load new tapes every month. The one drawback is that the library does not have the capability of creating its own local data sets.

LOCAL DATA FILES AT PENN

Using the criteria for database selection described earlier and drawing on information gleaned from bills to vendors for online search services, Penn chose to load three years of MEDLINE (English language only) and two years plus the current year building of ABI/INFORM. These databases were selected because of high interest among a wide spectrum of members of the Penn community. The database vendors (National Library of Medicine and Data Courier) were willing to negotiate license terms that were agreeable to both parties. The library had programmers on staff who were able to write the necessary conversion programs, define the BRS data file format, tailor the BRS/Colleague and Native Mode dialogues to meet Penn's own needs, and keep up with the monthly load processing.

In the future, the library may choose to use BRS/Onsite, especially for MEDLINE, which is an enormous database that is growing in size each year as NLM increases journal coverage and adds new features. The sheer volume of tape processing for the new records alone is awesome. Handling the correction tapes and changes to the MeSH vocabulary adds more problems. Further, it is important to our users that Penn's local implementation of MEDLINE be as close as possible to the commercially available product.

In addition to ABI/INFORM and MEDLINE, Penn has loaded two other local files that are publicly available using BRS/Search. The library had a card index to various Penn publications, such as *The Daily Pennsylvanian* (student newspaper) and *The Almanac* (primarily for faculty and administrators). This card file was entered mostly by students under the supervision of library staff and loaded into BRS/Search. Although not yet widely known, usage of this file in increasing. It is a valuable resource for student reporters and others who need access to information specifically about the university.

The second major local database to be made available is an index to a collection of television scripts housed at the Annenberg School of Communications. These scripts, a gift to the school from Walter Annenberg, have been indexed using a controlled vocabulary developed especially for this file. Information about individual episodes, series, actors and actresses, cameos, themes, and production details have all been carefully indicated. About 7,000 scripts have been processed so far. Users who need the scripts themselves may apply to the Annenberg School for access to the collection. The index is available to members of the Penn community as a local database.

LOCAL DATABASE USE

Figures 1 and 2 provide a clear picture of the use of these data files at Penn. Although the process began in early 1988, the matter of assigning user accounts and passwords was more troublesome than expected. The problem is in part caused by our operating environment, namely VM/CMS. VM requires a specific logon, that is, an explicit account and password for each user. It is impractical in an institution of Penn's size to assign 20,000 individual accounts and passwords; as a compromise, approximately 400 accounts were established. Several accounts were assigned to each library. Clusters of accounts were assigned to the biomedical library and the business li-

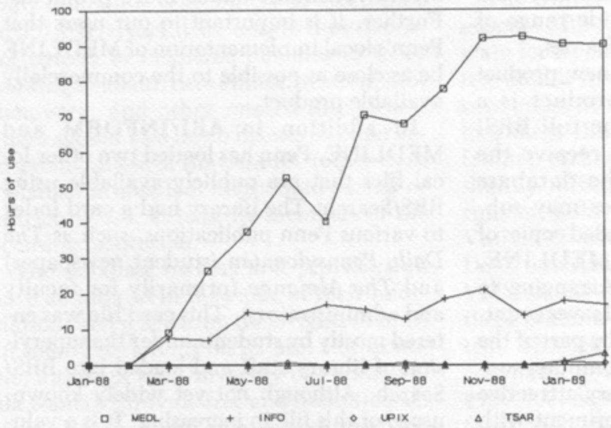

Fig. 1. Hours of BRS Use by Database (All Users).

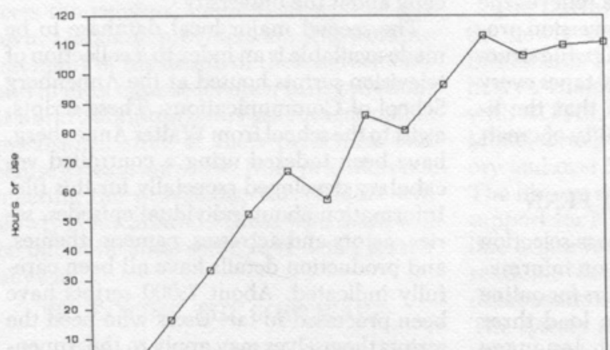

Fig. 2. Total Hours of BRS Use (All Users).

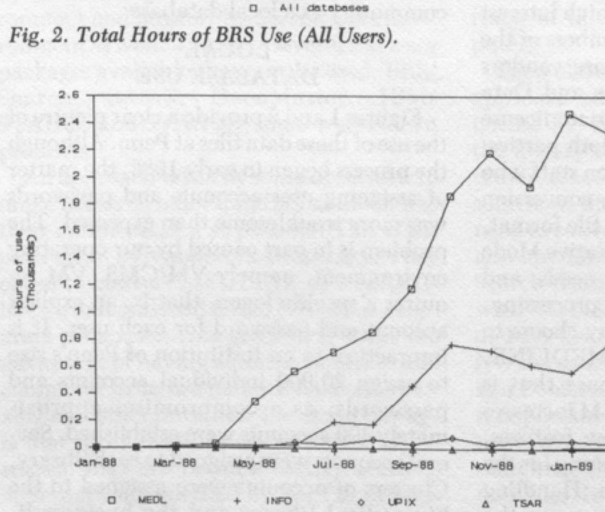

Fig. 3. Hours of BRS Use by Database (Non-Library Users).

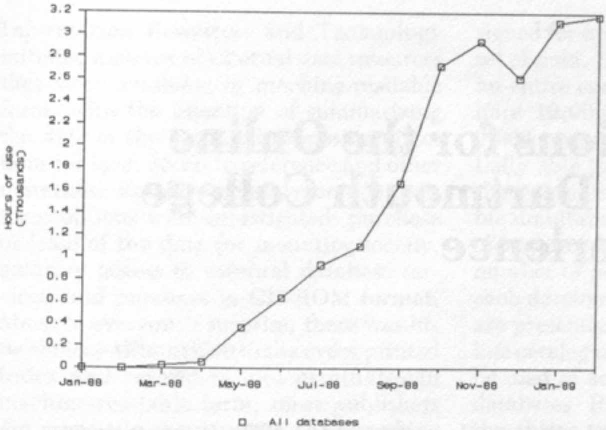

Fig. 4. Total Hours of BRS Use (Non-Library Users).

brary, where use was expected to be especially high. Other accounts were assigned to departments in the medical school, in the business school, and to others as available.

The arrangement is far from ideal. We are exploring other ways to control access. The goal is to provide unlimited access to the databases for members of the Penn community but to deny access to the public at large. Some alternatives may be to use a front-end computer to manage the accounts, to use IBM's CICS under VM, or as a last resort, to move to a different computer hardware platform with more suitable accounts-handling features. Response to the local databases has been very favorable. In fact, if we proposed removing this service as a cost-cutting measure, both users and libraries would suggest making other cuts first.

Penn set out to try to compare the relative merits of remote access, CD-ROM access, and of locally loaded data files. We anticipated that a clear choice would emerge. The biggest surprise of the entire project has been that for MEDLINE, at least, all three versions of the database are needed. Each serves different groups:

1. MEDLINE on CD-ROM is a superb teaching tool: it is located in the library, so staff are readily available to train users in formal or informal settings. It costs little to use, it is available all the time, and although it is not as fast or as powerful as online searching, it is adequate for teaching.

2. Local MEDLINE: Three years of English-language only MEDLINE meets the needs of most MEDLINE searchers. Having the file available locally means that users can get to it from their laboratories, classrooms, offices, and from the hospital. They find that being able to "turn off the clock" has significantly changed the way they use the database. They can do much more in the way of exploratory searching, tracking down ideas that may lead nowhere but may prove useful. Such leisurely access is far too costly using the various remote services.

3. Remote MEDLINE: For the serious researcher who needs an extensive search of the entire MEDLINE file, the remote search with the assistance of a trained searcher is the best alternative. This third mode of access meets the needs of a small but important group of users whose needs will be met with nothing less.

CONCLUSION

Locally loading databases are an attractive option for heavily used files that appeal to a large number of users. It leads to by far the most cost effective access for large numbers of users and changes significantly the ways users work with a database.

At the other end of the spectrum, having the ability to mount local databases makes it possible to provide online access to locally generated information that could not be provided any other way.

Both users and librarians are generally very enthusiastic about these new services. Of course, there are always questions about the trade-off between acquisitions and access, but that topic will have to wait for another paper. ■■

New Dimensions for the Online Catalog: The Dartmouth College Library Experience

Katharina Klemperer

The Dartmouth College online catalog originated in 1979. It currently includes monographs, serials, and on-order and in-process materials as well as an encyclopedia, a subset of the NLM MEDLINE file, and a number of locally created databases. The search engine (BRS/Search, Micro/Mini version) and the user-interface program make it easy to add new databases. The library and the computation center are embarking on a project to develop an integrated information system for the entire college.

The Dartmouth College online catalog has a long and interesting history. Work began in 1979 as part of a joint development project with BRS Information Technologies. BRS was interested in creating a marketable "search engine" for micro- and mini-computers (now known as BRS/Search, Micro/Mini Version), and Dartmouth was interested in creating an online catalog which would not simply replace its manual catalogs but provide significantly improved access to a wider array of information sources. A partnership was established. Dartmouth would develop a user-friendly front-end to the BRS search engine and would provide beta testing of the BRS software and, in return, would receive license to use the software. The first generation of the online catalog revolved around retrospective conversion of the entire collection, testing the BRS software and developing the user interface, and processing the OCLC and RLIN tapes, no small feat for a 1.7 million-volume collection whose first machine-readable records were created in 1972 and whose data-entry conventions developed along with the online catalog. By 1986 virtually the entire monographic and serial collection had been converted and loaded into a BRS database, which was accessible by means of terminals connected to the campus-wide local network or via dial-in. The current generation of the online catalog runs on a DEC VAX 11/785 computer under the UNIX operating system (BSD version 4.3) with BRS/Search, Micro/Mini Version 4.1, as its search engine.

EXPANDING THE ONLINE CATALOG

With a stable and mature online catalog, the library was now able to expand its dimensions. The first additions were the serials and the orders files, allowing users access to serials materials and to materials that are on order, in process, or cataloged but not yet in the main monographic catalog. Both the orders and serials files are now maintained by uploading weekly transaction files from the library's INNOVACQ system.

At this point the library's Committee on

Katharina Klemperer is Director of Library Automation, Dartmouth College Library, Hanover, New Hampshire.

Information Resources and Technology initiated a survey of external data resources that were available in machine-readable form, with the intention of summarizing the state of the field and proposing directions for local access to reference and other materials. For any given reference tool, three options were investigated: purchase or lease of the data for mounting locally, gateway access to external database services, and purchase in CD-ROM format. Much to everyone's surprise, there was little choice! Although virtually every printed index and reference tool originates in machine-readable form, most publishers are extremely conservative about making their raw data available for lease or purchase. There seems to be a fear of losing revenues from printed products, online search services, and CD-ROM products. A few publishers make raw data available only at exorbitant prices, and others have signed exclusive contracts with a single vendor to make their products available on CD-ROM or on microcomputers. What the Committee on Information Resources and Technology had originally intended as a proposal for purchase of a variety of data resources has turned into a status list, continually monitored and updated, of all the resources that Dartmouth would *like* to buy, with the current prices, if they are available at all. The document contains information about the online availability of such data resources as dictionaries (standard, specialized, and foreign language), encyclopedias (general and subject oriented), style manuals, journal citation indexes, and other reference tools. In fact, the summary document itself is available to library staff through the online catalog.

It is interesting to note that vendors, when they choose to sell or lease their data, have adopted a number of different pricing strategies. (1) *A flat fee*. This approach is most advantageous to a large institution with a potentially very large number of users. (2) *A fee based on the size of the institution*. This is probably the most advantageous to Dartmouth, a small institution with a relatively large number of users. (3) *A fee based on the number of terminals or ports into the system*. This is really only reasonable for systems that are specifically designed for non-networked access to a single set of data. How can one count ports when an entire campus is networked, with perhaps 10,000 hard-wired wall ports and 5,000 terminals, all of which are potentially able to access the online catalog? If the answer is to count the number of possible simultaneous users on the database machine, then how does one determine what number of ports should be apportioned to each database? In Dartmouth's case, there are presently thirty-two ports into the online catalog machine, any one of which can be used to access any one of over a dozen databases. It is difficult to justify paying the thirty-two port rate for one of this dozen. (4) *A fee based on usage*. This charging method is probably the fairest but the most complex, because it requires systems to monitor usage. Even then there is no clear-cut measure of "usage": does one measure connect time, number of searches, number of retrievals, number of items printed? To some extent, usage-based fees run contrary to the principle of locally mounted databases, which attempt to give free, easy, and unmonitored searching to all users. With usage-based fees it is tempting to influence users to search in a highly efficient manner, to reduce the cost, even if it restricts the user's freedom. It is reasonable to surmise that the difficulty in setting charges for local database use plays a large part in vendors' inclination to concentrate on CD-ROM products, where usage is much easier to assess.

In the end, Dartmouth elected to purchase and mount online data resources that were available at a reasonable price and did not require an excessive amount of storage space. These currently include Grolier's *Academic American Encyclopedia* and a subset of the National Library of Medicine's MEDLINE database. Dartmouth is experimenting with a test file from H. W. Wilson's *Reader's Guide* but has not committed itself to mounting any of the Wilson databases yet.

THE ENCYCLOPEDIA

The Encyclopedia project presented a number of new challenges and potentials. This was the first true full-text database to be mounted. Up to this point all the databases had been indexes to documents: the card catalog, indexes to special collections, indexes to journal articles. An encyclopedia article is, by contrast, the document itself. The chief differences between index documents and full-text documents are that full-text records vary greatly in length, some being very long, and that they are not structured as index records are. An index record contains a number of short fields that are identifiable in each record: author, title, place of publication, publisher, date of publication, collation, subject headings, OCLC number, LCCN, etc. An encyclopedia article contains a title, pronunciation, see and see also references, a bibliography, perhaps some tables, and a very large body of text which is sometimes tedious to navigate.

Because the entire text is searchable in the Dartmouth implementation, it is possible for a user to retrieve a large number of articles that mention a given topic, but it is not easy to locate the pertinent part of a long article. For this reason, two new display formats were added to the online catalog display. DISPLAY PREVIEW lists the title and first paragraph of the text of each retrieved record. DISPLAY CONTEXT lists the article title and all paragraphs that contain words that the user searched for.

The full-text nature of the encyclopedia allows BRS/Search to shine in the areas for which it is best suited. Because entire articles are indexed, users can retrieve all references to a particular topic. For example, a full-text search for *Ronald adj Reagan* will retrieve 117 articles, among them articles titled "space exploration," "Gorbachev, Mikhail," and "MX missile." A full-text database also makes full use of the proximity features that BRS provides. With the large size of encyclopedia records, adjacency searching often significantly improves precision. A familiar example is searching for *Dartmouth College*. In the encyclopedia, adding the adjacency operator eliminates several false retrievals of "Britannia Royal Naval College at Dartmouth [England]."

The encyclopedia has been well received by the student and faculty population. Every quarter, updated articles are received on tape and loaded, and these are indeed up-to-date. On January 20, 1989, the library received a tape containing an updated article for "George Bush," including mention of his inauguration.

THE MEDLINE PROJECT

During 1988 and 1989, a major area of online catalog development has been the addition of a subset of the National Library of Medicine's MEDLINE database. The purpose of this database is not to replace the existing use of online search services, nor to reduce the costs incurred by online search services. Instead, it aims to make available to the entire college community quick and easy access, at no charge, to the medical literature which is actually available on the campus. Only articles from journals actually owned by the library or medical departmental collections are included in the database; this represents approximately 59 percent of the current MEDLINE file. All titles in the Dartmouth serials file were searched in the NLM SERLINE file. The matches were reviewed manually, since there is enormous variation in cataloging between the two serials files, resulting in a considerable number of false matches and multiple matches. Each article being processed is passed against the resulting journal title list and only loaded if its journal title appears on the list. The entire MEDLINE article is indexed, including abstracts. To encourage searching by MeSH (Medical Subject Headings) and using the MeSH tree structure, the NLM MeSH file is available as a separate file, and MeSH tree numbers have been added to all the article records. Of course, casual users can search by keyword and will benefit from the exhaustive indexing.

As every other institution that has attempted to mount this database has discovered, the MEDLINE database is very large, and requires substantial investments in disk space. Because of the extensive indexing, including the addition of MeSH tree numbers, the library projects that the current year plus two previous years of Dartmouth's MEDLINE articles will easily consume 1.5 gigabytes of disk storage.

The Dartmouth MEDLINE file is the first that has really required extensive, specialized user education. Although adequate results can be obtained using only the standard techniques of the online catalog, the database is sufficiently distinctive that the library staff have written a user guide and provide training sessions for users. The most obvious distinctive elements are the use of MeSH and the formatting of author names, journal titles, and other fields.

LOCAL DATA RESOURCES

A large, untapped source of data turned out to be available locally. The library's Special Collections Department has undertaken several projects for briefly cataloging some of its theater and music collections. The first of these, called PLAYBILL, is an index to a collection of playbills, theater programs, and souvenir programs, only a few of which are cataloged in RLIN. A second is THESPIS, an index to a vertical file of a variety of theatrical information such as clippings, photographs, pamphlets, and drawings. The third is called G MAJOR and provides access to a very large collection of popular sheet music, which is cataloged nowhere else. All three of these databases are being constructed by keying the data into an Overvue database on the Apple Macintosh and then uploading the records onto the online catalog machine.

The library is also undertaking a preliminary project to provide an index to computer files available anywhere on campus. This particular project is tricky because it involves cataloging items that are not actually owned by the library. What the library would like to index in this database are programs, both commercially available and locally developed, and data files such as statistical data, again both purchased and locally created. Maintaining such a database will require continuous surveying of all departments to keep up with their acquisitions and development, and it is not being undertaken lightly. The value of such a database to the community will be immeasurable, however, in answering such questions as what software can be purchased at the computer store or whether someone has already written a program to perform a needed function.

The library is now undertaking to maintain a library policies database of its own internal and external policy documents online. These policies range from borrowing privileges for outside users to a personnel policy for released time to donate blood to a policy for hours during which online catalog maintenance may occur. Here the main technical problem is defining a document structure that can be applied, often retroactively, to quite a variety of documents. Equally important is deciding what policies ought to be made available to the public and which should remain internal.

A fascinating project, which is under investigation but temporarily on hold because of the excessive staff time required, is a "Tables of Contents" database, which would provide a current awareness service by giving online access to tables of contents of journals as they are received. Users would be able to search for authors' names, article titles, or journal titles and then to limit their retrievals by date. The result would be a quick survey of the current literature. The original intent was to scan the contents pages of incoming journals electronically and then convert the scanned pages into text files. While the technology exists to perform the scanning and conversion to ASCII text files, the library soon discovered that there is absolutely no consistency in the format and layout of contents pages, and considerable manual effort would be required to identify the author, title, and page numbers even after successful scanning. It would be faster to key in the information directly than to analyze the scanned text.

DESIGN FEATURES OF BRS/SEARCH AND THE DARTMOUTH ONLINE CATALOG

Technically, it has been easy for Dartmouth to add new databases to its online catalog for two reasons: BRS/Search is easy to manage, and the online catalog user interface has been specifically designed to allow easy addition of new BRS databases with no programming effort. The adoption of BRS/Search, Micro-Mini Version, has facilitated Dartmouth's ability to develop and enhance its online catalog with a fairly modest automation staff consisting of a director and three programmers.

It is extremely easy to build new BRS

databases. Once one has figured out the general features that BRS/Search offers, one designs the basic table, or "database form," that specifies in considerable detail the input format, indexing, and display formatting that BRS will provide for each field. If one has control over the data entry, as in the case of Dartmouth's Library Policies and Special Collections databases, it is easy enough to define the input format that the data entry operators must use. Data that is being acquired externally almost always requires preprocessing, however, and this can become quite complex, depending on the nature of the data.

Loading data is a straightforward operation for databases that merely add data; the preprocessed data is simply fed into the loader. However, if individual records require updating on an ongoing basis, database maintenance becomes a complex process, especially if chronology is important. This is the case in most data subscriptions, such as RLIN, OCLC, the encyclopedia, and MEDLINE, and is characteristic of the majority of information resources Dartmouth and other libraries want to incorporate. Each incoming record must be searched in the database in advance, then the dates must be compared, and if the incoming record is newer, both an "add" and a "delete" transaction must be created. BRS does support "replace" transactions, but they are extremely inefficient in all but the tiniest databases, so it is necessary to use the combination of "add" and "delete" instead. All the "add" transactions are loaded and, in a separate run, the "delete" transactions. A further complication is that BRS does not automatically reuse the storage space used when records are deleted. To recover this unused space, it is necessary to run a database reorganization utility periodically.

The loader is really a batch loader; that is, it works most efficiently when a group of records are processed all at once. Attempts to actually add and delete individual records in the data preparation program will result in a program that is too slow to use in a production system. When loading a batch of records, the loader does not process an entire record all at once. Rather, it first adds the text for all the input records and subsequently updates the indexes. If the load is interrupted for any reason, the database must be restored and the load restarted. Needless to say, for a database of any size that must be always available, this results in considerable overhead in maintaining disk backups during all loads. The database can be searched while a load is in progress, although any new records are available only after the entire load is completed.

BRS/Search is much more than a database manager; it provides a number of features that would ordinarily have to be included in a data-preparation program. Stopword lists are fed into the loader. The database owner can opt to "bind" fields; that is, index the entire string as well as its component words. A variety of ways of treating hyphens, periods, apostrophes, and other punctuation can be defined.

Librarians are also pleased at the functional capabilities of BRS as a search engine. It highlights retrieved terms. It supports searching of any combination of fields in a single command, as well as searching of the entire record; searching the entire record is in fact the fastest search. A variety of proximity operators, as well as truncation (single- and multi-letter, suffix, prefix, and internal) are provided. The index terms can be browsed.

A major selling point of BRS/Search, Micro/Mini version, is the fact that it supports real-time update; that is, the database records can be updated in place by an operator sitting at a terminal. This is undoubtedly a useful feature for some operations, but it is not helpful in the maintenance of large databases from external sources.

The fundamental difficulty which must be accommodated when using BRS as a database manager is that it does not support true relational databases. Basically, each database is a single flat file. This is a major hindrance in developing access to any kind of database that uses an authority file or thesaurus. Not only are space savings prevented, but the user cannot be provided with automatic access via cross-references, and authority normalization is difficult, if not impossible. While it is possible to mount a thesaurus as a separate file (Dartmouth has done this with the MeSH file as an adjunct to the MEDLINE article file), only one file can be open at a time in a given user's search session. Changing files is a slow process, and all searches are lost when the file is changed.

THE NATURE OF THE USER INTERFACE

Dartmouth has chosen to write its own user interface to the search engine, rather than use the nonstandard user interface provided by the vendor. This was written in C and makes use of BRS's unadvertised C interface. The Dartmouth command language is largely in conformance with the National Information Standards Organization's *Standard Command Language for Interactive Information Retrieval Systems* (balloted during early 1989), and it also provides menus which prompt the user for command words.

The Dartmouth Online Catalog user interface is generic; that is, nothing is hard-coded that relates to any particular database. The entire user interface is controlled by a table of "session variables," which contains a list of databases available and in service, the commands that have been implemented, the menus for each user condition, the notifications and error messages for each user condition, the locations of various help and explain files, the names of the remote printers accessible to the online catalog, and the default settings in effect for all users. Each database has a table of "database variables," containing such information as the location of database-specific help and explain files; special menu texts; a list of searchable indexes with their corresponding BRS names; a list of display fields with their corresponding BRS names; lists of sort, exact-searchable, and limit fields; and predefined display formats for the database, with instructions for constructing them. These tables not only allow for extreme flexibility in adding new databases, but they also allow nonprogramming staff to make modifications of wording and layout to the user interface quickly and simply.

A recently discovered benefit of this approach has been that it is easy for a nonprogrammer to design a display format that can be downloaded, using a terminal emulation program for the Apple Macintosh, into a Macintosh bibliographic reference manager. So far, special display formats have been designed for easy downloading into PBS's ProCite and Niles Associates' EndNote programs. These are the first steps toward using the Macintosh as a workstation with the online catalog rather than simply as a terminal emulator.

It is now a simple matter to add new BRS databases to the online catalog with no special programming. Databases that use a different database manager will require some program changes, but since all the routines that interface with BRS are clearly defined and isolated, it will not be a major effort to add parallel routines that interface with another search engine.

THE COMPUTING ENVIRONMENT AT DARTMOUTH

Most of the world knows that time-sharing was first developed at Dartmouth on its own Dartmouth Time-Sharing System (now known as Dartmouth College Time-Sharing) and that John Kemeny and Tom Kurtz first unveiled BASIC, still a heavily-used teaching language, at Dartmouth in the early 1960s. During the 1970s and 1980s, Dartmouth also committed itself to making centralized computing available to *all* members of the campus community, with the result that every office, classroom, and residence hall room now has a direct connection to the Kiewit Communication Network, and therefore to all the computers on the network, including the library's online catalog. In 1984 the Apple Macintosh microcomputer made its appearance on campus. Incoming freshman are not required to buy computers, but over 80 percent of them do, as have most of the faculty and staff, with the result that the entire campus is highly "computer-literate." For those who do not have their own computers, there are public clusters of workstations at a number of locations on campus, so that the entire community has access to a great deal of computer power, both on the desktop and at the computation center.

The library has also benefited greatly from the service philosophy of the college's Kiewit Computation Center, which itself is based on the "library model" of service. The computation center sees itself as providing computing service to all members of the campus community without expecting compensation for computing time. It has been very easy for the library to develop a cordial relationship with the computation center. The library owns its own computers and storage media but houses them in the computation center, which provides facilities management, including operator services, management of hardware maintenance contracts, system administration, and backup services. This has allowed the library's automation department to concentrate on systems analysis and application programming.

THE DARTMOUTH COLLEGE INFORMATION SYSTEM (DCIS)

The library is anxious to continue to provide online catalog access to its own collections and to purchase indexes and reference tools, but there is a wider constituency on campus with data resources that also deserve access. These are generally not under library control, nor does the library wish them to be, but in the eyes of the community they are of significant value. These include scholarly resources such as the Dante project, a multi-institutional project whose full-text database of selected commentaries on Dante's *Divine Comedy* resides at Dartmouth; instructional resources, such as the full text of course reserve readings, laboratory manuals, and viewing and listening assignments; administrative resources, such as the campus directory, class lists, and the prospectus of courses; "daily-living" information, such as a calendar of athletic and cultural events, rental housing listings, and employment listings; miscellaneous locally developed resources, such as an Afro-American studies database, the college newspaper in full text, and a "Dartmouth Almanac" of facts and history of Dartmouth; and external information resources, such as RLIN and other online retrieval systems. Many, although not all, of these information resources are already available in machine-readable form, but the method of access is varied and complicated.

In response to this need, the Dartmouth College Library and Dartmouth's Kiewit Computation Center are currently embarking on the next generation of access to information resources, going beyond the online catalog to an integrated campus-wide information system. The basic premise of the Dartmouth College Information System is that any user, from any place on the campus, should be able to access a variety of information resources, on a variety of mainframes and using a variety of database managers, using a single easy-to-use user interface from his or her desktop workstation. Information providers should be able to add their resources to the information system with a minimum of programming effort.

DCIS consists of three components:

1. *The user's workstation*, upon which the user interface resides. The user interface will be geared toward nontechnical users. An easy-to-use windowed dialogue will provide consistent access to the various databases in the system. Remote connections to information will be transparent: users will not be required to specify which machine they are accessing, nor will passwords be required except in the cases where access is for some reason restricted. The workstation will make available the basic functions of information retrieval (selecting and searching the correct databases), output manipulation (formatting the retrieved information in a number of ways, printing it at a number of locations, storing it on magnetic media either centrally or at the user's workstation, and sending it to other users either electronically or in paper format), and other specialized services, such as document request and delivery and storing searches for automatic periodic execution.

It is expected that the DCIS user interface will be fully integrated with the local electronic mail and bulletin-board services. While the interfaces to different databases may not be identical, all will have the same "look and feel," so that the user is not required to learn special procedures for each database.

The user interface design is taking a modular approach, with emphasis placed on building a "toolkit" of functions and dialogue components. This will allow data providers to assemble user interfaces that are tailored to their own types of data, with a minimum of assistance from central programming staff.

2. *A high-speed network.* The existing Kiewit network is a high-speed local data network which currently extends to nearly all the buildings on campus. All residence halls and all academic and administrative offices have access to 230.4 kbps AppleTalk networks. Bridges to other protocols are also supported, including Ethernet and asynchronous RS-232. Substantial improvements to the current configuration are planned in support of DCIS, including wiring an Ethernet backbone to connect the various Appletalk networks.

3. *The database hosts.* The data itself will reside in various databases on a variety of mainframes, using a variety of operating systems and database management systems. The initial implementations will include text and numeric data, but access to aural and visual data resources is envisioned in future generations. The DCIS employs a client-server model. A standard communication protocol has been defined that stands between the user interface on the workstations and the various database servers. The user interface produces queries in a standardized structure, which are sent to the appropriate server, which translates the queries into search commands appropriate to the database manager it is serving. A similar procedure occurs in reverse when the servers return retrieved information to the workstations.

ADVANTAGES OF THE CLIENT-SERVER MODEL

In the past, information resources such as library catalogs have been mounted on central mainframes, with the user interface program and the database residing on the same machine and the user logging into the database by means of a terminal or terminal emulator. Even if a smart workstation were involved, its effectiveness was reduced to that of a simple terminal.

The present model offers several advantages. Users need not be aware of what machine they are accessing a particular login or dialogue. The entire system is not dependent on the functioning of a single machine; loss of one database does not mean loss of the entire system. A variety of database management systems and operating systems can be used; each database server can act independently of the others. Growth can be incremental. As data becomes available, new database servers can be added. Alternative storage media can be explored. Rather then relying on magnetic disks as in the past, other media such as CD-ROM or laser disks can be included. Perhaps most welcome from an administrative point of view, the control of the data is now distributed. Once the server has been developed, individual data suppliers can manage the maintenance of their own data. There is no longer a centralized administrative body continually extending itself to embrace more and more data resources.

Dartmouth College has completed a pilot project to demonstrate the feasibility of the DCIS approach with two kinds of data, text and numeric. Both the library online catalog and a social science statistical database have been made available through a Macintosh user interface. The communication protocol has been implemented and a database server written for each of these data resources. The knowledge gained from these experiences will be used in the future development of DCIS.

CONCLUSION

The Dartmouth College Online Catalog, innovative since the time of its inception, now stands on the frontier of perhaps its most exciting era. The library already offers its services beyond its walls; the next generation of information systems will embrace a multitude of information resources that originate outside its walls. The coming months, years, and decades promise to offer increasing challenges and opportunities in cooperative development. ■■

TOC/DOC at Caltech: Evolution of Citation Access Online

Sandra Card

In November 1988 the Caltech Library made available to its users a new service which combines online access to journal citations located at Caltech with document delivery. This service evolved over several years, beginning with efforts to include the citation access in the library's integrated library system and culminating in the purchase of the powerful BRS/Search software to run on a small IBM mainframe specifically to support this project. Various decisions about implementation were made along the way, with no perfect solution being reached. Instead, the library learned to balance different factors against each other in order to determine the best solution.

California Institute of Technology (Caltech) is a small, private university with concentrations in engineering and the sciences. It currently enrolls about 900 undergraduates and 1,000 graduate students and has about 800 faculty. The university emphasizes research for students, from undergraduates through postdoctoral fellows. Caltech's reputation for scientific innovation and excellence is evident by the many Nobel prize–winners among its distinguished alumni and faculty.

The Caltech library system is very decentralized. A central building holds eight collections; there are ten branch or satellite libraries located in other buildings. Current volume holdings are approximately 500,000. Although the library adds about 9,000 monographs per year, it subscribes to approximately 6,200 serials on a current basis. The Caltech library has offered a document delivery service for many years, whereby campus users can request photocopies of single journal articles from the library to be delivered to their offices. The number of articles requested each year is similar to the number of items circulated. Use of the journal literature represents more than half of the library activity at Caltech.

TOC/DOC TODAY

In November 1988 the library instituted a new online table of contents/document delivery service called TOC/DOC. This system is available to all Caltech faculty, students, and staff through Caltech's local area network, CITNET.

To use the system, a Caltech faculty, student, or staff member must have access to CITNET. At the network prompt, they ask for connection to the TOC/DOC service. The main menu offers a variety of choices: searching for a specific journal by keywords, searching by author, searching by keywords in the title, or combined author/title searches (figure 1). One feature of the system is that we load only those titles on the SciSearch tapes which Caltech currently owns.

Therefore the system acts as an extension of the catalog in allowing people to locate the contents of the journals as well as the journals themselves.

Sandra Card is Systems Development Officer, California Institute of Technology, Pasadena.

When users search for a journal, they will see a display in table of contents format for a specific issue of a journal (figure 2). They can then reach a single article display (figure 3). The other types of searches (author and title keywords) lead directly to lists of articles found and then to single-article displays (figures 4,5). When users locate a journal article they want, they flag that title online for document delivery and fill in delivery and billing information at the end of their search session (figure 6). If the material is on the shelf, delivery of the photocopied article is guaranteed within forty-eight hours. In December 1988, the first full month of the service, 14 percent of the document delivery requests came by way of TOC/DOC.

There are also two types of help online: contextual help, which is customized to the screen that the person is on at the time help is requested (figure 7), and the menu help, which is a series of more detailed help screens available from the main menu (figure 8).

ORIGINS OF TOC/DOC AT CALTECH

In the January 1984 issue of *Engineering and Science* (47:3 pp.15–19) the new director of the library, Glenn Brudvig, first suggested the idea of citation access online for the campus. The library used OCLC for cataloging and interlibrary loan and used standard online services such as Dialog for reference. There were no other automated functions at that time. The library seriously needed full automation, in part because of its highly decentralized number of locations in relation to the size of the campus. As part of the envisioned automated system at Caltech, the director proposed online access to article holdings at Caltech. He stated:

> With some 50 million journal articles currently accessible online, this vast amount of information needs to be filtered so that we get only what is most relevant to the users on campus. If journals are selected carefully and quality is maintained, then the citations received by matching our current subscriptions against a large database should reflect that quality. Those 50 million articles would still be accessible online for a charge for individual searches, but the regular users of the library would have access without charge to the smaller numbers of citations in current journals. A file of the most cited articles could also be maintained in a similar manner.

With this in mind, Caltech hired a systems librarian to work on automating the library system. Convenient, easy access to the journal literature in the Caltech Libraries was desirable. Many of the scientists relied entirely on journal literature. This is not surprising, given that the scientists at Caltech not only need to keep up with current research but are frequently creating new areas of research themselves. The decisions to be made later involved the exact means of achieving this and the extent of the project.

Caltech's local area network, CITNET, has been in operation since the fall of 1983. This network uses Ethernet technology; specifically it uses Ungermann-Bass network hardware with their Net/One software. The entire campus was wired with network cables located in all buildings; therefore, we needed an integrated library system which could operate over a local area network. This also emphasized the need for article citations which could be accessed from any network connection on campus.

LAYING THE GROUNDWORK FOR TOC/DOC

In the spring of 1984, Caltech started researching automated integrated library systems. At the same time, we studied sources of data for journal citations. Although some libraries had in the past analyzed journal issues manually and filed cards at the article level in their card catalogs, we decided manual keying of data for online access was not affordable at Caltech.

One capability we definitely wanted was the citation access as part of the online catalog. We felt that author and subject searches should be conducted at both the monograph and the journal article level without the user having to distinguish types of formats. Moreover, it should enable users to check the serial holdings for the location and availability of the issue whose contents they were scanning.

We opened discussions with the Institute for Scientific Information about the lease-purchase of their SciSearch tapes. We de-

```
TOC/DOC          Welcome to TOC/DOC                        toc08
                 Main TOC/DOC menu
               Current Database:TY89
SEARCH OPTIONS:

  1 . Table of contents from a journal (keyword search)

  2 . Table of contents from a journal (exact one word title)

  3 . Papers by a specific author or authors

  4 . Keywords in article titles

  5 . Author name plus keywords in article titles

  6 . Date-restricted search

  7 . Main Help Menu and other search options

       Enter # to select:

H:HELP Q:QUIT  R:RESET FOR NEW SESSION
```

Fig. 1. Main Menu for TOC/DOC.

```
TY88 TOC/IEEE TRANSACTIONS ON ELECTRON DEVICES  36(1)1989     02/28/89

1   A NEW METHOD FOR THE WORK-FUNCTION DIFFERENCE DETERMINATION USING
    BURIED-CHANNEL MOS-TRANSISTORS
    by INIEWSKI-K    REFS: 10                                  152 - 153

2   ACCURATE PROFILE SIMULATION PARAMETERS FOR BF2 IMPLANTS IN PRE-
    AMORPHIZED SILICON
    by TASCH-AF SHIN-H PARK-C  REFS: 10                        149 - 152

3   BASE SPREADING RESISTANCE OF POLYSILICON SELF-ALIGNED BIPOLAR-
    TRANSISTORS
    by GREENEICH-EW     REFS: 5                                147 - 149

4   EFFECTS OF TITANIUM SALICIDE PROCESS ON MOSFET CHARACTERISTICS
    by CHANG-ST CHIU-KY    REFS: 7                             145 - 147

5   DESIGN AND FABRICATION OF 0.25-MU-M MESFETS WITH PARALLEL AND PI-GATE
    STRUCTURES
    by NAGARAJAN-RM VANHOVE-JM RASK-SD  REFS: 9                142 - 145

Enter # <cr>=select, <cr>=more, T<cr>=top, S<cr>=new srch:
```

Fig. 2. Table of Contents Display.

```
TY88                        Citation                        02/28/1989

2    ACCURATE PROFILE SIMULATION PARAMETERS FOR BF2 IMPLANTS IN PRE-
     AMORPHIZED SILICON
     NOTE (REFS: 10)
     by TASCH-AF (*R) SHIN-H PARK-C
        ALVIS-J NOVAK-S PFIESTER-J
        UNIV TEXAS,MICROELECTR RES CTR/AUSTIN//TX/78712 (*R)

   IEEE TRANSACTIONS ON ELECTRON DEVICES
        VOL:36 (1): 149 - 152(1989)
        LOC:ENGR APEE

Do You want this article photocopied? Y/N(N):
```

Fig. 3. Single Article Display.

```
TOC/DOC                          TY89                          toc08

        Words in Article titles

    Enter key words in article titles; truncate with '$'
        Examples:

             superconduct$ (high or tc)
             stabiliz$ flame$1
             wing$2 vort$7
             vitamin adj a (deficien$4 or metabol$4 or
                  transport$4)


  ENTER SELECTION:  archaeo$

H:HELP Q:QUIT P:PREVIOUS SCREEN M:MAIN MENU
```

Fig. 4. Search Screen for Keyword in Title.

```
TY89          38 ARTICLE(S) FOUND IN TY89                 02/28/89

1   THE NEXT ARCHITECTURE - DESIGN FOR THE 90S
  by WILLIAMS-T

   COMPUT DES VOL 27(22) : 27 - 28(1988)  EDITORIAL

2   DETECTION OF SEQUENCES HOMOLOGOUS TO THE HIGHLY-CONSERVED HSP70 GENE
   OF EUKARYOTES IN THERMOPHILIC EUBACTERIA AND ARCHAEBACTERIA
  by LAMBERT-E COWAN-D PIPER-PW (*R)

   FEMS MICROB VOL 56(2) : 157 - 160(1988)  ARTICLE

3   ASYMMETRIC CELL-DIVISION OF A TRIANGULAR HALOPHILIC ARCHAEBACTERIUM
  by HAMAMOTO-T TAKASHINA-T GRANT-WD HORIKOSHI-K

   FEMS MICROB VOL 56(2) : 221 - 224(1988)  ARTICLE

4   ARCHITECTURE OF A PACKET SWITCH BASED ON BANYAN SWITCHING NETWORK
   WITH FEEDBACK LOOPS
  by UEMATSU-H WATANABE-R

Enter # <cr>=select, <cr>=more, T<cr>=top, C<cr>=chge yr, S=new srch:
```

Fig. 5. Result of Keyword Search; Selection from this List Leads to Type of Display Shown in Figure 3.

```
                    TOC/DOC Copy Request

Enter # followed by a comma (,) then the information
Examples:
    1, John Smith
    2, F
    3, Biology

    1 , Name:
    2 , Status(F G U S):
    3 , Division:
    4 , Mail Code:
    5 , Phone number:
    6 , Account:


  Enter # and information: 1,Sandra Card

H:HELP  Q:QUIT  P:PREVIOUS SCREEN  M:MAIN MENU  S:NEW SEARCH
```

Fig. 6. Screen for Entering Billing Information.

```
                   ARTICLE DISPLAY HELP

The number of articles retrieved in your search is given at the top of
the screen.  No more than 5 authors per article are displayed and the
journal title is abbreviated per ISI conventions.
The articles are numbered in a sequence established by the search.
You must use the sequence number to the left of the article to obtain
a more detailed display before a copy request can be authorized.

While some other authors' names may be followed by the reprint source
indicator (*R), the citation display must be consulted for complete
reprint information.


STRIKE <cr> TO CONTINUE ...
```

Fig. 7. Contextual Help Screen for Article Display.

```
                    TOC/DOC HELP MENU

          1 . Database contents

          2 . Search instructions

          3 . Record contents

          4 . Display formats

          5 . Copy requests

          6 . Command keys

          7 . Trouble-shooting

        Enter # to select:

H:HELP  Q:QUIT  P:PREVIOUS SCREEN  M:MAIN MENU
```

Fig. 8. Main Help Menu.

```
TOC/DOC                 Journal Title File              toc08

          Table of Contents from a Journal (Keyword Search)

   Enter significant words from journal titles;
    truncate with $.
   Examples:

              ieee transactions acous$

              transactions ieee acoustics

              journal microbiology


   ENTER SELECTION:  devices

H:HELP Q:QUIT P:PREVIOUS SCREEN M:MAIN MENU
```

Fig. 9. Search Screen for Journal Table of Contents.

```
        3 JOURNAL TITLE(S) FOUND IN TITLE FILE                  02/28/1989
   To refine your search enter ' P ' and search with more title
   words. After selecting a title, please be patient. The system is
   processing a number of hidden steps.

1    IEE PROCEEDINGS-I SOLID-STATE AND ELECTRON DEVICES
2    IEEE CIRCUITS AND DEVICES MAGAZINE
3    IEEE TRANSACTIONS ON ELECTRON DEVICES

Enter # <cr>=select, <cr>=more, T<cr>=top, S<cr>=new srch:
```

Fig. 10. Display of Results for Journal Search.

```
TY88              460 ARTICLE(S) FOUND IN TY88                  02/28/198?

1    IEEE TRANSACTIONS ON ELECTRON DEVICES
         VOL:36 NO: 1 1989  LOC:ENGR APEE
41   IEEE TRANSACTIONS ON ELECTRON DEVICES
         VOL:35 NO:12 1988  LOC:ENGR APEE
167  IEEE TRANSACTIONS ON ELECTRON DEVICES
         VOL:35 NO:11 1988  LOC:ENGR APEE
214  IEEE TRANSACTIONS ON ELECTRON DEVICES
         VOL:35 NO:10 1988  LOC:ENGR APEE
242  IEEE TRANSACTIONS ON ELECTRON DEVICES
         VOL:35 NO:7 1988   LOC:ENGR APEE
303  IEEE TRANSACTIONS ON ELECTRON DEVICES
         VOL:35 NO:9 1988   LOC:ENGR APEE
328  IEEE TRANSACTIONS ON ELECTRON DEVICES
         VOL:35 NO:8 1988   LOC:ENGR APEE
363  IEEE TRANSACTIONS ON ELECTRON DEVICES
         VOL:35 NO:6 1988   LOC:ENGR APEE
378  IEEE TRANSACTIONS ON ELECTRON DEVICES
         VOL:35 NO:5 1988   LOC:ENGR APEE
399  IEEE TRANSACTIONS ON ELECTRON DEVICES

Enter # <cr>=select, <cr>=more, T<cr>=top, C<cr>=chge yr, S=new srch:
```

Fig. 11. Display of Issues for a Journal.

termined that the coverage of SciSearch (which includes titles covered in Science Citation Index and ISI's Current Contents series) would give us the greatest coverage of our collection in a single source. The expense of tapes at that time restricted us to mounting only one source of data; it would be too costly to purchase tapes from several different vendors for what would amount to a slight increase in the number of titles covered. We compared lists of titles covered by various vendor tapes with our current subscriptions lists.

We estimated that we would use about 1,800 of the approximately 4,500 titles covered in SciSearch, but those 1,800 titles would cover at least 70 percent of the titles in our science and engineering libraries. The remainder of the titles were in the Humanities/Social Sciences library or were newsletters, indexes and abstracting services, obscure journals, or very narrow, highly specialized journals. The 1,800 + titles pulled from the SciSearch tapes would cover the major portion of our science and engineering journals.

We immediately confronted our first problem. Some of the librarians felt that we should load the entire SciSearch tape, rather than pulling off only those titles with current subscriptions. This was in conflict with the original concept—expanding the online catalog to include article access to those articles located physically at Caltech.

We did not intend replacing online searching through such services as Dialog, BRS, and STN with complete local coverage.

Our budget could not support such activity locally, nor did we consider it desirable to duplicate available commercial services. We continued with the original concept of loading locally tailored files, but some of our staff still feel we should be loading the entire set of data.

In 1985 we started discussions with the Institute for Scientific Information about the use of the SciSearch tapes. At that time it was a relatively new use of such tapes. They had several customers using the tapes for batch SDI services, and one previous customer mounted a Current Contents file online (Washington University–St. Louis Biomedical Library). In the meantime we had purchased SIMS, an integrated library system from Sedna Corporation. SIMS is a database management system which, at that time, could be installed as a semi-turnkey system, replicating an existing installation, or in a customized fashion, where a library could do all its own file and screen design. We chose the latter option and began installation in April 1985.

We were occupied with starting up our SIMS system, bringing up an early version of the acquisitions module in October 1985. In the meantime discussions with the Institute for Scientific Information lagged behind because we were concentrating on starting the integrated library system. We installed a prototype of the online catalog for use in January 1986. Later in 1985 we received a sample tape to examine from the Institute for Scientific Information so that we could determine what was necessary in order to use the data locally.

DELAYS IN IMPLEMENTATION

We had numerous hardware problems with our integrated library system (which was running on a DEC PDP 11/44) and also encountered software problems. It was taking an inordinately long time to load and index our OCLC tapes. We therefore did not actually sign a contract with ISI until late in 1986 to start our tape subscription.

Due to continued problems with our integrated library system, we twice had to ask ISI to suspend the subscription; we had to concentrate on stabilizing the integrated library system and getting our OCLC tapes in for the online catalog. We finally picked up the subscription again in October 1987. We brought up a test version of the journal article file in November 1987 as part of the integrated library system.

We had no problem using the data on the ISI tapes, but we continued to have problems with SIMS. We had not yet finished indexing the call numbers nor were the OCLC tapes up-to-date. We could only have the online catalog up Monday through Friday 8 a.m.–5 p.m., because during evenings and weekends we were continuing the loading and indexing (both OCLC and Institute for Scientific Information tapes) and the processing development work. Moreover, the integrated library system could not yet handle keyword searches with Boolean logic for large files.

SEARCH FOR A SOLUTION

This situation disturbed the campus. We stopped loading tapes into the integrated library system in January 1988 and examined the situation. Two representatives from the library (the systems development officer and the head of technical services), a representative from the campus computing organization, and another nonlibrary computer expert were appointed to an Ad Hoc Library Automation Task Force to examine various aspects of the system. The library prepared documentation on what hardware and software enhancement would be necessary to bring the system up to par. There was no absolute guarantee of success, since much of the software work would have to be done by Sedna, with whom we had previous problems of missed deadlines and schedules. The task force examined the current marketplace and produced a list of potential vendors from whom a replacement system could be purchased.

The campus computing organization also examined various database management software that might be used to fully implement the journal access. The campus administration felt strongly that this was the single most important part of the library system and that a usable product should be available to the campus no later than fall 1988. The campus had Ingres operating on a VAX. However, the VAX, on which Ingres was installed, was already heavily used, and the Ingres programmer estimated a minimum of six months of programming work would be needed, given the functional requirements, including Boolean searching.

The campus computing organization also looked at BASIS and at STAIRS. Both would have required investment in hardware as well as programming time. Campus computing organization recommended the purchase of BRS/Search software to run on an IBM 9375 which the campus already owned; the 9375 had very little usage at that time. Their reasons were that the machine was already available and they had examined the BRS software at Jet Propulsion Laboratory and found it suitable for our purpose. Moreover, librarians were already familiar with BRS capabilities from using the dial-up search services.

The library expressed reservations about separating out the contents service from the online catalog. Separate logons to two different machines would result, and secondly, there would be no tie to holdings records. On the other hand, it was doubtful that the integrated library system would be reliable and able to accommodate user needs. The IBM hardware and the BRS/Search software were both reliable, established products. In June 1988 the campus administration decided to phase out the current integrated library system, replace it with a better system, and purchase BRS/Search for immediate implementation of the citation access function. We would send out an RFP in the spring of 1989. If the new system could handle journal citations, the ISI files would be moved from the IBM to the new system. If it could not, we would work locally on an interface between the two systems.

Additional reasons for transferring the ISI files from the integrated library system to the IBM were questions of file size and disk space, as well as stability and Boolean searching. The PDP was capable of handling one year's citation data (without keyword indexing), which was the original size file planned. The campus administration expressed strong interest in having more than one year of data available; it was doubtful that our integrated library system could handle that, even if upgraded. We found ourselves looking initially at a possible five-to-seven-year file.

STARTING WITH BRS/SEARCH

Based on the June 1988 decision the campus computing organization immediately purchased BRS/Search software. They also purchased additional disk space for the IBM 9375 and arranged for BRS to provide training on-site at Caltech. A management committee consisting of the manager of the campus computing organization, the library's systems development officer, and the head of reader services was appointed by the vice-provost to oversee the new project. The campus computing organization also arranged purchase of back years of SciSearch from ISI, although we were still not sure of how many years would be loaded.

Two members of the staff, one from the library and one from the campus computing organization, went back East to be trained in BRS' MNS menu programming language for BRS/Search.

The project was a joint library-CCO effort and had a target date of no later than November 1988 for availability of the online service. When the contents file was initially envisioned as representing an extension of the catalog holdings, the library planned to link the document delivery service to the contents file. This idea was carried over to the new project. The service was named TOC/DOC for Table of Contents/Document Delivery.

The library decided not to use the native mode in BRS for this service, since that would involve far more documentation and training for users. Instead, we would program our own menu interface using BRS' MNS language. BRS recommends that BRS/Search is used for six months before attempting to use MNS, but our time schedule did not allow for this. We had four-and-a-half months to get the contents/document delivery service online to the campus community.

Three things allowed us to succeed in this endeavor. First, our previous experience with loading the Institute for Scientific Information tapes into SIMS gave us a familiarity with the ISI format and with indexing loading decisions; second, the programmer assigned by the campus computing organization to the project quickly gained expertise with MNS; and third, various library staff were formed into subcommittees to spread out the workload and increase staff participation in the project.

DESIGNING THE DATABASE

Although we had previous experience with the Institute for Scientific Information tapes, we had not used all the information available on the integrated library system version. The first subcommittee looked at the fields available and made recommendations on treatment of the data in the context of the BRS software.

For example, some articles contain voluminous authors: one record is an article with a listing of over 180 authors. In SIMS we made the decision, during the loading process, to stop indexing after the tenth author. In BRS we decided to index all authors but only display some of them. BRS allows the option of double-posting hyphenated words, that is, indexing words both as individual words and as a single term. We chose this option in all fields.

Here we encountered a factor which was not present in the integrated library system, since it could not do keyword searching. More terms are hyphenated in the article titles as they appear on the tapes than in the original title; this is due to the permuterm indexing that ISI does for their products. This meant an increase in redundancy if we double-posted; if we did not double-post, we would lose certain kinds of access. We chose the redundancy over the loss in access.

BRS allows for a default operator for searches, such as OR and AND. We chose to set the default operator to SAME, which means terms not otherwise described must occur in the same sentence of the same field. Since each piece of data was stored as a separate field, e.g., the title of a journal is one field, this would work best for us. Operators such as OR or ADJ would override the default when specified by the searcher.

The exception was the author search. We decided that the author search would be an automatic ROOT function; "GOOD" would retrieve Good, Goode, Goodman, etc. "GOOD-" would retrieve anyone whose last name was Good no matter what their initials. Truncation could still be used as needed within the author field. However, if someone wished to search the occurrence of two authors together, they would need to specify the AND operator rather than let the SAME default apply, since we treated each author as a separate sentence within the field.

We also decided to use the reprint data and the reprint address data, something we did not attempt in the integrated library system version. Working with this part of the data was new to us. The BRS/Search software also allowed us to index data which was previously treated as descriptive and not as searchable, such as language, document type, number of references, update date, and so on. Some of these items are treated as indexed fields; some are used with the LIMIT command. For example, it is possible to search for any occurrence of an author who cites more than thirty references.

All decisions about the database structure and field handling had to be completed by September 9, 1988. The programmer took these decisions and wrote a program in FORTRAN to convert the ISI data into BRS load format.

CONTROLLING THE TITLES LOADED

She also wrote into this program the necessary steps for pulling off only those titles which we owned. Originally we had a title list in our integrated library system of those ISI titles which we owned, plus those titles showing up on the tapes which we did not own flagged with a NO (for not loaded). One of the subcommittees in the library upgraded this data. First, they reviewed the file for updates. Second, they added holding locations and certain kinds of notes. The holding locations would reduce the workload for the staff, since for most titles they would not need to look up the location in the serials list. Notes indicated recently cancelled items or ISI stopped indexing after a certain point, etc. We then produced a tape from our PDP which allowed us to transfer the data into the IBM. The programmer could then use this file for checking which titles to load. This file is referred to as the 'control file.'

Regular maintenance of the control file is necessary. If we add a new subscription, we must check to see if it is listed in SciSearch. If we cancel, we must update the control file accordingly. ISI sends us regular weekly updates to their listings, such as new titles added, title changes, ceased, etc. We continued to maintain the control file on the integrated library system through December. We found we could not easily edit it on the IBM because the file was too large to work with XEDIT, the text editor on the 9375. We therefore did a final transfer of the file to a VAX. The serials librarian who manages the file uses the VI editor to make changes to the file on the VAX. Whenever we want to update the control file on the IBM 9375, we use the FTP function to do a direct online transfer of the entire file from the VAX to the IBM. This process takes seconds since FTP is very fast.

Staff were concerned about missing titles on the ISI tapes. Some felt that there was insufficient time to check all our serial files (still in manual form) against the ISI lists, especially since our internal entries did not necessarily match those on the ISI list. We ended up creating another file, called the "exception file." This file contains titles which were missed on the initial run. As each tape was loaded, a list of titles was printed which were not loaded. These lists were checked by the librarians, and titles identified as "missed" were entered in an exception file. The tapes were then run a second time against the exception file. This, of course, increases the time needed for loading, as well as the time involved in working with the control file and the lists. But the librarians felt strongly that there should be a fail-safe for missed titles.

The control file actually exists on the IBM in two forms. One is used for matching against the ISI tapes during the conversion process. The other version exists in the BRS databases as part of the TOC/DOC function. In our implementation of BRS/Search each article is a document. Therefore, any index term points to documents rather than to journals. In order to have access to tables of contents at the issue level, we had to work out a strategy for bringing documents together other than through a journal title search. We could have done this as part of the document database, but it would have involved redundant indexing, that is, both keyword indexing of the journal title and a second field with the journal title and issue information treated as one term.

Instead we decided to create a separate file of journal titles. We used the existing control file for this purpose. A search for a journal title actually searches the control file (called TC88) and produces a list of journal titles (figures 9,10). Users then select the title they want to see. At this point the system automatically takes the journal title selected and switches databases to TYxx (where xx is the year, e.g., TY89) and performs a search which lists the specific issues involved (figure 11). From the list of issues they can then select a single issue to display the contents.

BALANCING DECISIONS AGAINST EACH OTHER

This design has several implications. It means that if we cancel a title, we must leave it in the control file (and flag it as a "no load"), since the control file controls journal title searching. It also adds overhead in response time, since the system must switch database files from TC to TY. Response time became a concern early on in the design phase. Although the software has very good response time, overhead in the menu system was escalating: first, the database switch; second, each journal title could be up to 255 characters long; and third, we were doing a sort for the issue display. BRS has the choice of FIFO or LIFO. If we chose FIFO, the issues would display from oldest to newest, since the oldest were loaded first. The contents would display in proper page number order, from lower to higher numbers. We therefore did a sort of the issues to display the most recent issue first. Moreover, the screen formatting part of the program added additional overhead. This was leading to unacceptably long search times for some entries.

As a result, we made several changes. We switched our BRS default to LIFO. This meant that the most recent issues, as the last in, showed up first. However, the page numbers of the contents were in reverse order, from highest to lowest. We discussed the problem of order of displays versus sorting versus response time with several faculty members. They expressed a willingness to have the page numbers in reverse order if it meant issues in reverse chronological order and no sorts to delay response. We decided on that change and on changing the searching method.

Early on, the programmer suggested using the eleven-character maximum ISI abbreviation as the matching mechanism for the tape conversion program. Eleven characters is much faster to match on than up to 255 characters. However, we disregarded that idea because we did not have the abbreviations in our locally generated control file. However, when the problem with response time arose, she recommended using that as a hidden search mechanism. In other words, when the user selects a journal title from the initial display of journals, the system takes the title abbreviation rather than the full title to match against the document database in order to produce a list of issues. We experimented with this in native mode. Based on what we discovered in searching, we decided to follow this recommendation. We entered all the necessary abbreviations into the control file, and we checked various source documents to obtain most of the abbreviations. For the last few we had to call ISI to find abbreviations we could not locate locally.

With these changes—dropping the sort of issues and switching to LIFO and changing to hidden abbreviation searching to retrieve a list of journal issues—we reduced the search time substantially. Our problem searches had been taking up to two-and-a-half minutes; this was reduced to under fifteen seconds.

We knew in advance that we would also have a problem with single-title word journals such as *Science* and *Cell*. Since many journal titles have the same words as part of their titles, a lot of false hits would likely occur. We created a special field for one-word journal titles. One-word journals are a separate option on the main menu. This option searches against only that special field, eliminating the problem of false hits and also the potential problem of response time if someone does a search on the single word "SCIENCE." A user can still search these titles in the regular journal title search, but they will encounter many false hits if they do so.

This is one example of the reiterative process that took place during the implementation period. Although we tried to make as many decisions upfront as possible, we did encounter problems along the way which caused us to change decisions. Of course, each change was weighed carefully to see how it would affect the system as a whole. Some erroneous decisions could not be reversed because they would entail a reload of the entire file of data. For example, over time, ISI has made changes to the data available on their tapes by adding a field, rather than by replacing the existing data structure.

In some cases, this allows a user to continue to use old programming if they are unable to change. On the other hand, it does mean keeping track of when changes are made. We made an outright error here, although not a serious one.

HOW NOT TO LOAD YOUR DATA

The language data on the ISI tapes originally was in a two-letter code which appeared in front of the article title when the title was translated, e.g., FR>> meant the article title had originally been in French and had been translated into English by ISI. They later added a separate field carrying that code and a separate field carrying the language spelled out in full. The earlier part of our 1988 tapes was in the old format, with the new format change taking place later in the year. The backfile we purchased was done in the new format. When we did our design for the database and the load program, we picked up the data from the new fields, forgetting that in the tapes for the earlier part of 1988 these fields did not exist. Part, but not all, of our files allow for restricting searches by language.

The original specifications we worked up for examining the use of Ingres allowed for the language being in the title field, but that item was dropped out after the decision to use BRS was made. In retrospect, if we had remembered sooner about the change, we could have picked up the language code in the earlier 1988 tapes from the title field. However, it turns out that when we examined the article titles for January through September 1988, fewer than 7 percent of the article titles were in a language other than English and thus a restriction of one's search to English language articles is not that useful.

A similar problem exists with the document type (e.g., article, letter, editorial, etc.). We programmed it to pick up the type in full in the new format, but in the earlier 1988 tapes it existed only in a coded form in a different field. This error is more harmful, since the document type can be a very useful aid in determining the value of an item listed. For example, an editorial may not be as useful as a regular article.

HARDWARE/SOFTWARE CONSTRAINTS AFFECT DESIGN

Another problem we encountered was the physical size of the files. We decided to place each year in a separate file. Our experience shows that we load approximately 45 percent of the data on the ISI tapes, and this adds up to around 380,000 article citations, which take up over 400 MB of storage for each year (401 MB for 1987, 435 MB for 1988). To load all the years into a single file was not possible, given our particular configuration of hardware and software. We therefore split the files up according to tape year. We started with 1988, which is now completed; and 1987 is now loaded. In February we started loading our 1989 tapes.

The difficulty with this system is that tape year does not reflect publication year. An issue which carries 1987 on its cover may actually appear in either the 1988 or the 1989 file, depending upon when it was actually published. This is a little difficult to explain to users. One suggested solution was to sort and load the data by publication year of the issue. Unfortunately, this is not possible in our environment, because it means keeping all database years as open files, which would require a sizable increase in disk space. As of now, 1987 and 1988 are treated as closed files.

WORKING WITH MENU DESIGN

The second subcommittee created the logical sequence for the menus, designating wording of the screens, options available, and the help text. We have two kinds of help: contextual tied to the screen and a help menu which leads to more extensive online documentation. The committee needed their work completed by early October to allow sufficient time for programming and testing. Certain recommendations were discussed with them, and adjustments were made. Either the logical path requested did not work well within the context of MNS, or it would be too slow.

For example, if an issue contains 170 articles, it takes many screens to display the list of articles. The committee wanted the ability to scroll back and forth between pages of a single issue. This function was programmed and tested and ended up being inordinately slow. It was actually faster to return to the beginning of the list and page forward, a function which is natural in MNS. It can take only a few seconds to page forward to the fourth screen or even more. Scrolling backwards and forwards is far more cumbersome, because it involves writing out temporary files to disk. Once again a reiterative process occurred as we learned more about how MNS worked for our particular needs.

GREMLINS

A far more serious problem we faced was the corruption of the database. We noticed part of the way through the 1988 database that search results were inconsistent. The indexes were becoming corrupted. We started working on the problem with BRS, but we could not discover the source of the corruption. Presently, we use an IBM 9375 with VM/SP release 5 operating system and 9335 disk drives. Although other BRS sites use the VM version of the software, we are the only site with this particular combination of operating system, CPU type, and disk drive type. We thought there might be a bug in the indexing program or in the reorganization program (which helps maintain the indexes). BRS suggested a lack of disk space. The reorganization creates a temporary working file equal in size to the original file, which means twice the disk space which you would normally need. We were not getting any error messages from the loads or the reorganizations.

We worked with BRS, but the source of the problem could not be identified and isolated (no other BRS VM site was encountering this problem). We finally arranged to send the converted ISI records to BRS for them to load and index on their machines (similar to their BRS/ON-SITE service) and send us complete database tapes which could be loaded onto our machine. We made arrangements to load 1988 and 1987 in this way, and we are loading 1989 locally, trying out various methods to avoid corruption of the files again.

First, when we convert the ISI data into BRS format, we split the data into load files of 500 records each, since it is better to use smaller files for batch loading into BRS. Second, we do a reorganization every four files (2,000 records) following BRS' recommendation. We have also restructured the disk space to leave more empty space on each disk. Finally, we do not update the public database directly. Although one would think it would be safe to load into a database which people are only searching and not editing, we cannot be certain that this is the case. We therefore load the records into a working file. Once a week we take over the system at the noon hour to copy the updated working file into the public version of the database. Hopefully, this combination of changes to our procedures will prevent further occurrences of index corruption. The working file is carefully tested after each update before we load it into the public files. BRS continues to investigate the source of the problem.

TRAINING AND DOCUMENTATION

We divided our staff training into three groups. First, ten staff members attended BRS training. Second, we held training sessions in native-mode searching for other staff members using documents we prepared locally, covering major aspects of searching but omitting more detailed or esoteric functions covered in the BRS' manuals. Third, we provided demonstrations of the menu system to all staff members once it was completed but prior to the November public release of the system.

Our goal was to ensure that all staff at public service points could field questions and assist users with the system. Although our intent was to make the menu system as self-explanatory as possible, we knew some questions would remain. The training effort was lead by the head of reader services, with the assistance of another subcommittee. Emphasis was placed on understanding the coverage of the files and the limitations in coverage, as well as the logical paths of the menus.

Public release was initiated by the Nov. 21 Faculty Board Minutes, which contained a brief handout on TOC/DOC explaining to the faculty how to access the system. Our intent was to use a limited group of patrons, the faculty, to test out the system in actual use, since library staff were too knowledgable to be good test subjects. As problems were resolved, we would advertise the service on a wider scope. The target date was March 1989. The Head of Reader Services prepared a brochure to be distributed to all faculty and students at Caltech. Articles appeared in the campus newspapers. A series of demonstrations of the system was scheduled for faculty, students, and staff to attend at their convenience.

Users can access an option in the system called "Feedback" to enter comments and questions about the system online. With the comments from this option and questions received, we made some changes to the menus. Some of the changes are cosmetic in nature, in terms of wording.

PROBLEMS USERS HAVE

The overall response to TOC/DOC is highly favorable. A few comments are requests for more titles or more years online. However, the vast majority of questions and complaints pertain to getting on the system. CCO is handling questions about network connections, since they are responsible for maintenance of CITNET and installation of network hardware. We were surprised at the number of people at

Caltech who were not yet on the network after five years.

When we first set up the system we discussed such matters as automatic timeouts if people left a terminal or network port logged on. We decided this was not a desirable feature. We also decided to avoid having people log on to the IBM, since this involves a separate protocol which causes them to be in the IBM's operating system until logon is completed. Ten public network ports for TOC/DOC were set up as "slave ports." These ports are continuously logged on. When a user connects to a TOC/DOC network port, they are dropped into a slave port on the IBM which is open to whatever screen the previous user left up. Users are asked to return to a "Home Screen" before disconnecting from the network port.

The use of slave ports results in three different problems. One is that we require VT100 emulation. Although the IBM can handle other kinds of terminal emulation, we only had the time and resources to set up TOC/DOC for one terminal type. Many of our users don't know if they can emulate VT100s or not; they also may have VT100 emulators which are not 100 percent compatible. We set up a standard using Kermit. Both the library and CCO supply floppy disks called TOCDISK, which has Kermit with customized keys. Use of the TOCDISK allows people to logon to TOC/DOC with a single macro command and ensures 100 percent compatibility.

The second problem was a direct result of slave ports being continuously logged on. Each account is assigned a working space (their A disk). Since logoff never occurs, the A disk accumulates and gets full, preventing further searches from taking place on that slave port. We therefore have to regularly go in to check the disk space in each TOC port and clear it if it is over a certain percent full. The third problem is that people often do not go to the Home Screen prior to disconnecting. We therefore also check the ports and return them to the Home Screen when necessary.

Despite these problems, we still think this is preferable to the overhead of setting up individual accounts, training people in the logon process for the IBM and BRS software, and worrying about what may happen if they get themselves lost in the IBM operating system. However, users with SUN microcomputers are assigned individual accounts, since for technical reasons SUNs cannot use the slave ports.

Our own staff also has access problems. We have various terminals in the library (Teleray, WYSE, VT), and the Telerays are not 100 percent VT100 compatible. They can be used with TOC/DOC, with the proviso that the backspace key does not operate as usual. What still confuses people is switching back and forth on the network between our library system and TOC/DOC. Our online catalog operates in block mode; the network expects character mode. We found that people, even with tailored documentation, had trouble learning to change settings on the Telerays. We set up function keys to change settings as needed. However, people still have difficulty with this. The WYSE and VT103 terminals are new to the library; users experience similar difficulties with them.

Most of our troubleshooting time is therefore spent on the technical/physical aspect of connecting to TOC/DOC. Once they are in the system, users experience few problems. Although Caltech is a very science-oriented campus, there are still many levels of faculty and staff expertise with respect to the network, terminals, and microcomputers. The students have the fewest problems.

FUTURE PLANS

Since we originally had only four-and-one-half months to bring the service up, we tailored our original plans to the time available and divided the specifications into Phase I and Phase II. Phase I was successfully completed with the basic contents/document delivery service online. We are now beginning the work on Phase II. Phase II includes the following enhancements:

- setting up profiles to produce SDI printouts for users (this is already available in BRS/Search but we need to determine the best way to implement it locally),
- setting up a blank form for entering document requests for items not in the files (both for document delivery and ILL), and
- setting up an accounting system for

photocopy billing to be produced automatically.

We will also look at other possible enhancements, such as adding more types of searches to the menu system (e.g., corporate source, document type), adding more titles, and training some users in native mode.

The TOC/DOC system has been very successful to date. Despite the short time available for the implementation of BRS/Search, we were able to accomplish our charge from the administration to bring up a journal citation access module by Fall 1988. We intend to improve on this achievement with the enhancements planned for Phase II. Our ultimate goal is to provide our users with a one-stop workstation for access to information. As our future plans develop, we intend that network stations will allow users to access (a), information physically available on campus such as the online catalog and citation access, and (b), information available from other sources through gateways and ILL services. With the success of TOC/DOC and the procurement of a new ILS, we will be able to move from the first goal of local access to the second goal of access without regard to physical location of information.

ACKNOWLEDGEMENTS

Many computing and library staff participated in the TOC/DOC project. Particular credit should go to Edith Huang, who programmed the menu system. Library staff outside Systems who contributed heavily to the project include Kimberly Douglas (head of reader services), Don Wassink (head of cataloging), Jim O'Donnell (geology librarian), and Mary Ito (serials librarian). A complete list of staff who participated is too long to mention here. ■■

Locally Loaded Databases in Arizona State University's Online Catalog Using the CARL System

George S. Machovec

Arizona State University installed the Tandem-based CARL integrated library system in 1987. This software supports all the traditional functions of an integrated system plus the ability to act as a platform for creating or loading multiple databases. The functionality of CARL software for this purpose is described, a brief rundown of databases loaded at ASU is provided, and the criteria for selecting databases is given. Public response and library instruction issues for locally loaded databases are briefly examined as is the future of the electronic library.

THE SETTING AT ARIZONA STATE UNIVERSITY

Arizona State University (ASU) is an emerging research university with over 43,000 students, of which 11,000 are graduate students. This makes ASU the fifth largest campus, in terms of student population, in the United States. The main campus is located in Tempe, Arizona, a part of the Phoenix metropolitan area, and due to the rapidly expanding community, a branch campus, called ASU West, was opened approximately thirty-five miles away in the northwest section of the city.

ASU Libraries have a collection of over 2.5 million volumes and 32,000 serial subscriptions in the main Hayden Library and five branches. The libraries have an acquisitions budget of over $3.6 million on the main campus and an additional $1.4 million at the new ASU West site.

BRIEF HISTORY AT ASU

In the early 1980s Arizona State University installed its first integrated library system, which was developed by Dataphase, Inc., and was called ALIS III (Automated Library Information System). By 1986 Dataphase went out of business and sold this product line to Utlas, Inc., and it was renamed Utlas T/Series 50. By 1986 this system had grown to four Non-Stop II CPUs with 128 terminals supporting circulation, bibliographic maintenance, and an online public access catalog.

By early 1986, although the Utlas T/Series 50 offered a superb circulation module, the online public access portion of the system received little development from the vendor, and a number of system bugs and limitations had never been resolved. During the fall of 1986, a task force was created to consider the pros and cons of remaining with the Utlas T/Series 50 (ALIS) integrated library system or changing to new software that would operate on existing Tandem hardware. Although several libraries around the country (University of

George S. Machovec is Head, Library Technology and Systems, Arizona State University, Tempe.

California at Berkeley and the Triangle Research Library Network, which consists of Duke, North Carolina State, and the University of North Carolina at Chapel Hill) have been developing integrated library software for the Tandem minicomputer, only one package was in an advanced state of development at the time—the CARL software that was developed by the Colorado Alliance of Research Libraries.[1,2,3]

SELECTION PROCESS AND SYSTEM MIGRATION AT ASU

During the spring of 1987 two integrated library systems were the leading contenders for the future of ASU: the Utlas T/Series 50 (which was already installed) or the CARL software.

A library automation committee was formed to examine each module of the CARL software and compare it with the existing Utlas T/Series 50. A test version of the CARL software was locally loaded, and side-by-side comparisons were made by staff and selected students who were willing to offer opinions. In addition, the computing services staff at ASU examined the CARL software in terms of its design to ensure its compatibility with existing hardware and its ability to handle the network load. After much discussion the CARL system was chosen to replace ALIS for several reasons:

1. CARL is a more efficient system written in TAL (Tandem Application Language) rather than COBOL (the language used for Utlas T/Series 50), thus providing improved system performance. This improved performance is reflected in faster terminal response time, a savings in disk space on the Tandem, and the ability to run more terminals with the same number of CPUs. With CARL, ASU has been able to drive over twice as many terminals with the same number of CPUs.

2. The ability to load multiple databases on CARL gives the option of loading our MARC bibliographic database as well as our locally produced indexes for various special collections, indexing/abstracting services, full-text information tools, directories, etc.

3. CARL is very user friendly and requires much less library instruction than the previous system. In addition the system offers a menu approach for the novice and advanced command-driven search features for the experienced user.

4. Dial-in access is supported by CARL software.

5. CARL allows for easy connectivity with other computer systems. For example, the ability to interface with the ASU campus broadband network opened up many new channels for faculty and students to use the libraries' computer. The system also allows the option to gateway with other commercial vendors or online catalogs.

6. Our contract with CARL gives us direct access to the TAL code so that if we want to make changes or enhancements to the software that option is possible.

7. An online bulletin board allows for many options in presenting information to users. This feature is used for such applications as library hours, policies and procedures, materials locator, system news, library locations and phone numbers, and descriptions of special services.

8. CARL has a demonstrated record of high reliability, whereas ALIS was notorious for terminal crashes due to depressing inappropriate keys. (These problems were never fixed after many formal requests.)

9. The CARL search engine provides many new access points not previously available, such as call-number browsing, series browsing, and the indexing of virtually every word in the MARC record for searching in the WORD or NAME mode.

10. CARL was available at a very reasonable fee through ASU becoming an associate member of the consortium. This arrangement offered all of the programming necessary to get ASU live on the system, software maintenance, software upgrades, and a voice in the future development in the software.

11. Since CARL was created in a multi-institution environment, it is a superb networking product which supports the individual circulation policies and MARC records of each participant while allowing the sharing of information in a networked setting.

12. Flexible and sophisticated transaction monitoring, as well as management and collection analysis reports, were available.

13. CARL provides the ability to process

searches with large set results (i.e., searches involving common but important words such as "history," "United States," and "university"). Many online public-access catalogs either cannot process terms with high postings or have poor response time.

CURRENT SYSTEM CONFIGURATION

As of May 1989 the ASU online catalog has the following system configuration:

1. Four Tandem Non-Stop II Central Processing Units (CPUs).
2. Two-hundred-sixty-four ports, of which twenty are allocated for connection to the campus broadband network (through which dial-up is possible, plus use from faculty offices and remote computing sites on campus).
3. A 9600-baud leased line link to the CARL system in Denver, Colorado. This is used for software updates plus can be used by the public as a gateway to the Uncover database being produced by CARL. (Uncover is a table of contents indexing service which indexes 10,000 periodical titles.)
4. Ten gigabytes of Winchester hard-disk storage is in use with additional growth planned for the near future.

CARL AS A PLATFORM FOR MULTIPLE DATABASES

The major modules that exist in the CARL software include circulation, bibliographic maintenance, public access catalog (PAC), serials control, electronic mail, and acquisitions. ASU is using all of the above modules except acquisitions and only a limited portion of the serials module.

In the PAC module, the patron is presented with a menu of database options from which to choose. Once a database is selected a welcome screen is displayed which gives a brief description of the database and the searching options for that file. (See Appendix A for a sample welcome screen.)

Since each database is searchable in CARL with the same search software as the MARC bibliographic records, it is very easy for a patron to switch from one database to another without having to learn different search protocols for each file. This is in contrast to some institutions that have locally loading databases on a separate minicomputer or mainframe using different search software than the integrated library system. The same problems exist for CD-ROM databases.

Another nice feature of the CARL software is the ability to transfer searches from one database to another without rekeying the strategy. This allows moving from file to file identifying relevant information. One problem which librarians and patrons must keep in mind when doing this is that different databases may use different controlled vocabulary, so that optimal results may not be obtained from each file. This is a library instruction problem which will need to be taught to patrons and has been long discussed in the literature in reference to multifile searching in databases through commercial vendors.[4]

CARL software supports both a menu- and command-driven modes for novice and experienced users. The command mode allows any type of search to be executed from virtually any screen, offering quick navigation through any search.

To load a database into the CARL system, the file must be mapped into a MARC format. At ASU a wide variety of databases have been loaded including regular OCLC/MARC records, periodical indexes from H. W. Wilson, full-text files (an encyclopedia), and several directory databases. A numeric file with statistical information is planned. Since each foreign database is stored as a separate entity, there is no compromise of the MARC records which have been cataloged on OCLC, RLIN, or other utilities. Almost any type of database may be mapped into a MARC-like structure, and the MARC tags merely serve as fields which determine how each record is displayed (MARC tags are displayed to the patron in tag order) and how each record is indexed (the MARC tags determine which fields will be indexed and in which index).[5]

CARL supports full keyword searching and has five basic indexes. Each library or database supported in the CARL software can determine how indexing will occur. The five basic indexes include:

- NAME—usually indexes authors, corporate names, or personal name subjects;
- WORD—usually indexes keywords from titles, subject headings, notes, or abstracts;

• TITLE BROWSE—indexes title strings and supports exact title searching;
• CALL NUMBER BROWSE—indexes call numbers; and
• SERIES BROWSE—indexes series.

In each of these cases the meanings of the above indexes can be modified or altered to fit the nature of the database being loaded or created.

Each database which is built by hand or requires maintenance has its own maintenance module in the CARL software. A general system screen (called newpex) used by library staff (several levels of passwording are available for security) allows staff to move into the various modules of the system including each of the maintenance modules. (See Appendix B for a sample screen of this menu at ASU.)

LOCALLY LOADED DATABASES AT ASU

One of the most exciting aspects of having an integrated library system which supports the loading or local creation of multiple databases is that the entire library staff and user community generate and suggest many wonderful ideas. As of June 1989 ASU has sixteen databases available for public use although another ten are in various stages of development. Some of the files are commercially produced and tapeloaded, some were transfers from other machine-readable files created in different departments of the library, and some of the databases are being hand-built through the CARL maintenance module.

Appendix C provides a sample menu screen which a patron may view when using the system at ASU. A brief description of each of the files is as follows:

1. General Catalog. All of the OCLC-cataloged MARC records are located in this database. At the present this includes over 1.3 million full MARC records, and the ASU libraries are approximately 95 percent converted. The majority of missing records are CJK (Chinese/Japanese/Korean).

2. News and Help. A submenu of read-only files are available to provide users with all sorts of general information, such as library hours, more detailed explanations of what is in certain databases, library code of conduct, and so forth.

3. Encyclopedia. This is the full text of the *Academic American Encyclopedia*, which consists of approximately 30,000 articles and is updated twice per year.

4. The Song Index. This database is locally produced by the ASU Music Library and indexes collected works of music. Each citation includes the composer, librettist, title of the song, first line of the song, and the title and call number of the collected work from which it came.

5. Career Services: Company. This database indexes over 2,000 company names and addresses for which the Career Services department on campus has information. This file was created in this nonlibrary department through terminals paid for by Career Services. As a nonbibliographic directory-type database this file provides information to students who are looking for jobs. The file includes when each company is interviewing on campus and what kind of information Career Services has on each company (i.e., videotapes, annual reports, benefits information, etc).

6. Career Services: Calendar. This database lists orientations, workshops, special events, jobs, advisement, and a recruiting master list for the Career Services.

7. Career Services: Books. The cataloged monographic collection from Career Services.

8. Humanities Index. This H. W. Wilson database covers 295 English-language humanities journals from late 1983 to the present.

9. Applied Science & Technology Index. This H. W. Wilson database covers 335 English-language journals in engineering and the applied sciences from 1983 to the present.

10. Social Sciences Index. This H. W. Wilson database indexes 300 English-language periodicals in the social sciences from early 1984 to the present.

11. General Science Index. This H. W. Wilson database indexes 111 English-language periodicals in the physical, life, and health sciences from 1984 to the present.

12. Education Index. This H. W. Wilson database indexes 353 English-language periodicals in all areas of education from late 1983 to the present.

13. Business Periodicals Index. This H.

W. Wilson database indexes over 300 English-language business journals from 1982 to the present.

14. Uncover. This database is not stored locally but is available via a transparent gateway between the ASU CARL system and the CARL system located in Denver, Colorado. The database provides access to title-level information on approximately 10,000 journals. The database can be searched using keywords, or a user can recreate a table of contents for a specific issue of a periodical.[6,7]

15. Map Index. The ASU Map Collection has a collection of over 160,000 maps. This collection had been indexed in a batch KWOC (keyword-out-of-context) system since the early 1970s. The file has been transferred to the ASU CARL system as a separate database and is now maintained online. Each map has very extensive indexing, and it is not unusual for a particular entry to have more than twenty-five subject headings.

16. Solar Energy Index. ASU has one of the most complete and comprehensive special collections on solar and alternative energy sources in the world. This locally produced file indexes journal articles, archive and manuscript materials, technical reports, and other pamphlets. Much of the material found in this collection is unique and not found elsewhere, making this file an invaluable national resource.

A number of other files are under development and will become operational over the upcoming year. Some of these include: Arizona Index (a locally produced file of state and local documents); the *GPO Monthly Catalog* (this will include over 250,000 MARC records that were obtained from MARCIVE, Inc.); *Grants Database* (Oryx Press); Southwest Index (locally produced index of manuscripts, archives, and photographs); *ASU Newspaper Index*; a collection of AMC MARC records from libraries around Arizona; a numeric statistical file; and others.

CRITERIA FOR DATABASE SELECTION

The creation or transferring of locally produced databases at ASU into the environment of the online catalog was one of the reasons for acquiring the CARL software. In this way, many of the special collections in the campus libraries could finally have equal availability and exposure as the main OCLC-cataloged collection. Special collections and units in the libraries are no longer treated as an afterthought or some distant goal on a planning document.

The leasing or purchase of commercially produced databases on magnetic tape has been an area of high interest. With the thousands of commercially produced bibliographic and numeric databases, there needs to be a way to narrow the field to a few select files from which to start.

To approach this issue, the head of library technology and systems began by collecting information on the availability, pricing, storage requirements, and licensing agreements for about fifty databases that were candidates. These fifty databases were initially selected based on frequently used printed indexes at the reference desks and on ten years of online searching statistics. Meant only as a starting point, it was realized that there may be some other databases which did not appear in the initial analysis but may be prime candidates for local loading.

This list of databases was brought to the Public Services Council at ASU Libraries, a group of public service department heads, who then developed a list of criteria for selecting databases from this list. The criteria have proved to be a valuable planning and management tool, not only for the initial selection of databases but also for long-range prospects. The criteria are as follows:[8]

1. Cost. The cost of purchasing or leasing a magnetic tape database was of very high importance. It was important that the database producer offer a leasing arrangement based on a flat annual fee and not include any connect time or hit charges. Another consideration was the cost of any backfiles. How much do backfiles cost? Are the backfiles paid for once, or must they be paid for again year after year? Obviously, a publisher that offers a one-time payment for backfile leasing was preferable.

2. Disk Storage. In some cases the cost of database leasing is minor compared with the cost of Winchester disk drives to store the database! How large is the database in

megabytes, and how much overhead will be required for indexing? Some of the most highly used databases were not considered for this reason alone, since the disk storage could not be afforded by ASU. Examples of these would include MEDLINE and *Chemical Abstracts*.

3. Backfile Availability. Is a backfile available for the database and if so how far into the past? Some topics need significant retrospective coverage, and this criteria may come into effect.

4. Size of Clientele Served. What clientele will a particular database serve? Since only a small number of databases are initially going to be loaded, it makes sense to load files which will have broad appeal.

5. Preeminence of Program Supported. Especially in academic and special libraries, the importance and prestige of certain users carries extra significance. Some departments on campus are targeted through the university administration for "excellence," and this may influence a decision.

6. Mapping the Database into the Local System. Some databases will be easier to map into the local system than others. Will the programming be done through local programmers or through the assistance from the vendor supplying the search software?

7. Searchability in Local Systems. What sophistication in search software is needed to successfully query the database under consideration? Depending on the sophistication of the local search software, some databases will perform better than others.

8. Workstations Used for Searching. Will dumb terminals or microcomputers (or a combination) be used as workstations for searching various locally loaded databases? Many dumb terminals do not support any graphics or non-Roman characters and may limit the loading of databases that have special requirements such as these.

9. Currency. How frequently is a database updated, and can the local institution keep up with regular database loads?

10. Multidisciplinary Coverage. If broad coverage for databases is desired, it may make sense to select initially databases that will meet a wide range of needs.

11. Overlap with CD-ROMs and Other Databases Locally Loaded. It does not usually make sense to purchase a database for local loading and then buy the same product on CD-ROM. In addition, does some other department in the institution already have other databases locally loaded? There are too many good databases to duplicate efforts.

12. Equality. Although it is impossible to purchase databases to meet everyone's needs, it does make sense to distribute the acquisition of databases across different disciplines to maintain some level of equity. A common complaint that is beginning to arise for locally loaded databases and CD-ROMs is, Why does that discipline get to have free searching, while we must still pay for searches through commercial search services?

13. Inexpensive Database Availability Through Commercial Search Services. Some databases are available at low regular rates through commercial search services, and it may not be economically practical to locally load them (e.g., MEDLINE). In some cases databases are available at great discounts to academic institutions (such as 90 percent off *Chemical Abstracts* through STN International), so that it again does not make sense to locally load them.

ISSUES ENCOUNTERED WITH DATABASE LOADING

As previously described, databases which are loaded into the CARL software must be mapped into a MARC-like format. Since most databases do not come tagged in this format, it is necessary to write a conversion program which will map each field from a record into a MARC tag. Thus authors become tag 100, titles become tag 245, subject headings may translate into tag 650, and so forth.

At ASU, the initial mapping for each database is usually done by the head of library technology and systems with assistance and recommendations from Original Cataloging as well as comments from public services staff. The comments from reference staff are critical since they are the librarians who know first-hand how a database may be used and what the patrons may want. This mapping process will determine what data is transferred, how it will be presented on the screen, and which fields will be indexed.

Some of the programming for databases loaded at ASU has been done by CARL staff in Denver as part of our associate member status. ASU also maintains its own applications programmer and systems programmer, who have been especially active in converting local databases into the CARL software.

For the loading of the H. W. Wilson indexes at ASU there was much healthy debate on the pros and cons of various indexing decisions. Eventually choices had to be made. Some people were not completely happy, but at least many issues were brought to light, and it gave the entire staff insight into the effects of indexing decisions.

Another problem with loading someone else's data is that the documentation is not always as clear as needed. This requires a close review of raw records and working with the database producer to resolve questions.

It has been found that the loading of a database is an iterative process. Once a mapping and indexing scheme has been developed, do NOT try and load the entire database. Rarely does everything work as planned on the first run. Load several hundred records, look at the results, and try again. Through several iterations it is usually possible to develop a tape loading program which will work with a reasonable level of confidence for the entire database. If library non-programmers are writing the specifications for tape loading, it is important to work closely with the programmers, since it is very easy to misinterpret what the librarians actually want.

DATABASE PRODUCER NEGOTIATIONS

Many database producers are just beginning to realize that a new market exists for the leasing of their databases for local loading. When ASU begin to contact database producers in mid-1987, it was discovered that many still had not considered this option and did not have a standard leasing fee or licensing agreements. This has been changing over the last two years, and a growing number of vendors now have regular price sheets and standardized agreements. As would be expected, there is little or no uniformity on pricing and legal contracts.

Vendors who still persist in very high leasing charges and demand connect hour fees and/or hit charges will lose out on this new library trend.

Database producers do have a number of legitimate concerns which must be addressed by a library before final agreements can be reached:[9]

- Migration away from the print, CD-ROM, or commercial search services. Database producers do not want to lose revenue, and their concern is that magnetic tape leasing may cause a drop in revenues from other sources.
- Dial-in and networking. Direct-dial ports into a library's computer system are a major concern for database producers who do not want free or low-cost access to their databases to undercut their files on commercial search services. Certainly the gatewaying of one integrated library system to another causes concerns since, if a database is linked with another system, users from different institutions are using it for free unless it is properly passworded or masked.
- Non-institutional walk-ins. Many publicly supported libraries are open to everyone, causing some of the same problems as dial-in fears.
- Most-favored nations clause. Some database producers have established contracts with commercial search services which require that any alternate royalty structure be offered to everyone. To get around this, the database producer must offer a modified tape product, lease to libraries on an "experimental" basis, or develop pricing strategies which have strict limitations.

Many databases producers and libraries are resolving some of the above issues, since both parties have a great deal to gain. Database producers who are willing to work creatively with libraries in establishing such contracts have increased potential for future earnings.

PUBLIC RESPONSE

In January 1988 the first database which was loaded aside from the cataloged MARC collection was Grolier's *Academic American Encyclopedia*. Although some of the library staff were reticent about offering this encyclopedia, it has proved to be a tremendous success. Almost a third of all online catalog users outside the library (faculty and students in their offices and remote computing sites are not near a printed encyclopedia) regularly use this database.

Since each subsequent database that has been loaded uses the same search engine, patrons have quickly taken to searching different files for different needs. The periodical indexes are very highly used, with patrons often asking for more.

As more databases have been added to the online catalog there has been an exponential growth in the demand for new terminals. When CARL was first implemented at ASU, 128 terminals were active. By June 1989, 264 terminals were in operation, with a demand for still more. Users seem to saturate every newly installed bank of terminals with no end in sight.

An added demand has been for printers to be attached to terminals. With the saturation of terminals by users and with such a variety of databases available, the lack of printers has become acute. Although the library has put four printers out for free public use, this has just fueled the demand for more. The major difficulty is funding, since it has been estimated that a quiet HP ThinkJet printer costs almost $1,000 per year to supply with ink cartridges and paper because of its heavy use. The library cannot afford to put printers on most terminals. The option of using debit-card boxes attached to printers is being explored, but the issue is not yet resolved.

Overall the public response to multiple databases loaded into the online catalog has been very positive. As more files are loaded greater expectations grow.

Dial-in ports are very heavily used, and faculty have been especially active in their use. The ability to search a truly "electronic library" from their offices has become very popular.

LIBRARY INSTRUCTION ISSUES

With the advent of locally loaded databases in the online catalog, CD-ROMs, and commercial fee-based search services there has been a growing need for quality instruction. This needs to be accomplished at several levels: librarians need to have increased expertise in the options available, and the library patron needs to be aware of the choices. Since ASU has over 43,000 students and 2,000 faculty, this is no easy task.[10]

Users must first be aware that multiple options for accessing databases exist and then learn how to use the specific tools when they are found.

Most reference librarians work ten to fifteen hours per week on the desk and are called upon to recommend printed tools, CD-ROMs, fee-based searches, and now databases in the online catalog. Commercial search services have traditionally been taught to librarians by the vendors themselves. In contrast, CD-ROM databases and locally loaded databases in the online catalog have required in-house training. This has primarily been done by a triad. The main reference department and the science reference department each have a coordinator of computer reference service who works closely with the head of library technology and systems. The training has primarily occurred at departmental meetings, informal small-group settings, and on a one-to-one basis.

Among the library patrons new databases have been advertised through special inserts in the campus newspaper, through local library publications, and by librarians as they give tours, orientations, and different types of library instruction. Dial-in access has spread very quickly by word-of-mouth, and the dial-in access brochure is one of the most heavily reprinted library publications.

Two diametrically opposed misconceptions sometimes occur with library patrons. Some patrons tend to believe that a computer is always correct and offers all possible information, not realizing the scope and limitations of databases they are searching. In other cases, the computer phobe is afraid and mistrustful of all information retrieved from one or more of the databases. Sometimes patrons do not discriminate between computer terminals and think that an optical disk workstation is the same as an online catalog terminal. One focus with instructional activities is to try and encourage patrons to seek skilled help.

One of the other major instructional difficulties is that when a patron selects to use an online database, that person is sometimes physically removed from the reference department and supervision. Online catalog terminals are distributed throughout the libraries on campus and are available many more hours than reference desk services. The "invisible patron" also includes dial-in users who are completely outside our service domain unless they call to ask for assistance.

Since instituting databases in the online catalog, the use of the collection has risen dramatically. Interlibrary loan has been especially affected. It is a continuing education process to let patrons know that many of the new databases loaded in the system are very broad in scope and that the library may not own everything listed.

The instructional goal is to articulate the availability and limits of both the databases which are available as well as the library collection.

FUTURE PLANS

The loading of the H. W. Wilson indexes and the opening of the Uncover gateway are merely starting points in this process. Every patron ultimately wants full text. Ideally this would be available at every terminal, and when the patron searches a database for information, the option would exist for retrieving the full article. An intermediate solution may include some form of user-generated online document ordering service where a patron can request a book, article, or report and have the item delivered to his home or office. The Colorado Alliance of Research Libraries is working in developing software for user-generated ILL plus full-text document retrieval.

In a more short term, ASU is examining a host of other bibliographic, full-text, and directory-type databases. The major limitation is funding for the tape leasing as well as for the disk space. It is hoped that the Wilson files and others which are currently operational will become indispensable for faculty and student use and that funding can be raised from external sources (i.e., campus departments or colleges) to continue to add more specialized databases.

Another area under consideration is the possibility of networking the ASU CARL system with other libraries in the state. This could be done by linking disparate systems via a leased line gateway or by additional libraries actually loading their MARC cataloging as a separate database in the ASU system. In the later case, these libraries could also use ASU's CARL for circulation and maintenance in the same way as the CARL network in Colorado.

The development of a true "electronic library" which offers a broad spectrum of databases is part of the vision of the future. There is no question that the local loading of databases as one means to this end will help fulfill this dream.[11,12]

REFERENCES

1. Judith Rice-Jones, "An Interview with Ward Shaw and Patricia Culkin: Designers of CARL PAC," *Colorado Libraries* 11, no.1:5–12 (Spring 1986).
2. "CARL (Colorado Alliance of Research Libraries)," *Library Technology Reports*, 418–21, (July/Aug. 1986).
3. Kenneth E. Dowlin, "Maggie III: The Prototypical Library System," *Library Hi Tech* 4, no.4:7–21 (Winter 1986).
4. Donald T. Hawkins, "Multiple Database Searching: Techniques and Pitfalls," *Online* 2, no.2:9–15 (Apr. 1978).
5. Ted Koppel and Ward Shaw, "What's That Doing Here? Non-Bibliographic Data in a Bibliographic Environment," *Reference Services Review* 15:15–19 (Winter 1987).
6. Gary M. Pitkin, "CARL's Latest Project: Access to Articles Through the Online Catalog," *American Libraries* 19, no.9:769–70 (Oct. 1988).
7. George S. Machovec, "Table of Contents Databases for Periodical Articles: Fad or Trend?" *Online Libraries and Microcomputers* 6, no.10:1–6 (Oct. 1988).
8. George S. Machovec, "Selection Criteria for Leasing Databases on CD-ROM and Magnetic Tape," *Online Libraries and Microcomputers* 6, no.3:1–4 (Mar. 1988).
9. Ibid., "Negotiating with Database Producers for Local Tape Loading," *Online Libraries and Microcomputers* 5, no.10:1–3 (Oct. 1987).
10. Robert A. Aken, "Meeting the Patron at the

OPAC Crossroads: The Reference Librarian as an Online Consultant," *RQ* 28, no.1:42–45 (Fall 1988).

11. William G. Potter, "Beyond the Online Catalog: Steps Toward Creating an Online Library," *Journal of Academic Librarianship* 14, no.6 (Jan. 1989). This is part of an insert called *Libraries and Computing Centers: Issues of Mutual Concern*, Issue 11:3–4.

12. Jon Drabenstott, "Beyond the Online Catalog: Great Potential and Profound Change," *Library Hi Tech* 6, no.1:101–11 (Spring 1988). A forum with six short articles by different library automation consultants. ■■

APPENDIX A. SAMPLE WELCOME SCREEN TO A DATABASE LOADED AT ASU

```
01/03/88
10:15 A.M.          SELECTED DATABASE:  Applied Science & Tech

The Applied Science & Technology Index covers 335 English-lang-
uage journals in the geneal fields of engineering, computers,
chemistry, applied mathematics, energy, and a wide variety of
applied sciences.  The database has records from late 1983 to
the present and is updated monthly.

Name searching retrieves proper names;  ord searching retrieves
words from article titles, journal titles, subject descriptions,
and other portions of the record.
         Enter    N  for  NAME search
                  W  for  WORD search

                  S   to   STOP or SWITCH to another database

         Type the letter for the kind of search you want,
         and end each line you type by pressing  <RETURN>

(Data copyright H.W. Wilson Inc.)
                    SELECTED DATABASE:  Appleid Science & Tech

ENTER COMMAND> >
```

APPENDIX B. SAMPLE NEWPEX MENU AT ASU

```
              Welcome to the ASU ONLINE SYSTEM EXECUTER
A project of the Colorado Alliance of REsearch Libraries and ASU
You may select:

              1. Public Catalog
              2. Circulation
              3. Maintenance
              4. Serials Control
              5. Fin
              6. Public Catalog (test system)
              7. Circulation (test system)
              8. Maintenance
              9. Mail Registration [coming soon]
             10. Mail System [coming soon]
             11. Maintenance (Career Services: Company)
             12. Maintenance (Career Services: Calendar)
             13. Maintenance (Career Services: Books)
             14. Maintenance (Music Song Index)
             15. Maintenance (Maps)
             16. Maintenance (Newspaper)
             17. Maintenance (AMC)

Enter the NUMBER of your choice, and press the <RETURN> key > >
```

APPENDIX C. ASU SAMPLE PAC MENU IN CARL

The following databases are currently available:

ASU LIBRARIES
1. ASU General Catalog
2. Library News and Help

REFERENCE WORKS
3. Encyclopedia
4. Song Index

CAREER SERVICES
5. Companiess
6. Calendar
7. Books

JOURNAL INDEXES
8. Humanities
9. Applied Sciences and Technology
10. Social Sciences
11. General Science
12. Education
13. Business Periodicals
14. Uncover - Article Access

SPECILIZED COLLECTIONS
15. Map
16. Solar Energy Index

Enter the number of your choice, and press the <RETURN> key:

Rethinking OPACS: The Design of Assertive Information Systems

Patricia B. Culkin

The notion of "Other Data for OPACs" raises certain prior questions. What exactly are OPACs in current rendition? How has their development diverged from conceptions of a decade ago? How can the divergence be managed to maximize current technologies? How can we ensure that OPACs play a pivotal and intellectually sound role in the educational process? These questions need to be addressed quite directly before specific and potentially very costly investments are made in electronic versions of "other data" for OPACs. In other words, maybe we need to address the "why" and the "how" as well as the "what."

OPACs came to be as a result of the library community's fascination with burgeoning computer technologies in the early 1960s. Libraries quickly recognized that the data ordering and high-volume processing capabilities of computers were completely appropriate to the management of cumbersome indexes and catalogs.

Economic and conversion issues prevented serious adoption of computer technologies, however, until the milestone—now fondly referred to as AACR2—forced the issue. As these revised cataloging rules were clearly dictating that manual catalogs be refiled or jerryrigged to comply, and as there were no guarantees that later revisions might not require similar effort in future decades, it became easy for many managers to decide that the time for computer-based catalogs had come.

Of course, there were other forces driving the development of early online catalogs. It was clear that ongoing catalog maintenance (introduction of new records, adding, changing, deleting entries, etc.) would be more efficient and less expensive in an electronic environment. It was also obvious that the computer's ability to consult multiple indexes and quickly assemble and sort results would improve searching for the end user. Computer-based catalogs could also be consulted from locations outside the library, thus offering faculty, students, and home users access to the library's holdings from remote sites. And finally, the computer environment would allow collections from multiple institutions to be viewed in an amalgamated environment. It was clear that OPACs, properly designed, could revolutionize access to library collections.

It is important to realize now, however, that the underlying thinking behind these early OPAC designs was to replicate the catalog. Computerization would make holdings easier to maintain and make searching easier to do, but, bottom line, the design intention was to recycle an existing tool into an electronic environment.

Herein lies the major divergence, or at least potential divergence, of today's OPAC from those of a decade ago. And if it is all in the name, then we may have to change the name. "Online," with its implication of computer-based, is still an accurate description of the technical base. "Public" and "Access" are also still relevant. They are descriptive of the audience to whom the systems are targeted and of the activity they are designed to enhance.

Patricia B. Culkin is Associate Director, Colorado Association of Research Libraries, Denver, Colorado.

"Catalog," however, is much too conservative. It has a finite connotation: its bounds are drawn by the holdings of the library. The word itself limits and perhaps even prevents creative thinking about how the resource can be developed to its fullest educational potential. Library information systems, if reconceived in modern telecommunications environments, can become portals to rich and diverse information sources that could never be replicated in home collections or in local computers. Yet the accessibility to this resource would have immense impact on user conception of how our intellectual heritage is organized. If these systems are thoughtfully and sensitively re-executed, it is possible that they could become an active, rather than passive, contributor to the educational process.

The educational process might welcome the contribution. We are all aware of declining college entrance scores and how basic reading, writing, math, and science skills are dwindling. The national literacy rate is said to have been falling steadily since the mid-1960s.

The electronic media have taken most of the fall on this one, from television to video programs to video games and even to PC technologies. People do not read any more; they prefer visual and/or aural communication of information, and they prefer it in brief, glitzy packages. They want information to be organized in ways that are familiar to them; they want access to be simple and straightforward; they want results presented in attractive, easily understood displays; and they want no physical inconvenience (the couch-potato syndrome).

In many ways, OPACs are positioned to redirect these characteristics of modern learning behavior back into literate pursuits. While observing the forms of information presentation that are comfortable to users, these systems can be used to organize and display the riches of our shared cultural heritage to a very diverse population. By requiring reading, analysis, and evaluation of the displays as the keys to understanding them, and perhaps by even offering commentary on the results, OPACs can contribute in a direct, context-specific way to the promotion of true literacy. They can satisfy, in a comfortable fashion, the immediate information needs of an end user and, at the same time, expose that user to a resource that can both inspire and extend his understanding. In other words, he will learn.

To illustrate the paradigm, we can borrow from a concept proposed by E. D. Hirsch, Jr., the author of *Cultural Literacy* (1987) and the *Dictionary of Cultural Literacy* (1988). Hirsch, an English professor at the University of Virginia, Charlottesville, has proposed that "true literacy depends on a knowledge of the specific information that is taken for granted in our public discourse."[1] He has based this contention on some very interesting theories about how learning occurs, most importantly that

> . . . learning depends on communication, and effective communication depends on shared background knowledge. The optimal way to fulfill this requirement of communication is simply to insure that readers and writers, students and teachers do in fact share a broad range of specific knowledge.[2]

He calls this broadly shared background knowledge "cultural literacy" and argues that the content of this literate background knowledge can be taught systematically to all students. Thus buoyed by this common base, students can go beyond it to the discovery of "further educational, economic, and social improvement." The major components of this base are presented narratively in the first book and organized in dictionary form in the second.

While his work is somewhat controversial in professional educational circles (Can the important elements of our culture be reduced to a single narrative presentation? Will familiarity with these elements make one truly literate?), his premise regarding the learning dynamic can still be applied to the design of the next generation of electronic information systems. If, as he contends, learning does require the sharing of a "broad range of specific knowledge," then what is better positioned to accomplish that sharing than library information systems? Libraries have been charged for centuries with preserving and presenting the elements of our cultural heritage. If they allow their electronic systems to embrace a diverse and physically remote universe of material and user interactions are designed in a sensitive, educationally sound fashion, they can become the provider of that common base. They are a natural for the role.

The above attempts to address the "Why" of an extended OPAC concept. The next section considers the "What." Some OPAC designers have already begun to think about how to make their current database management and search software work on data that are other than bibliographic. Recognizing that catalog management and access software is stable and effective in its limited MARC record purview, they have begun considering how to apply these techniques to other kinds of data and to make decisions about what that data should be.

The most obvious candidate is the journal literature. The research requirements of a variety of disciplines, especially in the sciences and social sciences, are based heavily in journal sources. The same can be said of less formal research required by undergraduate and secondary school curriculums and of that conducted by lay researchers in public library environments. Yet despite this requirement, and despite the substantial annual investment that journal acquisition requires (in academic libraries up to 70 percent of materials budgets), libraries have been unable to offer electronic access to these materials in fashions that are competitive with that applied to monographs. They have relied on combinations of paper indexes and fee-based, remote online access that have both physical and logical inconsistencies. It has long been suspected that this approach has stifled as much research as it has promoted.

To address this issue, certain institutions are experimenting with offering equal access to journal and report literature in the OPAC environment. Many of the indexing services, Wilson, ISI, EIC, Engineering Index, ERIC and MEDLINE among them, are offering their products in machine-readable (often MARC or MARC-like) structures. Institutions can site license these databases, load them into their local database management systems, apply their own indexing and access software, and allow users to search these files in parallel with the home collections. Vanderbilt University, using NOTIS system software, has loaded MEDLINE and is working on loading selected Wilson indexes. Arizona State University, using CARL system software, has licensed eight of the Wilson indexes. The Humanities Index and the Applied Science & Technology Index are now in release. The rest will be loaded during Spring 1989. California Institute of Technology offers certain of the ISI products using a BRS software interface.

The Colorado Alliance of Research Libraries (CARL), a consortium of Colorado-based research institutions, offers online access to the contents of current journals through its own database called UnCover. The database consists of tables of contents collected from journals owned by its members. UnCover now contains over 200,000 citations from 5,400 titles collected since fall 1988. It will include more than 600,000 citations and over 10,000 titles by September 1989. Data will remain online for at least five years.

The consortium manages centralized, online serials checkin for its members. As a check-in byproduct, tables of contents pages are transcribed into MARC record structures which are loaded and indexed to the Uncover file within seconds of handling. All issue check-in data, as well as article information, is made available immediately to users in PAC. Articles can either be retrieved by standard CARL searches (WORD, NAME, etc.) or can be reassembled as the table of contents for any specified issue. Any article retrieved from the Uncover database lists all consortium owners, so the database also functions as an online union listing tool. Data in the Uncover database can also be viewed from the owning library perspective.

There is, then, clearly some substantial effort being applied to OPAC access for journal literature. The reasons why there is not more development on this front have to do with cost, both for hardware and for site licensing; with technical issues; and with intellectual issues. Hardware costs are an issue because these files are big: they almost always require serious upgrade to the OPAC storage medium, a capital cost that is not easy for libraries to anticipate or underwrite. Site licensing costs are expensive, too. They are high in their own right for institutions of any size (license fees are usually based on number of terminals and/or size of user population) and because they are often accompanied by a requirement that the site continue its paper subscriptions.

Technical issues usually relate to the data itself. While the structures are often MARC or MARC-like, they tend to differ from the standard formats at the field and subfield level, so software work needs to be performed to make them load and index correctly into the local base. Intellectual issues also need to be addressed in projects of this type. Assuming limited capitalization, how can institutions make the best intellectual choices from the literally hundreds of online products? This is not an easy task in the relatively low profile paper medium; it is much more difficult in a highly visible, online environment.

OPACs are also offering access to other kinds of data in addition to journal and document indexing sources. These include standard reference materials such as Grolier's *Academic American Encyclopedia* (Carnegie Mellon University, Arizona State University, CARL, and others), the *American Heritage Dictionary* (Carnegie Mellon and CARL), and *Roget's Thesaurus* (soon at CARL). They also include many and myriad local databases: Career Services Data at Arizona State University; indexes to photography collections at Boulder Public Library, Boulder, Colorado; demographic data about the Denver Area supplied by a consortium of chambers of commerce to CARL; and a day care services directory and a voter information file at Pikes Peak Public Library, Colorado Springs.

It is even the case that certain commercially available software products come with embedded search software that can interface with the local OPAC command structure. Houghton Mifflin's Business Software Division provides "Roget's Thesaurus" with embedded "circular software," which automatically takes you from the primary word to related words to their related words, etc. It even preserves tense integrity for verbs. (If the search word is "went," the related words are displayed in past tense). The product was designed to be incorporated into word processing applications but its C software interface can be integrated into OPACs also. A similar release is planned for the *American Heritage Dictionary* in summer 1989.

Developments like these, reference tools and locally developed products, should probably be more prevalent than they are, because the associated hardware and licensing costs are small or nil; the software issues are under local design control; and the intellectual issues, while valid, are easier to manage because they do not have high price tags.

There are some other kinds of data the library community should be pursuing for local integration if it is to design truly assertive information systems. These include book reviews that are linked directly to collection holdings, citation counts that are linked to journal articles, circulation counts for individual titles, etc. Book review sources could be standard products such as *Book Review Digest*, retrospective *New York Times Book Reviews*, *Choice Cards*, even *Books for College Libraries*. These can be linked to the appropriate titles by ISBN number or by a locally managed editorial effort. The option to see a review could be inserted directly into the search session whenever the user selects a record for full display.

OPACs could also routinely report circulation histories for materials returned in a search. A title's circulation track record is arguably an important element for helping a user determine relevance. Of course, the ideal statistic would be a title's citation history, but it may be a while before we can systematically assemble and interject that piece. If there is published evaluation data about a class of materials, make the text available online, e.g., a review and comparison of encyclopedias or dictionaries. To the degree that OPACs can report statistical data or display critical data, they should do so.

These are just examples of ways in which OPAC technology can be used to extend the professional effort. Emily Fayen made the point in an article in *American Libraries*[4] a few years ago that librarians need to assert their professionalism a little more actively.

Other professionals, lawyers, doctors, accountants, etc., take direct and active responsibility for solving clients' problems. They apply their expertise to a set of issues and make strong recommendations for solution. Teachers, the better ones anyway, have a sense of the material they want their students to master. They may circle a bit in order to let students gain understanding through their own efforts, but they take responsibility for the process.

Librarians need also to assume responsibility for solution and need to ensure that this expertise is incorporated into OPAC design. This is the "How" of OPAC regeneration. INNOPAC software takes a step in this direction. After a user selects a record in INNOPAC, the system offers to display records with the same subject heading or to position the user for browsing in the classification scheme. While this is still "pointing" (it makes no recommendations), it has the effect of dynamically illustrating how information is organized in a bibliographic file. CARL software allows searches to be re-executed automatically in over twenty-two databases, from bibliographic to reference to demographic, etc. Not only does this return an abundance of information, it has the effect of demonstrating how different tools organize and present similar information. Enough searching in diverse environments reveals the underlying structures to the user.

The following may illustrate how diversity of resources can affect learning: a user is looking in the home library catalog for books on random number theory. His search returns a twenty-two record set. He notices that a lot of the books contain the word "stochastic." He does not know what that means, so he takes that "stochastic" to the dictionary (informative, but not comprehensive), to the encyclopedia (eight articles, one long one on "stochastic processes"), to the thesaurus (synonyms displayed in context), and to the journal index (lots of new articles in math journals he did not know existed). Then he goes back to the list of books, picks one, looks at its review, etc.

Does he have some information about random number theory as a result of this session? Yes. Does he understand how the "stochastic process" relates to random numbers? Probably. Is he intimidated, pressured, bored by the process? No—he was driving it; he could have quit at any point. Has he "learned" anything? What do you think?

OPAC creators also need to design intriguing user interfaces which entice and challenge users. OPACs need to be conceived of as communicators that have give and take with the user and that can guide and point users to expanded or alternative choices. They need to have a style of communication that is engaging. Just as newspapers have communication styles—the voice, tones, and rhythms of *The New York Times*, *The Wall Street Journal*, the *Christian Science Monitor*, the *Los Angeles Times*, *USA Today*, etc., are all unique unto themselves and quite distinct from one another. They are also quite different from chain newspapers and from small town, independently published papers. The same is true of broadcast media: a few minutes of listening is sufficient to distinguish between National Public Radio and AM talk radio station or a high-profile FM rock station.

OPACs, if they are to become an active participant in intellectual exchange, need to develop a voice and tone that both reflect institutional persona and accomplish quick rapport with users. The push to develop a universal command language ultimately disserves the user; it assumes he is too dense to understand anything but the most homogenous instructions. If anything, the system should err on the side of challenge. A stretch to comprehend the messages might extend to extra consideration of the data. A key element of the success of assertive OPACs will be their ability to both simulate and stimulate discourse.

While these projects and experiments with other data for OPACs do need to continue and multiply in local institutional settings, while we do need to pay serious attention to conveying structure and incorporating evaluative techniques in OPAC environments, and while we do need to consider the personality of OPACs as a design element, these factors comprise only part of the ultimate design solution. That solution is much more likely to occur through interconnection of disparate systems. Growing much faster than libraries' capital budgets for hardware acquisition or the pool of data available for local licensing is a dynamically developing national telecommunications network that is potentially a route to and from all OPACs and most online database services, regardless of origin.

This network, called Internet and administered by EDUCOM, is in itself a collection of regional telecommunications networks that are used quite heavily by the academic community for sharing and disseminating research activity. The regional components, Westnet, Tnet, and Nysernet, to name only a few, are the result of regional academic institutions and not-for-profit service organizations who decided to create telecommunications links among themselves. The EDUCOM Technical Committee and the Library of Congress Network Advisory Committee are currently engaged in activities designed to expand and upgrade this network and to promote and extend its resource to libraries.

Many in the library community already have individual access to Internet through their campus networks. Certain OPAC installations also have access to it, but the interconnections are usually cumbersome and the external resource available for consultation is thin. However, if OPAC installations and online database vendors and their subcontractors became providers on this network, and if OPACs offered transparent access to users as a featured main menu item, we would have cheap and relatively instant interconnect among library collections, as well as access to myriad online commercial resources. Assuming commercial databases will negotiate flat-fee subscriptions, we can open a wide variety of resource to users with relatively low initial capitalization.

This model would allow institutions a certain amount of flexibility in terms of local data acquisition without precluding their users' access to the external materials. It would also allow institutions creative leeway, in terms of editorial policy, presentation, and organization, for the things they do locally. In the true spirit of survival-of-the-fittest, certain institutions will probably become champions of data presentation, and others will fade away. Arizona State's handling of Wilson indexes may be preferred over XYZ U or Wilsonline itself. Another site may do an outstanding job with technical report literature. Still another might do a compelling job with a genealogical research system. A nationally networked system of OPACs removes the burden from individual institutions to attempt comprehensiveness and leaves them free to do a few projects really well. And, for the first time, we would have a fair method for allowing the market, i.e., the user community, be the arbiter of what really works in online searching.

By reconceiving OPACs in modern telecommunications environments, we accomplish several objectives:

1. Expand the scope and volume of research material available to any user.
2. Create a loose but effective organization of these materials.
3. Simplify the medium.
4. Convey the structure.
5. Become an active participant in the educational process.

The objectives described above are not revolutionary. They have been the fantasy target of most library computer system design for decades. They were touted in the literature as the ultimate solution before OPACs were invented. The astounding characteristic of today's design environment is that the pieces are extant and, in many cases, in release. It now remains to accomplish the links, identify the providers, and "jazz up" the messages.

REFERENCES

1. E. D. Hirsch, Jr., *Dictionary of Cultural Literacy* (Houghton Mifflin, 1988), p.ix.
2. Ibid., Preface, p.ix.
3. Ibid.
4. Emily Gallup Fayen, "Beyond Technology: Rethinking 'Librarian,'" *American Libraries* 17 no.4:240–42 (Apr. 1986). ■■

Public-Access Computer Systems: The Next Generation of Library Automation Systems

Charles W. Bailey, Jr.

Historically, library automation has focused attention on the automation of internal library processes, such as cataloging; however, a new generation of library automation systems intended for direct use by library patrons is emerging. These public-access computer systems are categorized and described. The types of public-access computer systems are: (1) catalog information systems, (2) electronic information systems, (3) information presentation systems, (4) instructional systems, (5) consultation systems, (6) information service and delivery systems, (7) conferencing systems, (8) integrative systems, and (9) end-user computing facilities. To be effective, these computer-based systems must be integrated with the traditional collections and services of libraries.

Increasingly, libraries are making computer systems available for public use. Online catalogs, end-user searching services for remote databases, CD-ROM reference databases, and local reference databases on mainframes and minicomputers are examples of public-access computer systems. A public-access computer system is defined, simply, as any library computer system that a patron can use directly.

Prior to the advent of public-access computer systems, library automation had focused primarily on improving the effectiveness and efficiency of internal library functions: acquisitions, cataloging, circulation, interlibrary loan, office automation, and serials control. The benefits of automating these functions were significant, but not always visible, to library users. Now, libraries are providing users with a powerful array of computerized systems, and the long-standing effort to automate library functions is, at last, becoming unmistakably apparent to users.

This paper will survey current and projected public-access computer systems.

OVERVIEW OF PUBLIC-ACCESS COMPUTER SYSTEMS

Public-access computer systems are still emerging, but it is possible to draw a preliminary map of the territory that they will cover.

Public-access computer systems are based on computer and telecommunications technologies, which are evolving at a very rapid rate.

Several major trends are shaping the current computing environment.[1] Historically, computer and telecommunications technologies have been characterized by dramatic increases in functionality and price/performance over time and by equally significant decreases in the size of system components. Different types of information (e.g., text, audio, and graphics) are increasingly becoming digitized. High-

Charles W. Bailey, Jr. is Assistant Director for Systems, University Libraries, University of Houston, Texas.

density storage devices, such as optical disks, are emerging that can store massive amounts of information at low cost. Evolving network technologies permit computers to communicate with each other, although data exchange across heterogeneous networks is still problematic. Artificial intelligence software is maturing, and a growing number of AI products and systems are becoming available.[2]

Public-access computer systems utilize specific computers, operating system and applications software, storage technologies, input/output devices, and data communications systems to provide services to users. The unique mix of technologies employed by a public-access computer system determines its capabilities. It is beyond the scope of this paper to consider fully the diverse technological infrastructures that could be used to implement each different type of public-access computer system. Nonetheless, there are certain technical capabilities of public-access computer systems that are worth highlighting briefly. Depending on underlying software and hardware configurations, these systems can provide users with: (1) rapid dissemination of new information; (2) powerful searching tools for quickly retrieving needed information; (3) sophisticated information manipulation and analysis tools (e.g., statistical and textual analysis software) to assist the user in creating intellectual works from retrieved information; (4) ability to download information to the user's computer for further manipulation and analysis; (5) simultaneous access to system resources by multiple users; (6) remote access to needed information and services from offices, homes, and other locations; and (7) round-the-clock availability of system resources.

Given these capabilities in the context of the overall information technology environment, the library can be envisioned as evolving into an online information utility that provides users with access to local public-access computer systems and acts as a gateway to remote systems. To be effective, this new function will need to be integrated with traditional library materials and services. A number of thoughtful authors have probed different aspects of such a transformation. Works by Battin,[3] Dowlin,[4] Downes,[5] Drake,[6] Horny,[7] Lancaster,[8] and Murr and Williams[9] investigated key issues related to this subject. Parsons foresightedly speculated about possible "online public access systems."[10]

Overall, it is important to recognize that impressive technological advances do not necessarily imply that successful, operational systems will emerge in libraries that make use of these technologies. Public-access computer systems efforts by libraries and vendors will be bound by fiscal, organizational, and legal constraints that will help shape the course of their development.

Equally important will be user demand for these systems. Convenient access and adequate user support will be necessary to nurture the use of public-access computer systems. Users are likely to continue to be cost-sensitive, and, if they are being charged, they will ration their use of public-access computer systems. Issues related to fees for service, equitable access, and library subsidies will become increasingly important as public-access computer systems become an increasingly visible and critical part of library services. When considering scenarios about technology-driven changes in library collections and services, it is important to remember that the library plays an important social role as a provider of no-cost or low-cost information to those who cannot afford to obtain this information directly from publishers and other information vendors.

Public-access computer systems offer exciting prospects for libraries and their users. Librarians should neither be discouraged by the obstacles that must be overcome in order to make them a reality nor expect that their libraries will be transformed overnight into the elusive "electronic library." As was the case in past automation efforts, progress will be made in a steady, evolutionary fashion. A variety of pioneering projects, most of which will be done with relatively modest resources, are likely to explore the benefits and problems associated with new types of public-access computer systems before they are adopted on a wide-spread basis by libraries.

As libraries continue to develop public-access computer systems and provide them to their users, they are likely to concentrate their efforts in the following broad areas.

CATALOG INFORMATION SYSTEMS

These systems give library users direct access to bibliographic, authority, summary and detailed holdings, and item-status information (or to subsets of these types of information) about materials in library collections. These systems typically offer users powerful retrieval capabilities, such as keyword and Boolean searching. A single-institution online catalog represents a typical contemporary catalog information system.

The long-term goal in the catalog information systems area will be to make information about remote library collections as accessible to the user as information about the local library collection.[11] This trend can be seen in the efforts of library consortia either to establish online union catalogs that reflect their joint holdings or to link the separate institutional online catalogs of their member libraries. On a broader level, libraries will increasingly provide public access to the regional, national, and international databases of bibliographic utilities, such as WLN, RLIN, and OCLC. As libraries provide users with instant access to information about remote library collections, issues related to libraries' ability to support increased resource sharing and to ensure rapid delivery of requested materials will become more critical.

An example of a local catalog information system is the UCLA Library's ORION system, which is written in PL/I and runs on a mainframe computer.[12] ORION, which is under authority control, offers users a diversity of searching techniques with which to retrieve bibliographic, authority, and holdings data, including exact phrase, keyword, Boolean, call number, and control number searching (e.g., ISBN) as well as search limitation by different criteria (e.g., location, date, and language).

Hildreth describes a variety of enhancements to online catalogs that could be incorporated in the next generation of these systems, including user interfaces that provide more feedback and prompting, system reformulation of search keys to correct user-input errors, search techniques derived from information retrieval research (e.g., relevance feedback, stemming, and term weighting), and expanded bibliographic records that contain information from the tables of contents and indexes of books.[13] These types of system enhancements are applicable to all types of catalog information systems.

ELECTRONIC INFORMATION SYSTEMS

These systems provide users with access to a growing variety of reference and source materials in digital form. Sophisticated searching techniques are commonly available in electronic information systems, and analysis tools may be incorporated as well.

Representative types of electronic information include: (1) indexes and abstracts; (2) full-text reference works, such as directories and encyclopedias; (3) full-text journals; (4) full-text newswires; (5) full-text books that are not reference works; (6) numeric databases, such as census data; (7) digitized audiovisual materials, such as graphic images; and (8) hypermedia, which organize other types of electronic information materials into a network of frames.

Electronic information materials may be derived from printed or audiovisual materials, or they may exist solely in digital form.[14] As electronic information continues to evolve, new information formats, which draw upon the unique capabilities of computer technology and have no equivalents in prior technologies, will emerge.

In the foreseeable future, electronic information systems are likely to complement, rather than dramatically displace, traditional materials in libraries, such as books. The development of electronic information systems by publishers is likely to be driven primarily by marketplace forces rather than strictly by technological capabilities. The provision of electronic information systems by libraries will be strongly influenced by economic factors and intellectual property rights issues, reflecting high costs relative to those of traditional materials and attempts by copyright holders to restrict use of electronic information materials.

The question of ownership is likely to be especially problematic, since neither licensing nor accessing a database gives the library any permanent claim on it. Maintaining parallel print and electronic collections is an expensive solution to part of this problem; however, it does not address the issue of ensuring permanent access to information that is available solely in electronic form. These databases may be discontinued by vendors for economic or other reasons. Without a national program that, at minimum, targets preservation of databases that exist solely in electronic form, we are in danger of losing an increasingly important part of our intellectual heritage. A potential model for such a preservation effort is the Knowledge Warehouse, a pilot project in England that is exploring legal, commercial, and technological issues related to the preservation of electronic information, including electronic working copies created as part of the print publication process.[15,16]

An example of an electronic information system is Carnegie Mellon University's Library Information System, which utilizes the STAIRS retrieval software running on a mainframe computer.[17] The mainframe computer is linked to the campus local area network. The Library Information System includes databases such as *Academic American Encyclopedia*, Computer Database, Houghton Mifflin's *American Heritage Dictionary*, Magazine Index, Management Contents, National Newspaper Index, and Trade and Industry Index. The system enables users to search these databases using keyword, Boolean, proximity, and search limitation (e.g., date limitation) techniques.

INFORMATION PRESENTATION SYSTEMS

These systems provide users with a structured sequence of screens that describe library-related topics, which users can review at their own pace. Typically, an information presentation system will be menu-driven, with the user moving up and down a hierarchy of screens; however, with the advent of hypermedia, systems are being created that use a network of interconnected information frames.

A representative information presentation system is the Information Machine at the University of Houston Libraries written in QuickBASIC and running on a microcomputer.[18] The Information Machine provides users with menu-driven access to explanations of library research strategies, materials, systems, facilities (complete with library maps), services, open hours, policies, and other orientation information.

INSTRUCTIONAL SYSTEMS

These systems interactively teach users about library-related topics, analyzing user responses and varying the information presented in accordance with them. Current instructional systems, commonly referred to as computer-assisted instruction systems, use drill-and-practice, tutorial, simulation, and game techniques to teach needed material.[19]

Prototype intelligent computer-assisted instruction systems, which are based on expert system technology, have been developed, and researchers are examining how these systems can be grounded in learning and instructional design theory.[20] These systems may provide much greater tailoring of instruction to individual learner differences than is currently possible in conventional CAI programs.

An illustrative instructional system is a CAI program at the University of Delaware Library, which employs the PLATO software as its authoring language and runs on a mainframe computer.[21] A microcomputer version was also developed which includes four lessons that treat the card catalog and LCSH, periodical indexes, newspaper indexes, and government documents indexes.

CONSULTATION SYSTEMS

These systems advise users, much as a professional librarian would, about library-related topics. Consultation systems are typically based on expert system and related artificial intelligence technologies. They are knowledge-based systems that embody the special expertise of library staff in using the library's collections, systems, and services.

Expert systems have three primary components: a knowledge base that contains facts, rules, and other representations of human knowledge related to a particular topic; an inference engine, which solves problems by manipulating information in the knowledge base in a way that mimics human reasoning; and a user interface, which permits system interaction with the user and explains system findings.[22]

A representative consultation system is the University of London's PLEXUS system, a prototype expert system that is written in Turbo Pascal and runs on a microcomputer.[23] This system, which is intended for public library use, is designed to identify relevant sources of information (e.g., individuals, books, and institutions) that can assist a user with a specific gardening problem. After building a profile of the user, PLEXUS allows the user to enter free-text questions about gardening, and it employs knowledge about the user, gardening, and gardening resources to identify appropriate resources to answer those questions.

INFORMATION SERVICE AND DELIVERY SYSTEMS

These systems allow users to request library services (e.g., mediated online search) and document delivery (e.g., local delivery or interlibrary loan of an item). This latter objective may be achieved through delivery of a physical item or the transmission of digitized information derived from a physical item.

As users employ more powerful tools for quickly identifying needed local and remote library materials in printed form, they will want equally convenient and speedy access to the information itself. Unless adequate attention is paid to information delivery, enhanced access to bibliographic and electronic information systems may create a "library with glass walls" instead of a "library without walls"—needed information can be quickly identified but not readily obtained.[24]

An example of this type of system is the Electronic Access to Reference Service system at the Health Sciences Library of the University of Maryland at Baltimore. The system is written in MIIS and runs on a minicomputer.[25] This menu-driven system allows users to search the online catalog, request a mediated online search, ask a reference question, submit an interlibrary loan request, request that an article be photocopied and delivered, submit book purchase requests, read electronic mail from library staff about service requests, and read library news notices.

CONFERENCING SYSTEMS

These systems give users a way of exchanging points of view on different topics in a public or private forum.[26] Unlike electronic mail systems, which provide one-to-one or one-to-many message services, computer conferences allow users to read ongoing dialogues by many participants on specific issues and to contribute to those dialogues. Computer conferences can be available to all or restricted to particular participants. A computer conferencing system can support a number of simultaneous conferences on different topics.

For academic and other research libraries, computer conferencing provides a new way of disseminating scholarly knowledge, potentially on a national or international basis, if appropriate network and inter-network links exist. Given their mutable nature, computer conferences offer interesting challenges in the areas of access and preservation.

An example of a library-sponsored computer conference is a private conference for librarians affiliated with the university library system, independent campus libraries, and the library school at the University of Michigan.[27] This system, which is one of many conferences supported by the University's CONFER conferencing software, enables librarians to electronically discuss a variety of issues of mutual concern and to make announcements of general interest.

INTEGRATIVE SYSTEMS

These systems provide users with convenient access to diverse local and remote public-access computer systems, new services based on these systems, and simplified techniques for utilizing these systems. Other types of public-access computer systems will be unified into a cohesive information network environment by integrative systems.

As the number of public-access computer

systems offered by a library increases, there will be a need to reduce the complexity of administration and use of these systems. Integrative systems will be developed to meet a variety of objectives, including: (1) permitting the use of a standard library workstation with a multiplicity of local and remote systems; (2) establishing a coherent networking strategy to provide access to these systems; (3) providing users with an interface that organizes the use of these systems; (4) assisting users in identifying appropriate systems to meet their needs; (5) teaching users how to utilize these systems; (6) linking heterogeneous systems to provide enhanced services to users (e.g., connecting an indexing database with an online catalog to identify locally held journals); (7) providing tools for further analysis and manipulation of information from different systems; and (8) furnishing a uniform user interface that masks underlying system differences in command syntax, retrieval capability, controlled vocabulary, and record structure as much as possible.

An example of an integrative system is the DoD Gateway Information System, which is being developed by the Defense Technical Information Center.[28] This prototype system offers the user: (1) a directory of available online databases, which can be searched by subject; (2) a variety of communication services, including electronic mail, online conversations with other users, and automated log-on to online systems; (3) a knowledge-based, common command language system for accessing heterogeneous online databases; and (4) post-search processing tools that can perform tasks such as converting records to a common format, eliminating duplicate records, sorting records by different keys, and analyzing search results.

END-USER COMPUTING FACILITIES

Microcomputer workstations housed in the library can provide users with access to a diversity of software tools, including business software (e.g., database management, spreadsheet, and other programs), programming languages and utilities, and scholarly productivity software (e.g., specialized foreign or scientific word processing, citation management, desktop publishing, and other programs). These workstations can be linked together in a local area network in order to share hardware and software resources. Given appropriate data communication linkages, these workstations can also access software packages, databases, and services on institutional or remote computers, including public-access computer systems provided by the library.

The linkage of increasingly powerful microcomputer workstations to public-access computer systems may bring into existence the long-awaited "scholar's workstation," an integrated computing environment for producing intellectual works. This may be achieved through a careful blending of resources: access to catalog and electronic information; a coordinated set of software tools to create, analyze, and manipulate information; and hardware to accomplish information transfer, production, and output.

Hess's detailed survey of seven microcomputer facilities located in academic libraries provides a good overview of contemporary efforts in higher education.[29]

CONCLUSION

Public-access computer systems represent the next step in the evolution of libraries. They provide an opportunity to synergetically blend printed and computerized information resources to create new, more effective library services. The library as we know it is not "dead," but it will be transformed.

Traditional library collections and services as well as existing library automation systems will provide a firm foundation for building public-access computer systems. In the near future, the library is unlikely to abandon print and metamorphose completely into a sophisticated system for providing electronic information and computer-based services. However, public-access computer systems will play an increasingly important role in libraries, and they will change the nature of the library in fundamental ways. Although the library as a physical entity will not disappear in the foreseeable future, the need for the user to walk through its doors will diminish over time, and increasingly its resources and services will be available in remote locations around the clock. Except in a small number of well-funded, innovative libraries, this change is likely to be incremental rather than swift and dramatic.

If the last generation of library automation systems is any guide, public-access computer systems could have a lengthy development cycle, gradually evolving from single-function to integrated systems. With careful planning, we can lay the foundation for eventual integration as we establish public-access computer systems.

An important step is to implement a high-capacity, library-wide local area network to permit access to a multiplicity of public-access computer systems from individual microcomputer workstations. With this essential infrastructure in place, local public-access computer systems can be added as network servers in a modular fashion and appropriate linkages to external public-access computer systems can be established as required. Gateway or bridge connections to institutional local area networks, dial-access ports, and other appropriate linkages can provide needed access to public-access computer systems for remote users.

Equally important is being cognizant, when developing or purchasing public-access computer systems, of the need eventually to integrate them. Standardization efforts in the computer, library, and publishing communities will play a key role in ensuring that systems can be effectively integrated, and libraries should be sensitive to whether their proposed public-access computer systems comply with applicable existing and emerging standards, such as the Common Command Language for Online Interactive Information Retrieval,[30] the Information Retrieval Service Definition and Protocol Specification for Library Applications,[31] and the Open Systems Interconnection[32] standards.

The challenge ahead is to balance our ongoing, increasingly expensive commitments to traditional information resources and services with the significant investment of human and fiscal resources required to make public-access computer systems a reality. Library automation systems aimed at computerizing internal library functions were costly; however, they did not alter the nature of the library's collections. Electronic information systems change library collections, providing better access to information at higher costs. These electronic information costs will constitute a significant and growing proportion of the total cost of providing public-access computer systems. Unlike equipment or software costs, which are primarily front-end costs, electronic information license or access fees will not drop significantly after the first year of use; rather, they are permanent costs subject to fee increases.

Libraries and library automation vendors have an overall record of success in creating complex, sophisticated library automation systems. Public-access computer systems offer exciting new challenges, which libraries and vendors will overcome to create a new generation of library automation systems that will provide library users with improved access to information resources and library services.

REFERENCES

1. Douglas E. Van Houweling, "Information Technology: A Survey From the Perspective of Higher Education," *Library Hi Tech* 4:23–28 (Winter 1986).
2. Linda C. Smith, "Artificial Intelligence and Information Retrieval," *Annual Review of Information Science and Technology* 22:43 (1987).
3. Patricia Battin, "The Library: Center of the Restructured University," *College & Research Libraries* 45:170–6 (May 1984).
4. Kenneth E. Dowlin, *The Electronic Library: The Promise and the Process* (New York: Neal-Schuman Publishers, 1984).
5. Robin N. Downes, "Journal Publishing, Resource Sharing, and New Information Technology," (Paper presented at The Impact of Rising Costs of Serials and Monographs on Library Services and Programs conference, Oklahoma City, Feb. 26, 1988).
6. Miriam A. Drake, "From Print to Nonprint Materials: Library Information Delivery Systems," *EDUCOM Bulletin* 23:28–30 (Spring 1988).
7. Karen L. Horny, "New Turns for a New Century: Library Services in the Information Era," *Library Resources & Technical Services* 31:6–11 (Jan./Mar. 1987).

8. F. W. Lancaster, *Libraries and Librarians in an Age of Electronics* (Arlington, Va.: Information Resources Press, 1982).
9. Lawrence E. Murr and James B. Williams, "The Roles of the Future Library," *Library Hi Tech* 5:7–23 (Fall 1987).
10. Jerry L. Parsons, "On-Line Public Access Systems: An Overview," *Technical Services Quarterly* 1:41–49 (Summer 1984).
11. Ibid., p.44–45.
12. James Fayollat and Elizabeth Coles, "Database Management Principles of the UCLA Library's Orion System," *Information Technology and Libraries* 6:102–15 (June 1987).
13. Charles R. Hildreth, "Beyond Boolean: Designing the Next Generation of Online Catalogs," *Library Trends* 35:654–62 (Spring 1987).
14. F. W. Lancaster, *Libraries and Librarians in an Age of Electronics*, p.54–62.
15. R. Williamson, "The Knowledge Warehouse: Legal and Commercial Issues," *The Electronic Library* 6:10–16 (Feb. 1988).
16. Michael Buckingham, "The Knowledge Warehouse: Technical Issues," *The Electronic Library* 6:6–9 (Feb. 1988).
17. Nancy Evans and Thomas Michalak, "Delivering Reference Information Through a Campus Network: Carnegie Mellon's Library Information System," *Reference Services Review* 15:7–10 (Winter 1987).
18. Jeff Fadell and Judy E. Myers, "The Information Machine: A Microcomputer-based Reference Service," *Reference Librarian* no.23:75–112 (1989).
19. Robert V. Bullough, Sr. and LaMond F. Beatty, *Classroom Applications of Microcomputers* (Columbus: Merrill Publishing Co., 1987), p.109–25.
20. Halyna Hajovy and Dean L. Christensen, "Intelligent Computer-assisted Instruction: The Next Generation," *Educational Technology*, 27:9–14 (May 1987).
21. *Bibliographic Instruction in ARL Libraries* (SPEC Kit 121) (Washington, D.C.: Association of Research Libraries, Office of Management Studies, 1986), p.64–66.
22. Kamran Parsaye and Mark Chignell, *Expert Systems for Experts* (New York: John Wiley & Sons, 1988), p.32–34.
23. A. Vickery and H. M. Brooks, "PLEXUS—The Expert System for Referral," *Information Processing & Management* 23, no.2:99–117 (1987).
24. Jane D. Segal, personal communication, July 1988.
25. Frieda O. Weise and Marilyn Borgendale, "EARS: Electronic Access to Reference Service," *Bulletin of the Medical Library Association* 74:300–2 (Oct. 1986).
26. F. W. Lancaster, *Libraries and Librarians in an Age of Electronics*, p.25–27.
27. William C. Manspeaker, Rebecca M. Dunkle, and Selden S. Deemer, "Microcomputer Installation and Support at the University of Michigan Library," *Information Technology and Libraries* 5:301–2 (Dec. 1986).
28. Allan D. Kuhn and Gladys A. Cotter, "The DoD Gateway Information System (DGIS): The Department of Defense Microcomputer User's Gateway to the World," *Microcomputers for Information Management* 5:77–87 (June 1988).
29. Anne Hess, "Notable Collections of Microcomputer Software and Hardware in Academia; The Findings of a Study: 1986–1987," *Library Software Review* 7:74–107 (Mar.–Apr. 1988).
30. Katharina Klemperer, "Common Command Language for Online Interactive Information Retrieval," *Library Hi Tech* 5:7–12 (Winter 1987).
31. Ray Denenberg, "NISO Draft Standard Z39.50: Information Retrieval Protocol," *Library Hi Tech News* no.39:1, 7–8 (June 1987).
32. Ray Denenberg, "Open Systems Interconnection," *Library Hi Tech* 3, no.1:15–26 (1985).

Special Sections

Optical Character Readers/Text Scanners: A Market Analysis

Mitchell A. Cahan

Optical character readers and text scanners have seen service in high-volume conversion of paper-based text into electronic format. Until recently, optical scanners saw heavy duty in banking, industrial production, and even in the retail industry. With the advent of desktop publishing throughout academia, optical scanners have now found a home in research institutions. Given the importance of publishing in the academic world and the high cost of textbook revision and other tasks requiring manual data entry, universities and research centers should take note of scanner applications. Libraries, as the information hub of research institutions, can play an important role in promoting the scanner as yet another indispensable technology for improving information management.

The following report examines today's text scanner market from the perspective of the academic health sciences community. The reader will find a description of applicants for the product, an examination of the technology (including hardware and software), a discussion of the different tiers of the market, advantages and limitations, future trends, and recommendations for further evaluation of specific products.

BACKGROUND

Modern scanning technology evolved from an early nineteenth-century effort to produce aids for the blind. Basic research in the field during the 1900s occurred in tandem with early efforts in both television and Morse code. By the middle of the twentieth century, researchers had focused work on reading machines and even produced a scanner that read one character per minute. By the late 1960s, scanners played an important part in the banking profession, automatically reading Magnetic Ink Character Recognition (MICR) font from bank checks. The last ten years have seen dramatic leaps in technology and a reduction in prices.

Until very recently the legal profession represented one of few areas of the corporate world that could justify the expensive undertaking of optical scanning; massive legal briefs posed high data entry costs and supported the move to conversion of text to magnetic form. With the recent explosive growth of the scanner market, other professions now recognize the potential of scanning devices.

Academic physicians, both researchers and clinicians, have long relied on manual labor to manage a vast array of laboratory, administrative, and patient-care data. From the small-scale office practice to the scientific laboratory, healthcare workers have coped with a massive amount of information routinely stored in paper format. As health professionals turn increasingly to computerized information for bibliographic searching, clinical data, and organization of laboratory files, academicians may find the optical scanner an answer to the laborious process of textbook revision; administrators may find it useful for records management; and the office-based practitioner for retrospective conversion of clinical files.

Specific applications for text scanners in the academic health sciences environment depend on the features and enhancements of a particular scanner. In general scanners allow conversion of information in a non-

Mitchell A. Cahan, Personal Information Management Specialist, at The William H. Welch Medical Library, The Johns Hopkins University, 1900 East Monument Street, Baltimore, MD 21205.

electronic format to a machine-readable format. Specific examples include digitization of medical textbooks, conversion of laboratory data from notebook to electronic files, and scanning of office files into computerized accounting systems. The first example provides one of the most exciting applications: in combination with telefacsimile machinery it would allow for decentralized editing and revision of the medical literature and creation of new online databases from existing knowledge bases.

TECHNOLOGY

Diversification and competition in the scanner market stems from the complex interaction between hardware and software required for successful marketing of a product. True distinction in the scanner market stems from the type of software that operates the machine.

Optical Character Recognition (OCR)

An OCR uses a light source to illuminate characters on a printed page. The black color of printed characters absorbs light, and the white portion of the page reflects light. The scanner's photosensor produces alternating voltage patterns from the black print and the white background of the page; variations in voltage translate into an analog signal. The scanner then processes the signal into a binary format, matches the digitized data against character patterns stored in memory, and finally creates ASCII characters.

The scanner may use one of two basic processes for conversion of text; sometimes a machine may use a combination of both methods. Character recognition technology plays the most important role in determining the scope of the machine in terms of the number of fonts and type faces that the scanner will recognize. Until recently most machines relied on template matching for character recognition: the machine reads a character and enters it successfully into ASCII format provided that it finds an exact match in its cache of stored fonts. Template matching machines only work if the font of the scanned document meshes with one of the stored fonts; such machines usually read only monospaced text (e.g., a typewritten document). Scanners that use pattern recognition technology surpass the limitations of template matching products. Pattern recognition, also known as feature extraction, involves the analysis of the constituent parts of each scanned character; the technique allows for reading of a variety of fonts, typeset, or proportionally spaced text and, in a sense, teaches the machine each new type set that it meets. Pattern recognition technology even allows for reading complexities of type, including ligatures, kerning letters, accents, and diacritical marks.

Clearly character recognition technique dictates that some machines on the market possess more intelligence than others. Several manufacturers offer proprietary versions of feature extraction technology. Examples include Kurzweil's Intelligent Character Recognition (ICR), and Calera Recognition Systems offers Intelligent Page Recognition (IPR). The sophisticated technology will mean increased price but offers increased flexibility. This is especially important for scanning older documents printed with obsolete type faces.

Image Scanning

Aside from scanning characters some scanners read images. Prospective scanner purchasers should contemplate the role of the machine: will it scan text or images or perhaps both? Design and engineering firms use specialized image scanners that range in price from $50,000 to several hundred thousand dollars for high resolution work. Machines for less do provide acceptable quality for desktop publishing but only when used in concert with expensive commercial printers.

Gray Scale

Image resolution plays an important role in the quality of the result. Successful image scanning requires a balance between two extremes: gray scale level and resolution. Gray scale refers to the number of shades of gray expressed in the image. Photographic images represent a continuous tone image; that is, a large number of gray shades from black to white appear in the picture. The fact that commercial presses use only black in printing work makes it impossible to print a continuous tone image without first changing the image to a halftone. A conventional halftone image results by taking a photograph of the original image through a screen. Halftone dots appear on the copied image; larger dots represent darker gray areas and smaller dots lighter gray. A scanner cannot discern differences in shape among halftone dots. As a result, the digitized image uses a basic unit, a spot, that consists of a cell of at least four dots to express each dot size recognized in a conventional halftone. A cell of four dots would allow for recognition of sixteen different levels of gray scale. Increasing the number of dots in the cell would allow for successful conversion of more gray levels; 8 dots in a unit would allow for 64 levels, 16 dots for 256 levels of gray.

Resolution and the Result

Unfortunately increasing the gray scale level decreases the resolution of the image. Resolution, usually measured in dots per inch (dpi) reflects the clarity or sharpness of the image. Most scanners priced under $50,000 provide resolution of 400 dpi or less. Because the digitized resolution does not equate with halftone resolution, actual resolution seen in a scanned image results from resolution (dpi) divided by the number of dots in each unit (four for sixteen gray scale level, etc.). Thus, a 300 dpi scanner using a 300 dpi laser printer reading sixteen different levels of gray will render a true image of seventy-five units per inch. Increasing the gray scale level will result in an image of less than seventy-five units per inch (which is less than newspaper quality). As a result publication quality printing (at least 150 units per inch for magazines) requires a commercial contractor. Commercial presses use printers with 2,540 dpi, which render magazine quality work even with images scanned at 256 levels of gray.

Important issues also arise in considering the feasibility of image scanning. First storage: one image scanned at 300 dpi requires 1 to 1.5 megabytes (MB) of disk storage. Furthermore, the quality of the result depends on the scanner's management of line art, continuous tones or half tone dots on an image. Finally, the buyer requiring color scanning restricts himself to few products; ScanMaster and Spectrafax market such machines.

LIMITATIONS

Consumers should recognize that scanners differ and each has some sort of drawback. Price often relates to flexibility of a particular machine; the more expensive machines offer the widest number of features.

Independence

Some scanners require more human interaction than others. For example, many scanners lack document feeders so that the operator must manually feed each page into the machine. Even scanners that offer document feeders offer problems of jamming and misalignment. Others prevent the host personal computer from multitasking during scanning operation. Such scanners obviously reduce the desired savings in manual labor offered by scanning technology.

Fonts, Font Sizes, and Spacing

Ideally all scanners should read all documents whether created within the last 500 or 5 years. Unfortunately scanners differ widely in recognition capability. Many scanners specify minimum and maximum document sizes, though most can read the standard letter-sized document. Only the most sophisticated scanners (generally beyond the desktop budget) scan handwritten documents; such technology serves the needs of the Internal Revenue Service and other government agencies. All scanners do not scan all fonts equally well. Most scanners fail to read regular dot matrix printing. Template matching scanners usually scan a limited number of fonts and read only monospaced text; more expensive, intelligent scanners handle a larger number of fonts, font sizes, and variously proportionally printed texts. Still even the most sophisticated scanners may experience difficulty with crumpled, illegible, and faded documents.

Text and Images

Most companies design scanners to read primarily text or graphics; reading both requires tradeoffs in the quality of one or the other. Scanners that can handle both text and images may require separate software products. Often scanning of text and images requires multiple passes of a document, increasing the total per page scan time. Hand scanners usually read one line of text or one 2-by-4-inch graphic per pass.

Storage Capacity

Extensive scanning, especially for textbook revision, will require a substantial amount of disk storage capacity. Scanning of images necessitates a significant cache of memory; a conservative estimate in the literature concludes that one 8½-by-11-inch color graphic needs at least one megabyte (MB) of storage. Scanning at high gray scale levels and resolution further strain memory needs. Serious scanning applications may justify purchase of a minicomputer.

Accuracy and Speed

Although some scanners compete well in accuracy and speed with a seasoned typist, the technology has not attained perfection. Even a track record of 99 percent accuracy still leaves the scanner operator with ten mistakes on a sixty-line page. In addition, most scanners still spend ten to thirty seconds reading each page. Thus scanners have not completely replaced the typist/proofreader.

MARKET STRUCTURE

In a general sense four different types of products constitute today's scanner market: high volume readers for data processing, sophisticated image scanners for design and engineering work, OCR page readers for word processing, and text and image readers manufactured for desktop publishing. This report focuses only on the last type, since the specialized scanners do not meet the diversified needs of the health sciences campus. One may further divide the electronic publishing scanners into three general classes: hand scanners, basic page readers, and sophisticated page readers. Market tiers arise from major differences in price and function. A discussion of the three scanner groups follows.[1]

SCANNER GROUPS

1. Hand Scanners

The first group, hand scanners, has made impressive progress during the last few years. Hand scanners, such as the Saba Handscan by Saba Technologies ($649.95), excel at reading small images of no more than two to four inches wide or one line of text in a single pass. Hand scanners no longer read only certain type faces; the new Transimage 1000 by Transimage Corp ($2,595; applications developer an addition $395) uses feature recognition technology to read most fonts, including proportionally spaced text. While hand scanners present an attractive option for those interested in scanning an occasional graphic or a few pages of text, hand scanners cannot meet the challenge of high volume work required in the desktop publishing environment. As a result specifications for hand scanners do not appear in the appendices.

2. Basic Page Scanners

Desktop publishing requires the services of the page scanner. Most page scanners fall into the intermediate price range, between $1,000 to $6,000. Most of these scanners operate by template matching instead of feature recognition. Such scanners use template matching to read a range of fonts, and some even read letter-quality dot matrix printing. When faced with a font not found in the template cache, the scanner will either create rejection or substitution errors (fail to read the font or insert incorrect characters in the electronic text). Until recently scanning of proportionally published text fell beyond the realm of the mid-priced page scanner. Several companies, including Apple, Calera Recognition, and Kurzweil Computer, have released machines that read typeset text.

Kurzweil brought the feature recognition software into a lower priced machine. Recently, Kurzweil reduced the price on its new Model 5 to below $4,000 and added a limited graphics capability. Calera Recognition Systems has joined the price war by launching Truescan as a high option upgrade package for moderately priced scanners. See Appendix B for some of the more popular, moderately priced page scanners.

3. Sophisticated Page Scanners

The most flexible scanners remain in the upper tier of the market. In the early 1980s the most sophisticated page readers stood beyond the budgets of most businesses, since machines started at $20,000. The market continues to change at an incredible pace. Four out of the five scanners listed in Appendix C cost less than $20,000. Machines at the high end of the desktop market provide true omnifont character recognition, graphics scanning, and high accuracy. Increase in price at the high end of the market offers additional features, including extensive document feeder trays, foreign language capability (Kurzweil machines), and greater flexibility in accepted document and font sizes. The upper level of the market provides for reading almost any type face, including proportionally spaced typeset text.

MARKET VOLATILITY

Prices of scanners have changed radically over a short period of time; prices will continue to fluctuate. In fact the scanner market represents one of the most volatile areas in the computer industry. Dataquest, the subsidiary of Dun and Bradstreet and publisher of a number of market survey periodicals, refuses to publish a scanner quarterly due to extreme variations in price structure. Most companies do not guarantee prices for more than thirty days. New scanner product advertisements appear in almost every new issue of the popular computer periodical literature. Already the market has consolidated; with a reduction in prices many features previously found on the most expensive page scanners have now surfaced on hand scanner models. The greatest competition appears in the moderately priced group of scanners where new products appear on a weekly basis. The leaders at the high end of the desktop scanner market, Kurzweil Computer and Calera Recognition Systems, have even introduced products in the medium price range.

THE MARKET PLAYERS

The market includes six major manufacturers of scanners: Apple, Calera Recognition Systems, Datacopy Corp, Dest, Kurzweil Computer, and Microtek. A list of scanner manufacturers appears in Appendix D. Some of the more popular and highly rated products appear in Appendix A and Appendix C. A comprehensive analysis of the entire market falls beyond the scope of this report. Appendix E lists reviews of scanner products.

Until recently the demarcation of three market tiers exemplified the diverse interests of scanner manufacturers. With the reduction in the cost of scanner technology, many companies have tried to compete in more than one tier of the market. CompuScan now produces both hand and moderately priced page scanners. Kurzweil has reduced the cost of its low end model in the 7320 series to compete with low cost manufacturers. Calera Recognition Systems has launched Truescan as an upgrade option for those interested in the lower end of the market. Xerox Imaging Systems recently acquired Datacopy Corporation and Kurzweil Computer; this will probably reinforce Kurzweil's position as leader, especially at the top of the desktop market. Other companies, such as CompuScan, Datacopy, and Dest, continue to concentrate on the moderately priced sector of the market.

Two trends characterize the recent history of the scanner: sales growth and price reductions. For example, Datacopy Corporation has posted a sales increase of 50 percent every year since 1984. Price reduction constitutes the other major industry trend. Before Kurzweil advertised its model 4000 (now the model 5000), the least expensive Kurzweil machine costs over $70,000. Calera Recognition Systems has cut the price of its 9000 series by almost $10,000.

PRODUCT SELECTION

With the turbulence in the scanner market successful identification of a particular product will only result from proper assessment of needs, price range, and timetable for purchase. Some of the following guidelines may simplify the process:

- Determine the primary application for the machine. Occasional scanning of documents of conventional size and font justifies purchase of a moderately priced page scanner. Textbook revision and intensive desktop publishing necessitate a more flexible scanner with omnifont capability.
- Will applications include image scanning or OCR only?
- Identify the minimum level of speed and accuracy required. Examine the mechanics of error correction, compatibility with word processing software, and availability of background operation.
- Question the hardware compatibility of the scanner. Many scanners not only work with IBM and Macintosh systems but with minicomputer workstations as well.
- Will the company upgrade the scanner as your needs change and expand?
- Does the company offer free demonstration, training, and hotline support? What is the learning curve for a particular scanner? Will the company supply a list of other institutions who have purchased the scanner?

FUTURE TRENDS

The scanner market will continue to change at a very rapid pace. Based on recent changes in market structure and the growing diversification of interest among participating companies, more change appears on the horizon. Prices will probably continue to fall; less expensive models will include more features. Machines will read more fonts, with better resolution and at greater speeds. Scanner manufacturers will continue to cater to the needs of both the IBM and the Macintosh market. In addition companies will try to package several machines into one; the future will probably bring a single box capable of sophisticated image and graphics scanning, telefacsimile operations, and conventional photocopying.

ACKNOWLEDGEMENTS

The author wishes to acknowledge the support and encouragement of Nina W. Matheson and Karen A. Butter. The author appreciates the generous assistance of Jayne Campbell for her critical reading of the paper. The article reflects the personal views of the author, who accepts responsibility for any errors or misrepresentations.

REFERENCES AND NOTES

1. The desktop market includes two other scanner technologies. The overhead scanner resembles an overhead projector with two light sources shining down on the document. The printhead scanner attaches to the head of a printer and scans much as a printer prints moving back and forth across the document. This report excludes the overhead and printhead scanners, since both technologies offer less accuracy than the more conventional scanners.

APPENDIX A. PRODUCT FEATURES

Each of the products listed in Appendices B and C include specifications of some of the scanner's features. An exhaustive list of all technical data falls beyond the scope of the report. A list of scanner manufacturers appears in Appendix D. Most companies will send telefacsimile copies of specifications. In addition many companies make referrals to local computer outlets or regional representatives for personal demonstrations. For additional information see Appendix E for a list of suggested readings. Appendix B and C include specifications for the following scanner features:

Company Name—The name of the scanner manufacturer.
Product Name—The series and model name of the product.
List Price—Approximate scanner price. Indication if price includes or excludes software.
Flatbed or Sheetfed—Flatbed indicates that the scanner mechanism moves under the document, which lays on a glass plate. A sheetfed scanner moves the document over a fixed scanner head. The flatbed scanner with a document feeder offers the best of both worlds, since it handles both bound documents and loose sheets.
Auto Document Feeder/Capacity—A bin that stores loose documents for feeding into the scanner.
Minimum/Maximum Sheet Size—Range of document sizes acceptable to the scanner.
Resolution (dpi)—Maximum resolution or sharpness with which a scanner scans a document. Some scanners have a range of resolution settings; the lower the setting the less storage capacity used by the scanned image.
Speed Rating (cps)—Scanning speed represents the manufacturer's specifications. Some manufacturers offer the rating in pages per second. Ac-

tual scanning time depends upon the proportion of text to graphics on the page. Most scanners range from 10 to 30 seconds per page.

Accuracy Rating (%)—The rating refers to the substitutions (the scanner uses an incorrect character) and rejections (the scanner fails to read a character) made by the scanner. Prospective buyers should test personal documents prior to purchasing a particular model.

Fonts Sizes Read—Range of font sizes read by the scanner.

Fonts Recognized—The font styles read by the scanner. Moderately priced scanners read only limited fonts, while the more expensive fonts have omnifont capability. Some manufacturers interpret omnifont as virtually all fonts, while others claim up to 20,000 fonts read. Actual demonstrations should offer realistic results. Many machines can read only letter quality dot matrix print. Relatively few machines read foreign characters.

OCR Method—Refers to the software technology used for recognition of text. IPR refers to Intelligent Page Recognition (Calera's feature extraction system), and ICR refers to Intelligent Character Recognition (Kurzweil's proprietary system). Less expensive models usually employ template or matrix matching, which offer a limited variety of recognized fonts.

Background Operation—Whether the host computer can perform other jobs during scanning operations.

Gray Scale—The maximum operating gray scale level.

APPENDIX B. BASIC PAGE SCANNERS

Company Name	Apple	Dest	Microtek	Calera Recognition	Kurzweil Computer	Datacopy
Product Name	Apple Scanner	2020	MSF-300G	Truescan Works with low cost scanner	Model 5 OCR and basic line graphics	JetReader
List Price (i.=includes software, e.=excludes software)	$1300i. graphics only;text w/optional software e.g. Omnipage $900	$2,690i.	$3,495i. graphics only; text optional*	$3,000i.	$3,995i.	$1,300e.*** software additional $1400.
Flatbed or Sheetfed	Flatbed	Flatbed	Sheetfed	N/A	Sheetfed	Sheetfed
Auto Document Feeder/ Capacity	No	Optional/(35 pages)	No	N/A	Yes/(10 pages)	Yes/(10 pages)
Minimum/Maximum Sheet Size	8.5 x 14 in. (max.)	3.5 x 3.5 to 8.5 x 14 in.	Definable/ 8.5 x 11	Scanner dependent	5.25 to 8.5 by 4.25 to 14 in.	5.75 to 10 by 4.25 to 14.25
Resolution (dpi)	75 to 300 dpi	38 to 300 dpi	300 dpi	300 dpi	300 dpi	300 dpi
Speed Rating (cps)	20.4 seconds per page	9.4 seconds per page	9.9 seconds per page	75-100 cps	30 to 120 seconds per page	30 seconds per page
Accuracy Rating (%)	Software dependent	90%	Software Dependent	99.9+	99.64+	95%
Font Sizes Read	• •	10, 12	• •	6 to 28 pt.	8 to 24 pt.	10-18 pt.
Fonts Recognized	• •	Limited	• •	Flexible	Omnifont	Flexible with Truescan
Dot Matrix		Yes		Yes	Letter Quality	Letter Quality
Typeset	• •	As a graphic	• •	Yes	Yes	w/Truescan
Foreign Characters		• •		No	No	Romance Languages
OCR Method	• •	Matrix Matching	• •	Intelligent	Intelligent	Template matching
Interface	MAC	IBM, MAC	IBM, MAC Sun compatible**	IBM compatible	IBM compatible, MAC forthcoming	IBM compatible, MAC compatible
Background Operation	No	No	•	No	Yes	Yes
Gray Scale	16 levels	256 levels	256 levels	Scanner dependent	16 levels	16 levels (730gs sells for $1300 offers 64 levels

*Omnipage available for MAC in 2-3 weeks @$795; Omnipage for PC requires purchase of additional card @$2,495.
**MSF300G for use with Sun costs $500 more than the MAC and PC model
***The scanner comes without software; separate packages available for OCR and graphics. IBM and MAC versions require separate software.
•Information not available at from company at present time.

APPENDIX C. SOPHISTICATED PAGE SCANNERS

Company Name	Kurzweil Computer	Kurzweil Computer	Kurzweil Computer	Calera Recognition	Calera Recognition
Product Name	Discover 7320 Model 20	Discover 7320 Model 30	Discover 5000	CDP 3000 Series	CDP 9000 Series
List Price (i.=includes software, e.=excludes software)	$4,995i.	$6,995i.	$15,950i	$15,000i	$30,000i
Flatebed or Sheetfed	Flatbed	Flatbed	Flatbed	Sheetfed	Sheetfed
Auto Document Feeder/ Capacity	Yes/(10 pages)	Yes/(30 pages)	Yes/(50 pages)	Yes/(50 pages)	Yes/(50 pgs)
Minimum/Maximum Sheet Size	5.25x4 in. to 8.5x14 in.	7.125x5.125 in. 8.5x14 in.	3x3 in. to 11x14 in.	3x5 in. to 8.5x14 in.	3x5 in. to 8.5x14in.
Resolution (dpi)	300 dpi	300 dpi	400 dpi.	300 dpi.	300 dpi
Speed Rating (cps)	30 to 120 seconds per page	30 to 120 seconds per page	30 to 120 seconds per page	100 cps	250 cps
Accuracy Rating (%)	99.64+	99.64+	99.64+	99.0	99.9
Font Sizes Read	8 to 24 pt.	8 to 24 pt.	6 to 24 pt.	6 to 28 pt.	6 to 28 pt.
Fonts Recongnized Dot Matrix Typeset Foreign Characters	Omnifont Letter Quality Yes 5 languages	Omnifont Letter Quality Yes 7 languages	Omnifont Letter Quality Yes 7 languages	Omnifont Yes Yes No	Omnifont Yes Yes No
OCR Method	ICR	ICR	ICR	IPR	IPR
Background Operation	Yes	Yes	Yes for text only	Yes with Sun only	Yes with Sun only
Gray Scale	•	•	64	•	•

•Information not available from company at present time.
•Both the CDP 3000 and the CDP 9000 do OCR and line art only and therefore do not have gray scale values. Calera offers the Recognition server 9000 ($27,000) for use with an external scanner for graphics other than line art.

APPENDIX D. DIRECTORY OF SCANNER MANUFACTURERS

Advanced Vision Research
2201 Qume Drive
San Jose, CA 95131
(408) 434-1111
1-(800) 544-6243

ANA Tech Corp.
10499 Bradford Road
Littleton, CO 80127
(303) 973-6722

Arc Laser Optical Technology
10 Victor Square Suite 600
Scotts Valley, CA 95066
(408) 438-7400

AST Research Inc.
2121 Alton Avenue
Irvine, CA 92714
(714) 863-1333

AT&T
Audiographics Communications Systems
85 Monmouth Parkway
West Long Branch, NJ 07764
(408)432-6000
1-(800) 227-6161

Calera Recognition Systems
2500 Augustine Drive
Santa Clara, CA 95054
(408) 986-8006

Canon U.S.A. Inc.
One Canon Plaza
Lake Success, NY 11042
(516) 488-6700

Computer Aided Technology, Inc.
7411 Hines Place, Suite 212
Dallas, TX 75235
(214) 631-6688

Datacopy Corporation
1215 Terra Bella Avenue
Mountain View, CA 94043
(415) 965-7900

Dest Corp.
1201 Cadillac Court
Milpitas, CA 95035
(408) 946-7100

Eikonix Corp.
(Division of Kodak Company)
23 Crosby Drive
Bedford, MA 01730
(617) 275-5070

Electronic Information Technology, Inc.
25 Just Road
Fairfield, NJ 07006
(201) 227-1447

Epson America, Inc.,
Computer Products Division
2780 Lomita Boulevard
Torrance, CA 90505
(213) 539-9140

Fujitsu America, Inc.
Storage Products Division
3055 Orchard Drive
San Jose, CA 95134
(408) 432-1300

Hell Graphics Systems
25 Harbor Park Drive
Port Washington, NY 11050
(516) 484-3000

Hewlett-Packard Company
1820 Embarcadero Road
Palo Alto, CA 94303
(415) 857-1501

Howtek, Inc.
21 Park Avenue
Hudson, NH 03051
(603) 882-5200

I-SYS Technology, Inc.
19151 Parthenia Street
Suite E
Northridge, CA 91324
(818) 886-9303

Imagitex, Inc.
77 Northeastern Boulevard
Nashua, NH 03062
(603) 889-6600

Kurzweil Computer Products
185 Albany Street
Cambridge, MA 02139
(617) 864-4700

Laser Connection
7852 Schillinger Park West
Mobile, AL 36608
(205) 633-7223

Lasodyne
2522 Chambers Road Suite 112
Tustin, CA 92680
(714) 544-2751

Micro Technologies, Inc.
921 Royal Heights Road
Belleville, IL 62223
(618) 233-2314

Microtek Lab, Inc.
16901 South Western Avenue
Gardena, CA 90247
(213) 321-2121

NBI, Inc.
P.O. Box 9001,
3450 Mitchell Lane
Boulder, CO 80301
(303) 444-5710

APPENDIX D. DIRECTORY OF SCANNER MANUFACTURERS (CONTINUED)

Princeton Graphic Systems
601 Ewing Street, Building A
Princeton, NJ 08540
(609) 683-1660

Saba Technologies
9500 SW Gemini Drive
Beaverton, OR 97005
(408) 432-6000

Shape, Inc.
Electronics Instruments Division
301 U.S. Route 1
Scarborugh, ME 04074
(207) 883-2931
1-(800) 247-1724

TEC America, Inc.
20 Walnut Street
Wellesley Hills, MA 02181
(617) 235-4422

Thunderware, Inc.
21 Orinda Way
Orinda, CA 94563
(415) 254-6581

TPS Electronics
4047 Transport Street
Palo Alto, CA 94303
(415) 856-6833

TransImage Corp.
910 Benicia Avenue
Sunnyvale, CA 94086
(408) 733-4111

Wang Laboratories, Inc.
One Industrial Avenue
Lowell, MA 01851
(617) 459-5000

APPENDIX E. SUGGESTED READINGS

"All about Scanners and OCR Devices," *Datapro Reports on Office Systems,* (Delran, N.J.: Datapro Research Corporation, 1988).

Donna Barron, "Scanners Improve Publishing's Image," *PC Week,* Jan. 30, 1989, p.s/8,10,20–26.

Stephen Beale, James Cuvuoto, and Aileen Abernathy, "A Touch of Gray," *MacUser,* Feb. 1989, p.257–67.

Pat Bellama, "OmniPage OCR Software Snares PC Users' Attention, *PC Week,* Jan. 9, 1989, p.27,32.

Steven Bobker, "Quick Clicks," *MacUser,* Feb. 1989, p.61–62.

Lauren Brown, "Hand-Held Scanners Offer Affordable Option," *PC Week,* Mar. 15, 1988, p.55.

________, "Seconds-per-Page Feature Blurs True Scanner Speed," *PC Week,* Mar. 15, 1988, p.55–60.

T.J. Byers, "Saba Handscan," *PC World,* Oct. 1987, p.188–91.

________, "Saba PageReader," *PC World,* Feb. 1988, p.185–7.

Mike Falkner, "Hand Scanners Inch Forward," *PC Magazine,* Nov. 29, 1988, p.277–309.

Curt Franklin, "The TransImage 1000: Versatile OCR in a Low-cost Package," *Byte,* Jan. 1988, p.98.

"Gammafax Gives Good Performance at Reasonable Price," *Computer and Software News,* Aug. 22, 1988, p.64,69.

Paul W. Hamby, "A Look at Saba and Compuscan OCR/Scanners," *Lawyer's PC,* Sept. 1, 1988, p.6–9.

Thom Hartmann, "Six Macintosh Scanners," *The Macintosh Buyer's Guide,* Winter 1989, p.18–27.

Jim Heid, "A Publisher's DEST Friend," *MacWorld,* October 1988, p.160–2.

Mark Henricks, "OmniPage: No Bit Player," *PC World,* p.80,82.

Richard Jantz, "The 300-Word-Per-Minute Typist," *Publish!,* April 1988, p.64–73.

Peter H. Lewis, "Here's an Easier Way for You to Process Mountains of Paperwork by Computer," *San Francisco Chronicle,* Dec. 27, 1988, p.B-4.

________, "Transferring Paper into Memory," *The New York Times,* Aug. 21, 1988, section 3, p.9.

Owen Linzmayer, "Picture This: A Comparison of Scanners," *Nibble,* Jan. 1988, p.38–40,43.

John McCormick, "Compuscan OCR Scanners Prove Fast and Accurate," *Government Computer News,* Nov. 7, 1988, p.31.

________, "OCR Scanner Uses Artificial Intelligence Techniques," *Government Computer News,* Sept. 12, 1988, p.26.

________, "Text Scanners for the IBM PC," *Byte,* Apr. 1987, p.233–8.

Clement J. McDonald, "Optical Readers for Clinical Data," *M.D. Computing,* Jan./Feb. 1988, p.10–12.

Brita Meng, "The Reading Edge," *MacWorld,* Feb. 1989, p.170–9.

Ted Nace, "JLaser Plus: Scanner/Laser Printer Controller Board," *PC World,* June 1987, p.138–41.

Christopher O'Malley, "What's New in Scanning," *Personal Computing,* Mar. 1989, p.103–10.

"Optical Character Recognition and Image Scanning," *Electronic Office Management* *and Technology*, v.2, Faulkner Technical Reports, Inc., 1988.

Jon Pepper, "Panasonic Enters Scanner Market with 400-dpi Model," *PC Week*, Sept. 1, 1987, p.87.

Lewis Perdue, "Saba's Page Reader Is an Easier-to-use OCR Scanner," *Computer and Software News*, Feb. 29, 1988, p.55,57.

Philip Robinson, "The Paper Chase," *PC Computing*, Nov. 1988, p.164–79.

Steve Rosenthal, "Hand-operated Scanners Capture Graphics Handily," *PC Week*, Jan. 9, 1989, p.97–8.

Tom Stanton, "From Fax to Finish: Scanners Build a Better Image," *PC Magazine*, Mar. 28, 1989, p.187–258.

________, "Handscan Reads Text Line by Line Directly into Applications," *PC Magazine*, June 9, 1987, p.33–35.

________, "Kurzweil Desktop Scanner OCR Breaks New Ground," *PC Magazine*, July 1987, p.33,36.

________, "Scanners Take Off," *PC Magazine*, Oct. 13, 1987, p.185–321.

Ben Templin, "Character Analysis," *MacUser*, Sept. 1988, p.295–301.

"TrueScan," *Byte*, Jan. 1989, p.330.

Bruce Webster, "Apple Scanner," *MacWorld*, p.220–21. ■■

READS: A Networked PC System

Jane B. Mandelbaum

More than thirty small libraries for the blind and physically handicapped are now operating "state of the art" local area networks (LANs) using a custom system developed by the National Library Service (NLS) for the Blind and Physically Handicapped of the Library of Congress. A LAN is generally defined as a network of interconnected personal computers with an operating system and hardware that permit sharing of data and peripherals. The NLS READS (Reader Enrollment and Delivery System) was specifically designed for small libraries—libraries with one to five full-time staff members. Since the development of the system, these libraries have been able to take advantage of high-technology applications as well as any other organization.

Jane B. Mandelbaum is Automation Officer, National Library Service for the Blind and Physically Handicapped, Library of Congress.

READS was developed by NLS to automate the functions of smaller libraries that are part of its network of cooperating libraries. Software development began in 1984, and a production version of the system has been in use since 1986. The basic system configuration, which has remained unchanged, is a central database on a local area network (LAN) file server accessed by multiple workstations. Both the file server and the workstations are IBM-compatible personal computers, and the LAN uses the Novell Netware operating system. READS is written primarily in PASCAL, with several commercial PASCAL-based utilities for data management. NLS continues to update the recommended equipment based on changes in the marketplace and also provides periodic upgrades to the application software. Although each new READS library procures up-to-date hardware and the newest Netware version, NLS ensures that the READS software supports all existing users.

SUMMARY OF THE NLS NETWORK

To understand the basis of the READS software, it is necessary to understand the operations of libraries for the blind and physically handicapped, whose automation needs are similar to those of other public libraries but with some striking differences.

NLS and its network of cooperating libraries provide blind and physically handicapped individuals nationwide with recreational reading materials similar to those in any public library collection. Both books and magazines are produced in braille and on recorded disc and cassette. To support the production and distribution of the books and magazines, as well as special playback machines to the libraries, NLS maintains central computer systems similar to those in any large organization.

However, much more information must

be maintained in the cooperating libraries than in libraries that circulate print materials to walk-in readers. The network of 56 regional (state) and more than 100 subregional (local) libraries circulates the books, special playback equipment, and other materials to a total readership of approximately 700,000 adults and children. A typical library in the network maintains a collection of over 40,000 titles, with several copies of each for a total of 100,000 items. The libraries vary in staffing from less than one full-time employee to more than fifty.

In libraries for blind and physically handicapped individuals, patrons do not typically come into the library and select their own materials. Most of the contact between the patron and the library is conducted over the phone or through the mail. Virtually all materials are sent to readers and returned through the mail.

Library staffing, office space, and all operating expenses are funded by the library's administering agency, while the reading materials and playback machines are furnished by NLS. Many of the libraries rely heavily on volunteers to perform the daily labor-intensive tasks of check-in, checkout, shelving, and book and machine inspection.

REQUIREMENTS OF A LIBRARY FOR THE BLIND SYSTEM

For each patron, the library staff generally performs the following functions:

- register the patron for the service;
- collect information on the patron's reading interests and preferences;
- register the patron for magazine subscriptions;
- select books for the patron based on the patron's reading interests, and record patron's selections from telephone calls or printed book order forms;
- mail non-circulating materials to patrons on request;
- check-out books and magazines and mail these to patrons;
- check-in materials;
- maintain information about playback machines on loan to patrons; and
- maintain records of all books the patron has been sent previously.

Both the book and patron information required is more extensive than in a "standard" library circulation system. The book record provided by NLS is cataloged from a modified MARC record that includes additional data items, such as the book medium, narrator (for recorded books), and a brief annotation that describes the book. The patron data maintained by each library includes extensive information about the patron's reading interests and preferences.

DEVELOPMENT OF AUTOMATED SYSTEMS IN LARGER LIBRARIES

In 1981 Cuadra Associates, Inc., prepared a study of automated circulation systems in the NLS network of libraries. Representative comments from library staff about the need for automation focused on the inability of manual processing to keep up with patron demand for timely service and on the "tedious and repetitive" nature of the paperwork. An important strategic goal for the network was to automate the large volume of paper-based transactions so that the libraries could provide more effective service to the patrons of the NLS program.

By 1983 most of the larger libraries were operating automated systems that greatly reduced the clerical tasks common to all the cooperating libraries. The automated systems required extensive customized programming to perform the complex tasks of matching patron interests and requests with books available, taking into account how many books the patron wants and how often and which books the patron has already read.

The specialized requirements were successfully written into a software package offered by library automation vendor Data Research Associates (DRA) for libraries with the resources for a minicomputer, software licensing fees, and a computer operator. Thirteen of the largest regional libraries currently operate with DRA software.

Approximately ten regional libraries have developed their own minicomputer- or mainframe-based systems, usually with computer and programming resources allocated from a shared computer center.

CONCEPT OF A MICROCOMPUTER-BASED SYSTEM FOR SMALLER LIBRARIES

Smaller libraries typically serve up to 6,000 patrons with a very small operating budget. The NLS network has more than 100 libraries in this category. Smaller libraries could obviously also benefit from automation but were unlikely to ever have the resources to support a minicomputer-based package or a full lifecycle of system development.

With the growth of the microcomputer, a new option for a network automation strategy was developed. NLS staff proposed a software package for smaller libraries, one that could operate on a network of microcomputers acquired by the libraries according to their staff size and resources. The system was designed and developed by NLS working with a contractor, Mobius Management Systems. System requirements were based on the analysis presented in the Cuadra report and NLS staff members' extensive knowledge of library operations.

NLS makes available executable copies of the software and user documentation without charge to qualified libraries that serve blind and physically handicapped individuals.

Original expectations in 1984 for this "small library" system were clearly in line with the optimism of the first stage of the microcomputer era. These expectations included:

- The system would operate on a network of microcomputers linked together; at that time, it was assumed that the system would not be able to fit on a single standard microcomputer, because the 40-100 megabytes for the files were larger than the hard disks available.
- With a simple installation guide to follow, libraries would be able to easily install the hardware and software on their own.
- The system would be "turnkey" so that once it was installed and started up, the library would be able to operate it without technical assistance.

During the initial design phase, the exact hardware components required by a user were not specified. However, it became clear quickly that because of the lack of standardization in the microcomputer world, the READS software could not be guaranteed to work on any hardware configuration other than the precise ones it was already operating on. Therefore, to provide guidance to prospective users when the system went into production in 1986, NLS began to describe READS as a "package" that included detailed specifications for the hardware and even the operating supplies. The original package included industry standards—the IBM PC/AT and IBM PC.

Personal computer hardware, LAN topologies, and Novell Netware have all been upgraded dramatically since 1986, providing new READS users with a greater variety of faster and more reliable products from which to choose. Because the basic configuration has remained consistent, existing READS users can easily upgrade one or more of their older pieces of equipment to enjoy the benefits of these improvements. NLS advises prospective users of the equipment that is known to work at existing libraries and provides test copies of the READS software for libraries that want to try untested products.

Currently, a typical new READS library acquires the following equipment and system software:

- a LAN file server using a 80286 or 80386 processor, which performs the database access and processing;
- an internal or external hard disk (sized according to the circulation volume of the library) attached to the file server to store the central READS database of the library patron and bibliographic files;
- multiple personal computers used as staff workstations;
- LAN hardware (Ethernet or Arcnet network interface cards and cables) used to connect the workstations to the file server;
- a shared dot-matrix network printer to produce mailcards for mailing materials to patrons, as well as to print management reports;
- an external cartridge tape backup unit compatible with the customized READS backup and restore programs;
- an uninterruptible power supply unit to condition the incoming power and to allow the users to bring the file server down

gradually in the case of a power failure;

- one or more light-pen scanners used to read OCR-A characters from book copy labels, mailcards, and patron order forms;
- Novell Netware operating systems software (SFT version 2.1); and
- third-party commercial software utilities used by READS, including a data management package, a sort package, and MS-DOS.

OFFERING THE SOLUTION

The READS approach has broad appeal to the libraries in the network. Several key elements contribute to this appeal:

Integrated library concept. The system design includes a full range of library functions, from circulation to tracking of machine use by patrons and generation of machine-readable transactions for the central NLS master system used for direct distribution of magazines and catalogs. Therefore, READS presents an integrated automation strategy that addresses all aspects of a library's workflow. By design, the system offers opportunities for additional modules and enhancements in the future.

Pre-approved standardized features. Because the libraries follow guidelines established by NLS, a library could assume that software developed by NLS would follow applicable processing guidelines and that the system would have the NLS "stamp of approval."

Complete automation solution. With integrated software and a specified hardware configuration, a library has a full-blown existing library "solution" to present to their management for approval and funding.

Pre-packaged systems analysis and implementation process. READS is pre-tested, pre-approved by NLS and integrates most library functions, so a library can easily skip almost all phases of the system planning lifecycle up to installation. The library does not have to perform a cost-benefit study or evaluation of options, development, or testing. There is a checklist of required steps for READS planning, and NLS provides guidance and information at every step. Even data conversion is facilitated, since NLS extracts data from three master files for the library's primary data files.

Reduction of paperwork. Although no specific claims were made that a library could reduce staffing or costs, READS is specifically designed to eliminate many clerical tasks and almost all paper recordkeeping and filing.

THE LURE OF THE LAN

Some of these appealing features are characteristic of any commercial software package produced by a reliable vendor. However, READS goes beyond this level by offering a LAN—a hardware configuration that had other advantages that are not as common in the library world. The LAN configuration offers the following:

Reasonable and flexible costs. READS is a package of low-priced standard components, none of which cost more than $5,000 and all of which can be purchased separately if needed. A library can control the costs by controlling the number of required workstations. The library can add to the configuration easily over time, and it can upgrade individual components as needed.

Minimal impact on physical environment. The READS LAN fits into the small library environment easily; the microcomputers fit on desktops without any changes in the physical layout or library configuration.

Automation with PCs, rather than a "real" computer. The selling points of a network of microcomputers are obvious: the cost of each processor is low, microcomputers are now standard office commodities available at retail outlets, and they are even adaptable for other purposes. The ordinariness of this machine compares favorably with typically perceived requirements for a "computer," such as a special computer room, air-conditioning, and additional electrical wiring and security. The exclusive use of the lowly microcomputer increases the "comfort" factor of READS. The microcomputer is perceived as a user-friendly tool that can be used right out of the box like a typewriter.

"Turnkey" hardware package. When READS was first offered as a production system, NLS specified the brand names and models for all required equipment as well as exact quantities of supplies. Even hard-disk capacity is calculated for the library by

the use of a spreadsheet formula incorporating twenty readership and circulation statistical parameters.

Future flexibility. Since READS consists of modular components, each component can be replaced as needed because of equipment failure or a need to upgrade. These incremental costs compare very favorably to a system based on a single centralized minicomputer or mainframe, which is costly and risky to replace or upgrade and which almost always requires a conversion.

Customized approach. Since READS was designed specifically for the needs of a cooperating library for the blind, library staff can justify the requirement for an independent system that is not simply an extension of a "standard" library circulation system. The exacting requirements of the READS solution and the exotic nature of the LAN makes it easier for the prospective READS library to justify their unique functional requirements for funding by their administering agency or "friends of the library" groups.

BENEFITS

With these appealing features, READS was launched in 1984 with pilot installations at four libraries. After a successful pilot period, READS is now installed in more than thirty libraries, with an estimated thirty more installations likely over the next three to four years. NLS continues to support the READS software in place at the libraries by providing limited technical assistance and periodic enhancements from a systems contractor. Support and guidance are also available for prospective READS user libraries that are planning to implement READS.

Overall, READS is a carefully designed system that provides a library with extensive functional benefits. It has easy-to-use screens with standardized function keys, and a user can quickly learn to perform data entry and retrieval without referencing the user manual. It provides:

Improved service to patrons. READS provides libraries with most of the expected functional benefits: the library has better control of its collection and is able to increase circulation significantly without increasing staff. Automation relieves staff of clerical tasks and therefore gives them the opportunity for more interaction with existing patrons, more outreach to prospective or new patrons, and more familiarity with the collection. This is particularly valuable in a small library with one or two professionals, who can now perform more of the librarian functions.

Higher quality of information. By reducing the dependence on paper, staff start to see the importance of the quality of the information itself. READS is designed to collect and manage statistical and summary data used by both the library and by NLS. By making data management painless, the library has more reliable and current information about its service.

UNEXPECTED ISSUES

The READS software is an example of a customized integrated library product that performs its functions almost flawlessly and provides the benefits of sophisticated automation to many users in small libraries that have absolutely no experience with computers. Careful planning and the use of pilot sites for testing were also instrumental in the fact that the functions of the software operate well. The screens are well-designed, and the system is easy to learn and to use.

This aspect of the system has worked and has worked well. Libraries are now automated that could never become so without READS. In most situations, READS would seem to be an automation success story beyond our wildest dreams.

The difficult issues that have arisen have been related to the "turnkey" nature of the system and the intention to provide an integrated package that included planning support, equipment specifications, and a step-by-step checklist. Developing the tools for users is not enough; the users need to be able to take the tools and take responsibility for them. Looking back, these issues are typical of discoveries in the computer industry during this period.

In well-intentioned efforts to make READS an easy system to acquire and to ensure it would operate properly, the original READS package specified all the hardware and system software, without making provision for upgrades to the hardware or allowing for deviations dictated by unavailability of particular products or mandates from the libraries' administering organizations. Responsibility for the equipment after the installation date is at the individual library. As a result, users can be, and are, operating an automated system that is critical to their daily work, without acquiring an understanding of how to move ahead with that system.

Based on the experiences thus far, important questions need to be addressed by a small library considering, or operating, a system such as READS:

How Is Data Integrity Maintained?

Microcomputer users soon discover that independence from a centrally controlled computer system means that they are responsible for the content and fate of their own data. On a LAN with a centralized database, this responsibility is even more important. The LAN is used in READS as a substitute for a multi-user minicomputer, and data management has to be treated as seriously as it would be on a minicomputer. No software package can guarantee that the data will always be in the state the user expects it to be.

Use of READS requires daily monitoring of operations and outputs to ensure that the database is accurate. Many actions can result in unexpected changes to the database; for example, failure to exit cleanly from the application software can leave files open, which can cause file corruption. The READS library cannot tolerate this possibility, and the staff must be aware of the consequences and must be able to take appropriate action. Unfortunately, this situation usually requires restoration of a prior copy of the database.

In most microcomputer-based applications, such as word processing or spreadsheets, there are no large interconnected database files that require this kind of "mainframe" recovery processing. Therefore, the READS library staff is not usually prepared to operate with these kinds of constraints and concerns.

Who Will Manage the LAN?

Most LAN advocates (and vendors) say a LAN manager is not needed for a small stand-alone network. However, LAN users will qualify this statement when talking about a LAN running a multi-user database processing system. When a LAN is really a centralized computer system, a trained manager with administrative and technical skills is needed to ensure that the network is operating day-to-day and to deal with strategic issues for the future. If the manager cannot personally repair or diagnose LAN problems, the manager must have the resources available to get problems resolved.

In the READS environment, it is impossible for the small library staff to hire or train a LAN manager or even dedicate staff resources on a regular basis. The administering or funding agency is usually skeptical that a manager or even staff training is required for a "few PCs." However, a successful installation requires someone to be able to diagnose and evaluate both daily and unusual problems. This process includes understanding of the equipment, the software, *and* the database. When the LAN is not operating, the library is not operating.

Complex Technical Configuration: Who Will Be Responsible?

A network adds layers of complexity to any microcomputer-based application. Because of this complexity, LANs cannot yet be taken for granted by a non-technical organization such as a library. LAN hardware selection, evaluation, and operation are technical issues that are usually not the responsibility of the end users. Even the experienced LAN users discover that LAN problems and upgrades are not easily taken care of by professional computer technicians. The components and configurations change at a furious rate, and vendors have tremendous difficulty in keeping a high level of technical support for their products.

However, the READS libraries perceive the LAN as just another component of the READS package. Since the LAN comes with READS and because READS itself is easy to use, the library staff are often not prepared for taking necessary steps to learn more about the LAN and how it works. The inexperience of the staff becomes a disadvantage in both solving LAN-related daily problems and in determining ways to upgrade or replace LAN components as technology improves.

How Will Changes in LAN Hardware Be Addressed?

LAN hardware changes constantly. Instead of a stable environment, READS has unexpectedly become part of the most advanced technological edge of the computer world: local networking. Libraries often do not have the expertise or ability to evaluate options for growth in this environment.

Because of the profusion of options in LAN topologies, cabling, and microcomputers, even LAN installers and vendors are not reliable sources of information, products, or analysis. For example, the first version of a key hardware component from Novell (the LAN software/hardware vendor used for READS) was unreliable, and the components had to be recalled. Products are declared obsolete and are not supported within two years of their initial sales. In this environment, NLS can no longer recommend or specify many of the components for READS, since it is impossible to test every combination of options. This situation puts unexpected responsibility on the library to select options or to find a vendor that will ensure that a configuration will work with READS.

CONCLUSIONS

READS has proven to be a fast way of bringing library staff into automation. The LAN configuration provides flexibility and easier upgrade paths than a centralized minicomputer.

However, the READS "turnkey" concept was a product of the initial optimism of the personal computer era. Each year since then has shown that the development and maintenance of a networked system requires not only planning and implementation but continuing skilled maintenance, management, and, most important, growth. Skills are required in an awesome range of fields, from cabling to user requirements analysis. These skills are simply not in the job descriptions of librarians managing "one-room" libraries.

As LAN technology continues to advance, LANS will become common fixtures in libraries, even in small libraries. READS has positioned libraries for the blind in the automation vanguard, with the opportunity to continue to improve services to patrons while also taking advantage of the opportunities to manage their own long-term automation goals. ■■

Communications

Characteristics of Subject Heading Records in the Machine-Readable Library of Congress Subject Headings

Diane Vizine-Goetz and Karen Markey

Since April 1986 the Library of Congress has been distributing the machine-readable Library of Congress Subject Headings (LCSH-mr) in the form of a cumulative master tape and weekly tape update service. LC's distribution of LCSH-mr has motivated library systems staff, bibliographic service system designers, and integrated library systems vendors to begin developing subject authority capabilities for their respective systems. These capabilities will improve these systems with what patron users want most—the ability to view lists of terms related to their search terms. To launch development activity, system designers need to determine the characteristics of authority records in LCSH-mr.

This paper details the characteristics of authority records for subject headings (MARC tag 150), which are the majority of records in LCSH-mr. Statistics are given on occurrences of variable-length fields and the length and maximum length of such fields in subject heading records. System designers can use these statistics to determine record structures and estimate space allocations for LCSH-mr.

Diane Vizine-Goetz is Research Scientist, Office of Research, OCLC Online Computer Library Center, Inc., Dublin, Ohio. Karen Markey is Assistant Professor, the University of Michigan, School of Information and Library Studies, Ann Arbor, Michigan.

LCSH-MR USED IN THIS STUDY

The LCSH-mr used for this project is a cumulative master tape that contains 160,706 authority records created or modified through July 6, 1987. This version of LCSH-mr includes headings found in the tenth edition of LCSH and additions and changes published in the 1985 annual supplement and LC Subject Headings Weekly Lists through spring 1987. The version of LCSH-mr used in this study almost corresponds to the eleventh edition of the printed LCSH, which contains headings and references established by LC through August 1987.

The Library of Congress distributes LCSH-mr on magnetic tape in the USMARC Communications Format for Authority Data (Library of Congress 1987). The project staff restructured the LCSH-mr records into the OCLC Authority Record Internal Processing Format (ARIPF). Consequently, this paper provides statistics on the variable-length fields of LCSH-mr records. Systems staffs can determine the length of subject authority records in their systems by:

1. Studying how their systems restructure the directory and fixed fields.
2. Adding the results to the average length of subject heading records given in this paper and to the average length of other types of LCSH-mr records covered elsewhere (Markey and Vizine-Goetz 1988, 13-66).

The project team enhanced the version of LCSH-mr with reciprocals of *see also from* tracings (MARC tags 5xx). The reciprocals are *see also* references and have been added to LCSH-mr records in local 9xx fields. The 9xx range of tags is not authorized for use in the USMARC format for authority data and thus does not conflict with authorized fields.

TYPES OF RECORDS IN LCSH-MR

LCSH-mr contains 160,706 records, the majority of which are for subject headings. General explanatory *see* reference fields (MARC tag 260) occur 837 times in LCSH-mr records. Subject heading records containing general *see* references are not included in subsequent analyses of variable-length fields because the established heading fields (MARC tags 1xx) of these records do not actually bear an established heading. The number and percentage of records for the six different types of established headings (MARC tags 1xx) in LCSH-mr are given in table 1.

LCSH-MR DISTRIBUTED BY LC

Summary field statistics for tagged fields of 128,367 topical subject heading records (MARC tag 150) appear in table 2. The headings for each column in table 2 are defined here:

Tag: three-digit field tag
Total Occurrence: Total number of this field in records
(%) Records: Percent of records that contain this field
Occurrence/Record: Occurrence of this field divided by the total number of records
Length/Record: Given that this field occurs once in a record, the average length of the field, including subfield codes and delimiters

Table 1. Established Headings

Tag/Name of Established Heading	Total Occ.	(%) Records
100/Personal name	9,454	5.91
110/Corporate name	2,279	1.43
111/Conference or meeting	26	0.02
130/Uniform title	525*	0.33
150/Topical subject	128,367†	80.29
151/Geographical name	19,218‡	12.02
Total	159,869§	100.00

* Three records for unestablished headings are omitted, yielding 528 uniform title records.
† 831 records for unestablished headings are omitted, yielding 129,198 topical subject heading records.
‡ Three records for unestablished headings are omitted, yielding 19,221 geographical name records.
§ 837 records for unestablished headings are omitted, yielding 160,706 LCSH-mr records.

Length/Occurrence: Average length of this field each time it occurs, including subfield codes and delimiters
Characters/Record: Length of this field averaged over all records, including subfield codes and delimiters
Maximum Length: Maximum length of this field found in all records, including subfield codes and delimiters
S.D. Length/Occurrence: Standard deviation of length/occurrence

The following 20 MARC-tagged fields occur in subject heading records distributed by LC:

LC record control number (MARC tag 010)
Cataloging source (MARC tag 040)
Suggested LC classification number(s) (MARC tag 053)
Established heading for topical subjects (MARC tag 150)
General *see also* reference (MARC tag 360)
See from tracings (MARC tags 400, 410, 411, 430, 450, 451)
See also from tracings (MARC tags 500, 510, 530, 550, 551)
Information found and not found fields (MARC tags 670 and 675)
Scope note(s) (MARC tag 680)
Note/example under field(s) (MARC tag 681)

The project team's enhancement of LCSH-mr with *see also* references results in an additional six fields bearing local tags 900, 910, 911, 930, 950, 951.

LCSH-mr contains 128,367 records for topical subjects (MARC tag 150). Of these, 43.02% contain one or more suggested LC classification numbers (MARC tag 053); 2.41% bear one or more scope notes (MARC tag 680). These percentages are slightly higher than the 36.33% of all LCSH-mr records (i.e., MARC tags 1xx) with LC classification numbers and the 2.12% of all LCSH-mr records with scope notes (Markey and Vizine-Goetz 1988, 13-30).

See from tracings for topical subjects (MARC tag 450) are the most plentiful of the six different types of *see from* tracings in LCSH-mr records. *See also from* tracings for topical subjects (MARC tag 550) are also the most plentiful of the six different

*Table 2. Field Statistics for Topical Subject Heading Records**

Tag	Total Occ.	(%) Records	Occ./ Record	Length/ Record	Length/ Occ.	Char./ Record	Max. Length	S.D. Length/ Occ.
010	128,367	100.00	1.000	14.00	14.00	14.00	28	0.15
040	128,367	100.00	1.000	10.44	10.44	10.44	15	1.41
053	62,587	43.02	0.488	14.30	12.61	6.15	64	6.44
150	128,367	100.00	1.000	24.84	24.84	24.84	109	11.76
360	3,034	2.36	0.024	117.14	117.14	2.77	587	71.92
400	40	0.03	0.000	47.27	39.00	0.01	92	15.39
410	75	0.05	0.001	39.88	36.69	0.02	70	10.44
411	1	0.00	0.000	47.00	47.00	0.00	47	0.00
430	209	0.14	0.002	24.20	21.19	0.03	56	8.58
450	96,980	42.48	0.755	39.08	21.97	16.60	217	8.95
451	277	0.19	0.002	39.78	35.76	0.08	78	12.32
500	214	0.15	0.002	44.93	40.73	0.07	80	14.04
510	249	0.18	0.002	37.94	34.89	0.07	88	11.85
530	122	0.09	0.001	28.78	27.36	0.03	69	12.93
550	139,355	61.25	1.086	38.63	21.79	23.66	99	9.18
551	2,903	1.84	0.023	41.44	33.69	0.76	80	13.08
670	10,676	4.42	0.083	125.53	66.64	5.54	861	53.66
675	1,306	1.02	0.010	55.71	55.71	0.57	340	48.92
680	3,309	2.41	0.026	223.95	209.19	5.39	721	114.69
681	5,161	3.95	0.040	40.78	40.06	1.61	183	16.69
Locally Generated *See Also* References								
900	438	0.27	0.003	60.98	47.75	0.16	99	14.53
910	1,337	0.49	0.010	74.58	35.42	0.37	72	10.60
911	32	0.02	0.000	43.43	31.22	0.01	75	13.20
930	201	0.12	0.002	35.03	25.79	0.04	71	10.03
950	139,233	29.74	1.085	77.86	21.35	23.15	99	9.97
951	10,056	2.58	0.078	91.28	30.11	2.36	85	10.14

*Excludes 831 records bearing unestablished forms of the subject referred to in established heading fields (MARC tag 150).

types of *see also from* tracings in such records. The average number of variable-length fields in LCSH-mr records distributed by LC is 5.545 (the sum of column three, table 2, MARC tags 010-681). Two fields occurring in every subject authority record, i.e., LC record control number (MARC tag 010) and cataloging source (040), do not clarify the meaning of the established heading (MARC tag 1xx). On average, fewer than three of the following fields are given to the established heading (MARC tag 1xx): LC class numbers (MARC tag 053), general references (MARC tags 260 and 360), tracings (MARC tags 4xx and 5xx), and information fields (MARC tags 680 and 681). These fields contain terms and information that clarify the meaning of the subject in the established heading field.

The average number of characters per record in variable-length fields of LCSH-mr records distributed by LC is 112.64 (the sum of column 6, table 2, MARC tags 010-681). Scope notes (MARC tag 680), general references (MARC tags 260 and 360), and source data found fields occur infrequently in LCSH-mr records. When they do occur, they are the longest fields in LCSH-mr records. The maximum number of variable-length fields in records distributed by LC is 71 for the established subject heading *Noctuidae*. The longest subject authority record is the established heading *Examinations* with 3,506 characters.

LCSH-MR ENHANCED WITH *SEE ALSO* REFERENCES

When LCSH-mr records are obtained from LC, *see also from* tracings are contained in MARC tags 5xx. Reciprocals of *see also from* tracings are made in library catalogs to refer searchers from established subject headings to one or more other established headings; such reciprocals are called *see also* references. The project team added *see also* references to LCSH-mr records in local tags 9xx. When systems staffs add *see also* references to LCSH-mr records in their local online systems, they add only those references used in their catalogs.

Summary statistics for LCSH-mr records when *see also* references are added to records appear in table 2. Most *see also* references are for topical subjects (local tag 950).

The average number of fields per topical subject heading record is 6.723 (the sum of column three, table 2, MARC tags 010-951). The average number of characters contributed by variable-length fields per topical subject heading record is 138.76 characters. Fields for established topical subjects (24.84 characters), *see also from* tracings (23.66 characters), and *see also* references (23.15 characters)—MARC tags 150, 550, and local tag 950, respectively—contribute the most characters per topical subject heading record. The maximum number of variable-length fields in records enhanced with *see also* references is 432 for the established subject heading *Indians of North America*. This is also the longest subject authority record with 10,015 characters.

CROSS-REFERENCE STRUCTURE

Coding in control subfield $w of *see also from* tracing fields (MARC tags 5xx) indicates a related or broader term relationship between the tracing and established heading. When reciprocals of *see also from* tracings are generated, the coding indicates a related or narrower term relationship between the *see also* reference and the established heading.

The number, percentage, and occurrences per record of related and broader terms are listed in table 3.

The majority of all *see also from* tracings in topical subject headings are broader terms. When *see also from* tracings for subjects (MARC tag 550) and geographical names (MARC tag 551) occur in topical subject heading records, 88.04% and 98.45%, respectively, are broader terms. The greatest number of occurrences (0.956) is for broader subject terms (MARC tag 550). The number, percentage, and number of occurrences per record of related and narrower terms in topical subject heading records are listed in table 4.

For *see also* references, 88.03% are narrower subject terms, and 99.54% are narrower terms for geographical names. The greatest number of occurrences (0.955) is for narrower subject terms. Narrower terms are desirable in LCSH-mr records because catalog searchers often consult terms that are broader than their topics of interest (Bates 1977; Knapp 1944; Markey and Demeyer 1986, 157-58).

FIELD STATISTICS FOR UNSUBDIVIDED AND SUBDIVIDED SUBJECT HEADINGS

Established, topical subject heading records are distributed into three separate and mutually exclusive categories: (1) unsubdivided subject headings, (2) once-subdivided subject headings, and (3) subject headings that have been subdivided

Table 3. Related and Broader See Also From *Tracings in Topical Subject Heading Records*

Tag	Occ. of Broader Terms	(%) of Refs.	Occ./ Record	Occ. of Related Terms	(%) of Refs.	Occ./ Record	Total No. of References
500	152	71.03	.001	62	28.97	.001	214
510	234	93.98	.002	15	6.02	.000	249
530	118	96.72	.001	4	3.28	.000	122
550	122,692	88.04	.956	16,663	11.96	.130	139,355
551	2,858	98.45	.022	45	1.55	.000	2,903
Total	126,054	88.25	.982	16,789	11.75	.130	142,843

two or more times, i.e., multisubdivided headings.

The majority (70.48%) of topical subject heading records bear unsubdivided headings. Once-subdivided subject headings are 25.90% of subject heading records, and multisubdivided subject headings are 3.62% of such records.

Summary field statistics for tagged fields of the three categories of topical subject

*Table 4. Related and Narrower See Also References in Topical Subject Heading Records**

Tag	Occ. of Narrower Terms	(%) of Refs.	Occ./ Record	Occ. of Related Terms	(%) of Refs.	Occ./ Record	Total No. of References
900	376	85.84	.003	62	14.16	.000	438
910	1,322	98.88	.010	15	1.12	.000	1,337
911	32	100.00	.000	0	0.00	.000	32
930	197	98.01	.002	4	1.99	.000	201
950	122,571	88.03	.995	16,662	11.97	.130	139,233
951	10,010	99.54	.078	46	0.46	.000	10,056
Total	134,508	88.90	1.048	16,789	11.10	.000	151,297

*Totals differ between tables 3 and 4 because of 202 blind references in *see also from* tracing fields (MARC tags 5xx) of LCSH-mr.

Table 5. Field Statistics for Unsubdivided and Subdivided Topical Subject Headings

	Unsubdivided			Once-subdivided			Multisubdivided		
Tag	Total Occ.	(%) Records	Occ./ Record	Total Occ.	(%) Records	Occ./ Record	Total Occ.	(%) Records	Occ./ Record
010	90,471	100.00	1.000	33,246	100.00	1.000	4,650	100.00	1.000
040	90,471	100.00	1.000	33,246	100.00	1.000	4,650	100.00	1.000
053	54,456	52.82	0.602	7,584	20.77	0.228	547	11.44	0.118
150	90,471	100.00	1.000	33,246	100.00	1.000	4,650	100.00	1.000
360	2,389	2.64	0.026	629	1.89	0.019	16	0.34	0.003
400	39	0.04	0.000	1	0.00	0.000	0	0.00	0.000
410	34	0.04	0.000	40	0.11	0.001	1	0.02	0.000
411	1	0.00	0.000	0	0.00	0.000	0	0.00	0.000
430	195	0.19	0.002	13	0.04	0.000	1	0.02	0.000
450	83,190	50.59	0.920	12,937	24.47	0.389	853	13.57	0.183
451	144	0.14	0.002	123	0.35	0.004	10	0.22	0.002
500	210	0.21	0.002	4	0.01	0.000	0	0.00	0.000
510	225	0.23	0.002	22	0.06	0.001	2	0.04	0.000
530	113	0.12	0.001	8	0.02	0.000	1	0.02	0.000
550	126,130	75.91	1.394	12,375	27.77	0.372	850	15.33	0.183
551	2,548	2.23	0.028	270	0.80	0.008	85	1.72	0.018
670	9,251	4.96	0.102	1,122	2.71	0.034	303	6.04	0.065
675	1,225	1.35	0.014	72	0.22	0.002	9	0.19	0.002
680	2,882	2.95	0.032	408	1.22	0.012	19	0.41	0.004
681	2,626	2.84	0.029	2,307	6.84	0.069	228	4.88	0.049
Locally Generated *See Also* References									
900	337	0.29	0.004	75	0.17	0.002	26	0.47	0.006
910	222	0.18	0.002	1,097	1.38	0.033	18	0.39	0.004
911	15	0.01	0.000	12	0.04	0.000	5	0.09	0.001
930	122	0.11	0.001	72	0.13	0.002	7	0.15	0.002
950	120,614	32.58	1.333	17,196	23.38	0.517	1,423	20.02	0.306
951	456	0.29	0.005	9,341	8.84	0.281	259	2.49	0.056

heading records are given for records enhanced with *see also* references and appear in table 5. The headings for each column in table 5 are the same headings used in table 2.

Unsubdivided subject heading records have the highest percentage of LC classification numbers (MARC tag 053, 52.82%); only 20.77% and 11.44% of once-subdivided and multisubdivided subject heading records, respectively, have LC classification numbers. LC classification numbers generally occur in 43.02% of subject heading records (table 2). Unsubdivided subject heading records also feature the highest percentage (2.95%) of scope notes (MARC tag 680) of the three records categories. This percentage, however, is not much greater than the 2.41% of topical subject heading records generally with scope notes (table 2).

Unsubdivided subject heading records contain greater percentages and numbers of occurrences of cross-reference fields for topical subjects (i.e., MARC tags 450, 550, and local tag 950) than the other two record categories. Although cross-reference fields for geographical names (MARC tags 451, 551, and local tag 951) occur infrequently in topical subject heading records (MARC tag 150), once-subdivided subject heading records contain a greater percentage (8.84%) and number of occurrences per record (0.281) of *see also* references than the other two record categories. The once-subdivided headings in such records are usually topical subjects subdivided geographically (i.e., subfield $z), e.g., *Arches—Missouri* or *Bridges—California*, with *see also* references naming the arch or bridge, i.e., *Gateway Arch (St. Louis)* or *Golden Gate Bridge (San Francisco, Calif.)*, respectively. Generally, unsubdivided subject heading records have greater percentages and numbers of occurrences per record of all tagged fields than the other two record categories.

Summary statistics for topical subject heading records generally and for the three categories of such records are given in table 6.

LCSH-mr contains 128,367 records for topical subject headings (MARC tag 150); these records contain 26 types of variable-length fields and 107 types of subfields. The subject heading record with the greatest number of variable-length fields (432) and characters (10,115) is *Indians of North*

Table 6. Summary Statistics for Topical Subject Heading Records

Criteria	All	Unsubdivided	Once-Subdivided	Multi-Subdivided
Number of records	128,367.00	90,471.00	33,246.00	4,650.00
Number of different variable fields	26.00	26.00	25.00	23.00
Number of variable fields	862,896.00	678,837.00	165,446.00	18,613.00
Average number of variable fields per record	6.723	7.50	4.98	4.00
Average length of each variable field, including subfield codes and delimiters	20.64	20.11	22.44	25.69
Average length of variable fields in each record, including subfield codes and delimiters	138.76	150.83	110.75	102.76
Number of different subfields	107.00	96.00	84.00	64.00
Total word count	2,447,312.00	1,876,751.00	518,320.00	67,241.00
Average number of words per record	19.06	20.58	15.59	14.46
Average number of words per field	2.84	2.74	3.13	3.61
Average number of words per subfield	1.93	1.92	1.96	1.89
Average number of subfields per record	9.88	10.71	7.96	7.63
Average number of subfields per field	1.47	1.43	1.60	1.91
Average length of subfields in each record, excluding subfield codes and delimeters	118.80	129.30	94.59	87.46

America. (Of LCSH-mr records generally, the record with the greatest number of fields (432) is *Indians of North America* and with the most characters (14,145) is *Japan—Antiquities* (MARC tag 151)). Once-subdivided headings contain 26 types of variable-length fields and 96 types of subfields; multisubdivided headings contain fewer fields and subfields. Topical subject heading records average 6.723 variable-length fields and 138.76 characters per record. For unsubdivided headings, these numbers are greater; for subdivided headings, these numbers are generally less than topical subject heading records and unsubdivided heading records. Subdivided headings exceed unsubdivided headings in the average length of each variable field, average number of words per field, and the average number of subfields used per field. The main reason why these statistics are higher for subdivided than for unsubdivided headings is the increased number of characters, words, and subfields needed in the former records to represent subdivided subject headings.

DISCUSSION

The results of statistical analyses of topical subject headings are given in this paper to help system designers who are now planning to add LCSH-mr records to their systems. Generally, the statistical analysis increases our understanding of LCSH-mr as a file of subject authority records.

The statistical analysis presented in this paper marks the completion of the first phase of a three-phase project sponsored by the Council on Library Resources, OCLC Online Computer Library Center, and the University of Michigan to study the accessibility of LCSH-mr in online bibliographic systems. The second and third phases of the project focus on comparisons of subject headings in LCSH-mr with assigned subject headings in bibliographic records and with the terms patron users enter into online catalogs. Combined with first-phase analyses, the second and third phases of the project will provide additional data, analyses, and recommendations about forging LCSH-mr into a user's tool for subject access to online bibligraphic systems.

BIBLIOGRAPHY

Bates, Marcia. "System Meets User: Problems in Matching Subject Search Terms." *Information Processing and Management* 13:367–75 (1977).

Knapp, Patricia B. "The Subject Catalog in the College Library." *Library Quarterly* 14:108–18, 214–28 (1944).

Library of Congress. *USMARC Format for Authority Data Including Guidelines for Content Designation* (Washington, D.C.: LC, Network Development and MARC Standards Office, 1987).

Markey, Karen, and Anh Demeyer. "Dewey Decimal Classification Online Project: Evaluation of a Library Schedule and Index Integrated into the Subject Searching Capabilities of an Online Catalog." Research report no. OCLC/OPR/RR-86/1 (Dublin, Ohio: OCLC: 1986).

Markey, Karen, and Diane Vizine-Goetz. "Characteristics of Subject Authority Records in the Machine-Readable Library of Congress Subject Headings." Research report no. OCLC/OPR/RR-88/2 (Dublin, Ohio: OCLC: 1988). ■■

Survey of Artificial Intelligence and Expert Systems in Library and Information Science Literature

Cynthia C. Hsieh and Wendy Hall

A basic understanding of AI's changing nature is useful in recognizing the complex issues, questions, and concerns now facing librarians in the information industry. This survey examines the definition and history of artificial intelligence and investigates the body of literature on AI found in two major resources in the library and information science field from 1976 to 1987. The idea is

Cynthia C. Hsieh is Technical Service Librarian and Wendy Hall is Reference Librarian, Columbia College Library at Chicago.

to provide an inside look at this new technology and its impact on the way information is managed.

INTRODUCTION

We are at a new beginning of the information revolution. Computer and electronic technology is making a major breakthrough from "pure" numerical calculation to "intelligent" problem solving, known as artificial intelligence (AI).

Interests in artificial intelligence and in its branch, expert systems, have gained tremendous attention in the last decade and will continue to be the focus of future computer and electronic technology research. Today, AI-related studies have emerged from academic research laboratories into the "real world" marketplace and from the computer sciences to almost every discipline of human knowledge. This is evident by the growth of products based on AI research which have become an important segment of the market and by the appearance of various seminars, conferences, and workshops aimed at special audiences from different occupations and professions.

The history of library and information science experienced a pattern of rise and decline influenced greatly in response to changes in computer technology: the arrival of the OCLC system in the 1970s changed the nature and even the concept of cataloging in libraries and information settings, and the emergence of online databases changed the way we store and retrieve information. These are only two examples. As interest in AI research grows, librarians and information specialists—being key characters in the information profession—need a basic understanding of this new trend in computer technology. Hence, we may be able to apply new AI products which appear to enhance our services.

The purpose of this paper is to provide valuable information for practicing librarians, as well as for library and information science researchers, who are interested in the application of AI in library and information settings. This will help them recognize the potential importance of this new research endeavor as well as to suggest future research possibilities and directions by a systematic analysis of library and information science-related literature.

DEFINITIONS

Definitions of artificial intelligence are unclear and vague. The term "artificial intelligence" was first formally introduced in 1956 by John McCarthy and a few other scientists who organized the first AI conference.[1] However, its origin extends back into the 1930s and 1940s when the British mathematician, Alan Turing, theorized on the possibility of teaching computers to perform tasks based on logic. Turing's ideas are fundamental to AI. The following is a list of the most commonly recognized definitions:

> Artificial intelligence is the branch of Computer Science that deals with ways of representing knowledge using symbols rather than numbers and with rules-of-thumb, or heuristic, methods for processing information.[2]

> Artificial intelligence is the part of Computer Science concerned with designing intelligent computer systems, that is, systems that exhibit the characteristics we associate with intelligence in human behavior—understanding language, learning, solving problems, and so on.[3]

> The subfield of Computer Science concerned with developing intelligent computer programs. That includes programs that can solve problems, learn from experience, understand language, interpret visual scenes, and, in general, behave in a way that would be considered intelligent if observed in a human. A growing set of computer problem solving techniques that are being developed to imitate human thought or decision making processes or to produce the same results as those processes.[4]

> Artificial intelligence is the branch of Computer Science that deals with using computers to simulate human thinking. Artificial intelligence is concerned with building computer programs that can solve problems creatively, rather than simply working through the steps of a solution designed by the programmer.[5]

From the definitions above, one can conclude artificial intelligence is as follows:

1. It is a branch/subfield of computer sciences.
2. Its goal is to act "intelligently," like a human being.
3. It deals with problems which require techniques not only to match letters and

numbers or to do calculation but also to make decisions.

4. It should be able to combine new information, previously known facts, rules of thumb, and even intuition to come up with an appropriate response to a problem.

HISTORY

Patrick H. Winston, in his article "The AI Business: a Perspective," divided the history of AI development into six ages:[6]

1. Prehistoric age—The period started in 1842 and extended to about 1960. It was the age of conceptual breakthrough and speculation.

2. Dawn age—In the 1960s AI scientists tried to simulate the complicated process of thinking by finding general methods for solving broad classes of problems. The most famous example is the General Problem Solver (GPS) devised by Newell and his fellow researchers in 1960. However, despite some interesting progress, this approach proved to be too difficult and unrealistic.

3. Dark age—Since the mid-1960s there was not much happening because of overexpectation from the Dawn age. Until early 1970, AI had only limited success. Natural language translation had already collapsed. Well-constructed problems such as game playing proved tractable, but real complex problems proved to be beyond the techniques thus far devised or resulted in an explosive search that exceeded the then current capabilities.

4. Renaissance—During this period, researchers began to make systems that caught people's eye. Due to the lessons from the Dark age, AI scientists decided there must be another way to make a computer program intelligent. There was, therefore, a shift in emphasis toward the investigation of the role of knowledge in intelligent solving of real-world problems. It became apparent that shortcuts or rules-of-thumb were essential in order to narrow the problem space to a manageable size. This conceptual breakthrough led to the development of many impressive expert systems, among them MYCIN, a medical diagnosis and treatment consultant devised by Shortliffe (1976) at Stanford University.

5. Age of partnerships—According to Winston, the years between the mid-1970s and the early 1980s are the period when researchers in AI began to admit that there were other researchers, particularly linguists and psychologists, with whom people working in AI could form important liaisons. This was also when library and information professionals began to show interest and make contributions to AI and its subfields.

6. Age of the entrepreneur—This age emerged in the 1980s. The big news of this age is that AI went "commercial." Companies involved in AI application product development shifted from a very few companies founded by AI researchers in the earlier days to hundreds of companies today owned by large corporations. The existing AI technology is now becoming codified and therefore widely available to everyone in every discipline, not just the core group of several hundred researchers in the early 1970s.

RESEARCH METHODS

A method is described in which a scan of two major online reference sources in the library and information science field from 1976 to 1987 was conducted to retrieve articles in English relevant to AI and expert systems: *Library Literature* via Wilsonline and *Library and Information Science Abstracts (LISA)* via Dialog.

When exploring the possibilities of how to best access the AI literature related to the information profession from among numerous resources on the market, the authors' choice was narrowed down to *Library Literature* and *LISA* in response to the following merits: (1) both resources present literature targeted toward information professionals; (2) both indexes appear in print and online formats, which complement one another enabling greater access; (3) *Library Literature*, which indexes 185 leading library publications with emphasis on U.S. publications, is unique in focusing on the traditional aspects and applications of AI; and (4) *LISA*, the British counterpart of *Library Literature*, is unique in its technological focus on AI research and development.

This survey chose 1976 as the starting point, since the earliest article discovered in a preliminary online comprehensive search that discussed AI in relation to library science literature was published in 1976.[7] As the investigation into the literature began in early 1988, 1987 is a reasonable cut-off date for this research purpose.

The overall strategy involved in the online search was using AI and expert systems as subject headings/descriptors: in *Library Literature* using "(SH) ARTIFICIAL INTELLIGENCE OR EXPERT SYSTEMS" and in *LISA* using "ARTIFICIAL() INTELLIGENCE/DE OR EXPERT() SYSTEMS/DE." The results were then limited to English language publications only.

The search in *Library Literature* yielded a total of 120 postings, and *LISA* retrieved a total of 65 postings, for a gross of 185 references.

A manual search was appropriate in *Library Literature* from 1981 to 1984, as the online version started coverage later in October 1984. The subject term ARTIFICIAL INTELLIGENCE officially appears in *Library Literature* in 1981 and EXPERT SYSTEMS shortly thereafter. Prior to 1981 articles related to AI were scattered under various subject headings, such as COMPUTERS, INFORMATION SCIENCE, INFORMATION THEORY, INFORMATION STORAGE AND RETRIEVAL, etc. The degree of difficulty and the investment of time involved made it unfeasible for the authors to undertake a thorough manual search for the period prior to 1981. However, by comparing combined results from both the manual and online searches of these two resources, the authors were able to sift through the results by manually processing them to see if any pattern or relationships developed and thereby excluded a total of twenty-eight duplicates.

RESULTS

Despite the difficulties encountered during the research process, the authors would like to state that, as is typical with most computer searches, there are no guarantees to retrieving "every" relevant reference available on a topic. And the results retrieved in this survey are no exception.

Since the onset of artificial intelligence, the literature on AI has sharply increased, and, as anticipated, the search for AI-related literature in this survey also reflects this trend. A gradual increase can be seen from the year 1983 onward, as shown in figure 1. The exception to this is the year 1979, where an impressive showing of conference papers increased the total for that year.

Figure 2 identifies in descending order those journals that contributed the most articles related to the subject of AI in library and information science literature.

The two most reputable library and information science journals, *American Libraries* and *Library Journal* had very low retrieval totals, with the former showing four and the later only one article.

It is also interesting to note that both *the Special Libraries and Information Bureaux* (the top one on the list) and the *Proceedings of the American Society for Information Science* (third highest on the list) contributed a total of fifty-one articles. And *NTIS Reports*, a technical report-type of publication, placed fourth on figure 2.

In contrast to conference papers and technical reports, only two books were retrieved from the total number of resources located: one is a revision of a 1978 Ph.D. thesis by Carole D. Hafner from the University of Michigan;[8] the other was published in 1979 by A. M. Mulvehill.[9] This may be because AI and expert systems are relatively new in library and information science; therefore, there are more conference papers and technical reports than books. (Monograph-type of publications often indicate the maturity of a knowledge or technology.) However, the authors also do not exclude the fact that *LISA's* current selection policies emphasize journal articles.

Figure 3 provides a percentage of these various publications broken down by library application. Out of a grand total of 179, only 52 publications were specifically targeted toward applications in the library.

Twelve articles addressed AI applications in the technical services area, especially for the operation of cataloging. This may be due to the ease with which AACR2 rules can be easily adapted to manipulation. This is not the case, however, when approaching the areas of acquisitions or indexing and abstracting, where there are no set rules to guide the creation of expert sys-

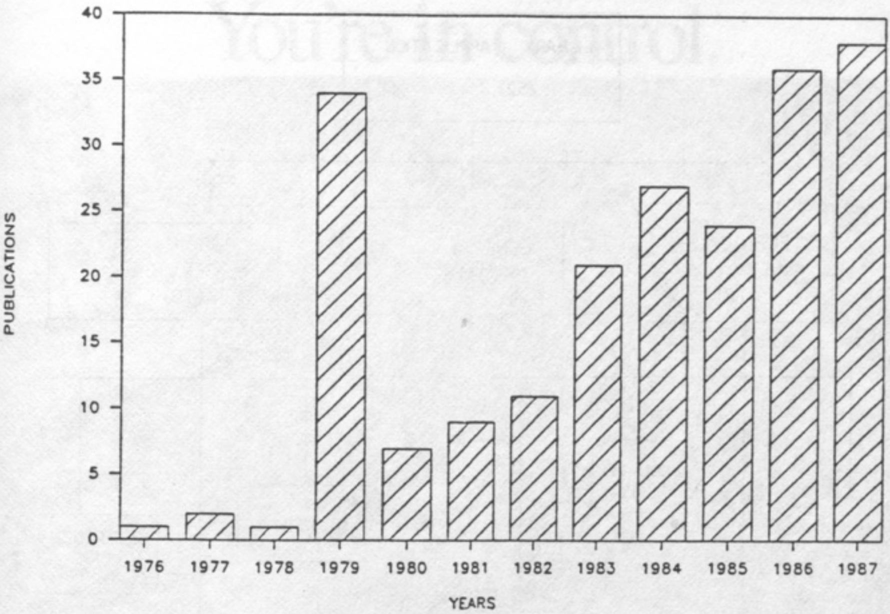

Fig. 1. Publications by Year.

JOURNAL TITLES BY NUMBER OF PUBLICATIONS

JOURNAL TITLE	NUMBER
Proceedings of the Special Libraries and Information Bureaux	38
Information Processing and Management	18
Proceedings of the American Society for Information Science	13
NTIS Report	11
Journal of the American Society for Information Science	10
Bulletin of the American Society for Information Science	6
Library Software Review	5
Online	5
Canadian Journal of Information Science	4
Information Technology: Research and Development	4
Journal of Information Science	4
Library Journal	4
Online Review	4

Fig. 2. Journal Titles by Number of Publications.

tems. Twelve articles were directed toward AI applications in computerized online searching, which lends itself well to this new technology due to their similar natures.

However, from the results culled in this survey, artificial intelligence and expert systems still have made only a minor impact as yet in the library. It takes a great deal of money, time, and knowledge—all of which are precious commodities—to incorporate AI in the library. This may change with the "age of the entrepreneur" heralding an historic new era in the evolution of artificial intelligence by speeding the availability of AI application products and services. The moderate integration of AI into the business world and eventually

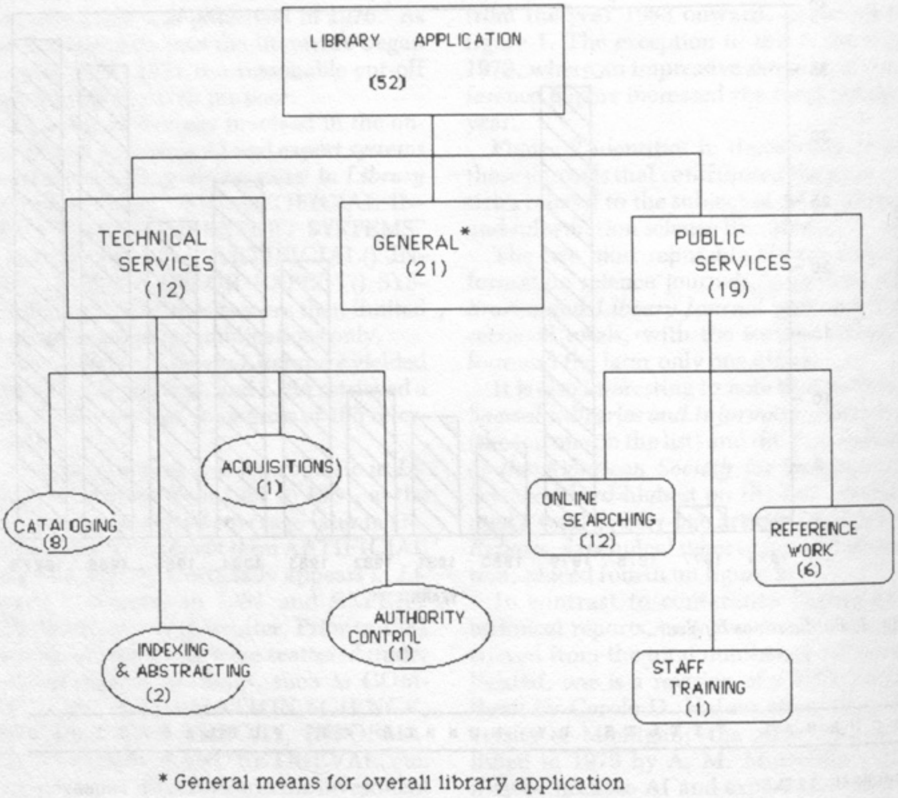

Fig. 3. Number of Articles by Types of Library Application.

into conventional computing could make remarkable new library applications possible in the near future.

REFERENCES AND NOTES

1. Donald T. Hawkins, "Artificial Intelligence (AI) and Expert Systems for Information Professionals—Basic AI Terminology," *Online*, p.92, (Sept. 1987).
2. Henry C. Mishkoff, *Understanding Artificial Intelligence* (Indianapolis, Ind.: Howard W. Sams, 1985), p.11.
3. Avron Barr and Edward A. Feigenbaum, *The Handbook of Artificial Intelligence* (Los Altos, Calif.: William Kaupman, 1981–82), v.1, p.3.
4. V. Daniel Hunt, *Artificial Intelligence and Expert System's Sourcebook* (New York, N.Y.: Chapman and Hall, 1986), p.152.
5. Douglas Downing, *Dictionary of Computer Terms* (Woodbury, N.Y.: Barron's, 1986), p.14.
6. Patrick H. Winston, "The AI Business: A Perspective," *Manufacturing Engineering* (Mar. 1985), p.75–78.
7. The earliest article that the authors found is: Linda C. Smith, "Artificial Intelligence in Information Retrieval Systems," *Information Processing and Management* 12, 3:189–222 (1976).
8. Carole D. Hafner, *An Information Retrieval System Based on Computer Model of Legal Knowledge* (1978).
9. A. M. Mulvehill, *Study of the Data Organizational Techniques Underlying Pattern Discovery in Living and Artificial Pattern Recognition Systems* (1979). ■■

News and Announcements

OCLC and AMIGOS Announce Joint Venture

AMIGOS Bibliographic Council, Inc., and OCLC Online Computer Library Center have announced plans to enter into a joint venture to market a collection analysis service to libraries.

The OCLC/AMIGOS service brings together AMIGOS's batch-oriented Collection Analysis Service (available since 1984) and a compact disc-based collection analysis system now under development at OCLC. AMIGOS will undertake nationwide marketing and sales on behalf of the coordinated product line, which will be offered to individual libraries and resource-sharing groups.

The compact-disc collection analysis option will allow a library to compare its collection development activity against national practice, as reflected in a standardized subset of bibliographic and holdings data derived from the OCLC Online Union Catalog. An interactive microcomputer format will permit users to conduct hands-on analyses at the local level.

A batch option provides customized statistical reports that categorize the individual or overlapping holdings of libraries, within client-defined subject categories. The holdings data are compiled by comparing magnetic tapes of library bibliographic records against a machine-readable classification table.

Information about product features and pricing of the compact-disc option will be announced in coming months. ■■

EBSCO and Data Trek Announce Interface

EBSCO and Data Trek, Inc., have announced the completion of an interface between the two systems. Over a period of time, Data Trek has developed the following capabilities: database conversion, online claiming, online ordering, and a feature that compares invoice data on diskette to the records in the database.

Database conversion is achieved by extracting data from the Invoice on Diskette provided by EBSCO, free to qualifying customers. Fields such as Title of Subscription, Start Date, Term of Subscription, Subscriber Address, EBSCO Title Number, and Frequency are extracted from the diskette in machine-readable form to create similar fields of information in the Data Trek database. This eliminates time-consuming keying of data by library staff to create the serials database. Claims and orders are created on the Data Trek system and uploaded electronically through Tymnet directly into the EBSCO computer. This speeds the processes by preventing delays caused by postal delivery as well as eliminating unnecessary rekeying of information. The invoice program compares data from the diskette with records in the customer's database and issues a discrepancy report when the title, price or other data differs. Further development will provide for downloading of invoice data.

To learn more about this interface, please contact Data Trek, Inc., 621 Second St., Encinitas, CA 92024; (800) 876-5484. In California, call (619) 436-5055; FAX (619) 436-0514. At EBSCO, contact Sharon Cline, Director of Library Services, (205) 991-1368; FAX (205) 991-1479. ■■

Personal Bibliographic Software, Inc., Announces BRS Biblio-Link Upgrade

Personal Bibliographic Software, Inc. (PBS), has announced the release of a new version of the Biblio-Link to BRS (BRS Information Technologies' BRS/Search Service) for IBM personal computers and compatibles.

BRS/Search Service provides users access to over 200 computerized databases. Information is available on numerous subjects, including business, education, life sciences, medicine, social sciences, and the humanities.

Biblio-Link is a program which transfers references downloaded from online database systems directly into Pro-Cite, an advanced bibliographic database- and text-management program offered by PBS. Biblio-Link automatically puts each reference into the correct Pro-Cite workform type (such as journal or book) and transfers each field of information into the correct Pro-Cite field. After the transfer to a Pro-Cite database, records can be sorted, searched, edited, indexed, and formatted according to any bibliographic style. Biblio-Links and Pro-Cite are tools for any professional who must manage reference information.

The new version of the BRS/Biblio-Link works with standard BRS records as well as BRS/Colleague records and BRS/After Dark records. The program has the ability to parse the Source (SO) field, which is necessary when BRS databases have multiple types of journal information in this field. The Source field may contain information such as the journal title, volume, issue, and date. With parsing turned on, this information is separated, and each bit is transferred into the appropriate Pro-Cite field. The new BRS/Biblio-Link supports lower casing for each database or lower casing in any particular field. Biblio-Link will capitalize the first letter of every word in the author, index, and affiliate field and the first letter of all other fields. The new version also includes a completely rewritten manual.

Other PBS products for IBM personal computers and compatibles include: Pro-Search, a specialized front-end program for easy searching of BRS and Dialog; Pro-Cite, a database management program for managing reference information; and Biblio-Links to Dialog, MEDLARS, US-MARC, STN, OCLC, and RLIN. Pro-Cite and Biblio-Links to BRS, Dialog, and MEDLARS are also available for the Macintosh.

Biblio-Links are available from Personal Bibliographic Software, Inc., for $195 each. Upgrades are $45; Pro-Cite programs are $395; and Pro-Search is $495. Free customer support and a 30-day return policy are available for all PBS products. ■■

GPO and UCLA Join in OCLC's LSP Authorities Implementation

The United States Government Printing Office (GPO) and UCLA began contributing new and changed authority records using OCLC's Linked Systems Project (LSP) Authorities Implementation on February 6, 1989.

These OCLC members join seven other institutions participating in the OCLC/National Coordinated Cataloging Operations (NACO) to contribute authority records to the Library of Congress via the Linked Systems Project link.

In 1977 GPO was the first participant to join the NACO project and annually contributes approximately 4,000 new authority records.

The UCLA Film and Television Archive and the Physical Sciences and Technology libraries are also contributing authorities records via the LSP link.

There are currently nine institutions that are active as OCLC NACO/LSP participants and contribute new and changed authority records to the Library of Congress via OCLC's LSP Authorities Implementation. They are:

- Eastman School of Music (participating as part of the NACO Music Project);
- Indiana University;
- Indiana University Music Library (participating as part of the NACO Music Project);
- Library of Congress Serial Record Division;
- Minnesota Historical Society;
- OCLC;
- United States Government Printing Office;
- University of California at Los Angeles; and
- University of Illinois at Urbana–Champaign.

From the beginning of contribution via OCLC in March through December 1988, these institutions have added 8,151 new and changed authority records to the national authority file.

The Linked Systems Project involves

computer-to-computer links between the Library of Congress, the Research Library Information Network (RLIN), and OCLC. The National Coordinated Cataloging Operations program (formerly the Name Authority Cooperative project) is in the initial phase of the applications component of LSP. ■■

IMPACT Inter-Library Loan Module Introduced by Auto-Graphics

The Inter-Library Loan (ILL) module of IMPACT, a CD-ROM public access catalog system, has been introduced by Auto-Graphics, Inc. The ILL module supports a variety of functions from request generation and electronic transmission to recordkeeping through the loan request's completion.

Two options are available for the ILL module. The first offers automatic generation of the ILL form directly from a search of the CD catalog or from a blank request form. The expanded functions of the second option provide for requests to be stored on disk for transmission by electronic mail. In one automatic sequence, the software connects to the communication network, transmits newly batched requests, and records incoming requests and reports to disk for local processing. Request maintenance operations include internal recordkeeping and reports between libraries on loans in process, reviewing the history of a request, and automatic assignment of a lending library for an individual transaction. A referral form is also provided for transmitting reference questions among libraries.

With some 1,800 workstations throughout North America, IMPACT is a leading system for CD-ROM public-access catalogs. Noted for its patron-oriented search screens and its many features for customizing displays, news and information files, and location-based searching, IMPACT is widely used not only for individual libraries but also multi-branch libraries, state and regional consortia, and multitype cooperatives. In addition to the Patron and Expert Level search modules and interlibrary loan, other components of IMPACT include modules for system administration, for cataloging, and, forthcoming in spring 1989, for Research Level searching.

For more information on IMPACT, contact Auto-Graphics, Inc., 3201 Temple Ave., Pomona, CA 91768; (714) 595-7204; (800) 325-7961 (U.S.); and (800) 828-9585 (California). ■■

University Libraries of Mississippi to Install CLSI System

CLSI has announced the signing of a $1.9-million contract with State Central Data Processing Authority for the University Libraries of the State of Mississippi—one of CLSI's largest academic contracts to date.

Two significant components of the contract include CLSI's Custom Item Conversion Program, which provides barcode labels in shelf-list order to speed up the task of labeling the libraries' combined three-million volume collection, and the implementation of the Open Systems Interconnection (OSI) Reference Model. The OSI Reference Model is a networking interface developed by the International Standards Organization, and allows disparate computer systems to communicate. The contract calls for implementation of the model within two years.

The libraries will be connected through DataLink, CLSI's communications module, allowing them to share information and resources. In addition, DataLink provides access to the bibliographic databases of other libraries, as well as a variety of online databases, including ALANET, BRS, and WILSONLINE. The State of Mississippi Universities system is comprised of eight schools, including the University of Mississippi, known as "Ole Mis." The University Library at "Ole Mis" is the repository of several noteworthy collections, including a "blues" archive with recordings, manuscripts, oral histories, monographs and folklore, and the William Faulkner Special Collection, which includes first editions and manuscripts of the author's work. ■■

WLN to Offer MARC Record Conversion Service

The Western Library Network (WLN) has announced plans for a new MARC Record Conversion Service that libraries can utilize to upgrade existing machine-readable records or to convert their shelf-lists to machine-readable form.

The new service accepts records from a variety of sources: tapes or floppy disks from local systems in MARC or non-MARC format, OCLC archival tapes, or hard-copy, such as shelfist cards, catalog cards, worksheets, or physical items. Records may come from one or more institutions and may be in mixed formats. The service also features the ability to extract individual libraries' records from multiple institution sources.

The MARC Record Conversion Service will provide these features to meet today's library automation needs:

- Checking and correction of tags, indicators, and subfields.
- Deduplication and merging of records from multiple institutions and multiple sources (tapes, floppies, etc.) according to library-defined criteria. The library can also determine the precedence for retaining and merging local information.
- Upgrading of bibliographic records to current LC standards.
- Conversion of headings to current LC subject or name authority headings using the NAF/SAF. Optionally, the library can use WLN's or its own authority file. Conversion includes changing abbreviations and initialisms to full forms, changing or deleting obsolete headings and subdivisions, and flipping headings or parts of headings from invalid to valid forms.

WLN also offers manual review and problem resolution for exception records. The components of the new service are available individually or as a full-service package. Libraries can receive processed records back in one or more formats: CD-ROM using WLN's well-known *LaserCat* software, magnetic tape in USMARC communications format, or microforms. Ziegman said the new service will begin operation in early 1990 and will be competitively priced. ■■

Sun Microsystems and OCLC to Exchange Technology

Sun Microsystems and OCLC are negotiating a technology-exchange agreement to explore the feasibility of using Sun hardware as a platform for the prototype of a local campus information system.

Under the agreement Sun will provide grants and discounts to OCLC for approximately $300,000 in computer hardware, which the OCLC Office of Research will use to conduct experiments and projects. The OCLC Office of Research will provide Sun with results of its research and investigations.

Sun Microsystems will install a Sun 4/280S data center server at OCLC. This hardware configuration is based on Sun's RISC (Reduced Instructions Set Computer) architecture, called SPARC. SPARC stands for Scalable Processor ARChitecture, emphasizing performance, efficiency, simplicity, and scalibility. Scalibility means that SPARC can be implemented in newer and faster semiconductor technologies. SPARC systems have an open-computer architecture. The design specification is published, and other vendors are producing microprocessors implementing the design. OCLC research scientists will develop prototype systems on it.

Carnegie Mellon University and OCLC are working together to develop Mercury, a prototype electronic research library, and applications developed under the Sun technology transfer agreement will be tested in the Mercury environment. ■■

UnCover Passes 5,000 Titles

CARL's article access service now covers more than 5,000 journals. Plans for UnCover were first announced in May 1988 and production began in September. Today, UnCover provides online access to the articles contained in nearly 5,300 journals, with a database of over 180,000 article records.

Growth in coverage continues at a rapid pace. At current production levels, UnCover is projected to reach its goal of covering 10,000 unique titles by summer of 1989.

Titles included in the database correspond to the collections of CARL Member Libraries: Auraria Library (serving the University of Colorado at Denver, Metropolitan State College, and the Community College of Denver), the Colorado School of Mines, the Denver Public Library, the University of Colorado at Boulder, the University of Denver, and the University of Northern Colorado. Records are created from actual issues received and are current at the time of issue check-in and processing.

Article records are based on the tables of contents of these issues, including authors, titles, and descriptive sentences or abstracts wherever available. Records can be retrieved individually using CARL's NAME searching for authors and WORD searching for topics as listed in titles or abstracts.

Specific journal titles can be retrieved using CARL's TITLE search (a part of the BROWSE capability prompted on the opening menu of UnCover). Once a user has located a desired journal title, the system will display a detailed listing of all issues checked in on the system along with the next expected issue. Users can then select a particular issue and ask to see its contents "recreated" on screen. More detail on individual articles is then available with a single keystroke.

UnCover's TITLE search capability produces a Union List of Serials on screen. And users have been quick to note the unexpected and added capability to use this "Union List" to find articles they need from anywhere in the combined CARL collections. In fact, it is this tie to actual journal title and issue holdings which will become most important for UnCover's next phase—full document delivery.

Plans call for users to be able to request delivery of any document discovered in the UnCover database from their terminals by the end of 1989. Delivery options will include ASCII full text, telefacsimile, and hardcopy, depending on the document requested and user equipment capabilities. Copyright permissions will be tracked and observed through careful system design, and both users and publishers will be provided detailed information in this area.

UnCover has been publicly released to CARL primary users since December 1988, and reaction has been enthusiastic.

CARL invites library professionals and other interested individuals to experience UnCover and the full CARL System directly without charge. Simply dial The CARL System at (303) 830-0106 (terminals should be set to 8 bit, no parity, 1 stop bit, full duplex; 1200 or 2400 baud modems are supported; CARL will automatically adjust to your baud rate; press any key several times to allow us to determine your modem speed). CARL is also accessible through the Internet at NIKE.CAIR.DU.EDU or 130.253,1.14; enter "carl" (small letters) at the login prompt.

UnCover is password protected for dial users. Those who are affiliated with a CARL member library or who have subscribed to a year's personal or institutional password will be prompted to enter a user ID. Those accessing UnCover for the first time can follow system prompts to access a demonstration database and learn more about the system.

The Colorado Alliance of Research Libraries was formed in 1978 to coordinate and extend the research resources provided by its members. CARL introduced its online system in 1981 and currently supports over 600 terminals and 3 million bibliographic records. Over 35,000 people use the system each day.

The CARL System is also installed at Arizona State University, Boulder Public Library, MARMOT (serving Colorado's western slope), Pikes Peak Library District in Colorado Springs, and Montgomery County Library System in Rockville, Maryland. Sites are interconnected for ready online access to all databases.

CARL Systems, Inc., was formed in 1988 to market and support CARL Systems and Services. For more information about CARL or CARL Systems, contact Rebecca T. Lenzini, CARL Systems, Inc., 777 Grant St., Suite 306, Denver, CO 80203; (303) 961-5319. ■■

Recent Publications

Software Reviews

Bib-Base/Acq. Bib-Base/Core + /Acq + /MARC modules, version 4.1, c1989. $1,485 as tested. Library Technologies, Inc., 1142E Bradfield Rd., Abington, PA 19001. Hardware requirements: MS-DOS compatible computer; hard disk, 10 MB min.

INTRODUCTION

Bib-Base/Acq is the acquisitions "arm" of the Bib-Base family of library software developed by Library Technologies (formerly Small Library Computing). Besides the /Acq and /MARC modules reviewed here, there is /Subject (for online subject retrieval), /CDROM (for accessing *Bibliofile* discs), /Multi-user (to allow access through a LAN by several work stations), /Public (a search-only, patron-oriented interface to Bib-Base), and /Cat (to make catalog cards and labels from Bib-Base records). The /BISAC module, to allow sending orders electronically to vendors, is in development.

This review focuses on the /Acq and /MARC software, which our library has been running for a year. Bib-Base is a powerful database because it uses the MARC record format for the bibliographic portion of the Bib-Base acquisitions record. MARC tag numbers are used for all bib fields. (The Bib-Base manual gives ample help for those unfamiliar with 100 being author, 245 being the title, and so on.)

Great advantages come from using the MARC format for acquisitions. The bib fields—indeed, the whole record—can be any length. Bib-Base can make acquisitions records from ANY source of MARC records. All Bib-Base records, even records which have been typed in from the computer keyboard, can be output as MARC-formatted records. Bib-Base will supply default values for required MARC fields.

Bib-Base/Acq can do many of the function one expects from any acquisitions system. Some it does well, some it does poorly. Some it cannot do at all. The distinctive features of Bib-Base are (1) *its generic functions* and (2) *its reliance on the Status field and free-text notes in the Bib-Base acquisitions record to deal with the changes in the acquisitions cycle*. These virtues give it great flexibility to be customized for a specific library's unique work flow.

GENERIC FUNCTIONS

A look at the Bib-Base main menu reveals some of the generic functions. There is a Search function, a Scan function, and a List function, among others. All of these functions operate on the file of acquisitions records. The Search function will isolate acq records that satisfy the *particular* bibliographic record field searched (name, title, ISBN, classification number, etc.) or *particular* system number (Bib-Base Record ID number, OCLC/RLN record number, invoice number).

The Scan function will isolate acq records matching the desired *combination* of special acquisitions fields (Status(es), data or range of dates, Fund(s), Supplier(s), Type(s)). Then those records can be viewed, deleted, or have their special acquisitions fields globally changed to other values. The List function uses an enhanced Scan function to select, sort, and then print (or write to a disk file) the matching acq records.

Through the Search function, the user can create an order, check in a received item, make note of a supplier's report on the status of an outstanding order, etc.,—anything to do with individual acquisition records.

Through the Scan function, the user can review records for claiming, delete a group of records that are no longer active, activate a group of special-priority records for ordering, globally change the acquisitions fields on an identifiable group of records, and so on. Using the List function, the user can create a bibliography, sorted by subject heading, for example, print the list of titles to be claimed with a single supplier, or print a list of out-of-print titles for an OP dealer to quote on.

RECORD FORMAT

The acquisitions record in Bib-Base is a simple and flexible, basically the bibliographic record—complete with MARC tags—enhanced by fourteen hardworking, fixed-length, Bib-Base acquisitions fields, plus any notes of unlimited length that one cares to add. A single acquisitions record will carry most transactions through the whole cycle—from "request" to "cataloged," with changes in the record's Status field, price fields, date fields, and notes heralding the stops along the trip.

Bib-Base requires the acquisitions staff setting up the system to know their own work flows, decide on their own uses for Bib-Base, configure the fixed fields, set up the input/output specifications, and consider what kinds of free-text notes will be needed.

None of this is unusual for setting up an automated system, although the reliance on notes is sensible. Those notes are no more than what one would write on a printed order residing in a manual on-order file. They indicate the order history of that title even when the order Status (e.g., "ORD'D") doesn't change.

ORDERING

Sources of Bibliographic Records

Bib-Base offers lots of options for obtaining bib records besides typing them in on the keyboard. As a system that developed independent of any vendor or bibliographic utility, Bib-Base *has* to offer those options to survive.

Many libraries these days are doing preorder search and verification over phone lines from bibliographic utilities (OCLC, RLIN, etc.) or vendors' in-house databases, or from CD-ROM copies of bibliographic databases like *BIP Plus* installed in their players. The advantage to using records from these sources in an acquisitions system is that the bibliographic record does not have to be rekeyed, presumably saving some time and definitely decreasing the errors associated with retyping.

Bibliographic records can be loaded into Bib-Base from a disk file of OCLC Savescreen records, RLIN Pass records, *BIP Plus* records, or "standard MARC" records. An added Bib-Base module, called Bib-Base/MARC, is needed to do this. This module is seamlessly integrated into the Bib-Base menus and functional options, which is as it should be.

All of the source options above require the records to be searched on the other database, written to disk file, and then loaded into Bib-Base. Library Technologies has developed, and now offers, software to search the Library Corporation's *Bibliofile* CD-ROM discs *from within Bib-Base*! The Bib-Base module that performs this feat is called Bib-Base/CDROM.

When adding bib records from one of these databases, staff can choose to strip off any unwanted bibliographic fields from the bib record (except for the title). In the same step, the system offers the ability to have common or default values entered into the acquisitions fields on record input. Once the parameters are chosen for loading and stripping the incoming bib records and adding values to the acquisitions fields, those parameters are saved for future use.

Similarly, there are five possible templates that can be defined and saved for typing in a new record from the keyboard. My library uses one for firm orders for which there is not an OCLC record and one for certain gift receipts. The other three are awaiting future needs.

Performance: The OCLC Savescreen and "standard MARC" record loading options have worked extremely well for my library. Student assistants do all our preorder searching on OCLC, downloading each Savescreen record onto a floppy disk at the same time they print out the record from OCLC for review by regular searching staff. The loading time into Bib-Base is relatively slow (ca. 8 min. for 25 Savescreen records) on our IBM XT clone, but we do not strip much off the WHOLE OCLC record. The processor speed (8 MHz), hard disk access speed, the length of the record to be translated into MARC format, and the number of access points that need to be indexed all contribute to this time. Batches of records are loaded at a low-use time of the day.

```
═══════════════════════UPDATE FUNCTIONS═══════════════════════
  F1 Search       F2 New Rec    AF2 Pattern     F3 Edit Bib     F4 Edit Fixed
  F9 Print Rec    CF2 Dup Rec   CF7 Delete      CF4 Edit MARC   F8 Ck-In Item
  AF3 Brieflist   F10 Exit

RecID 104         Status INV'D    SDate 06-DEC-88     Format BK
OPric 19.00       Supplr COUTT    ODate 31-MAY-88     Prepd? NO       #Cop 1
RPric 17.45       Fund   C-5      RDate 06-DEC-88     ActYr  CURR     Type
>010 __  85-30589
>020 __  0934802270 : |c $19.95
>040 __  DLC |c DLC
>050 0   QH26 |b .V53 1986
>082 0   508.32/2 |a B |2 19
>100 10  Vickery, Jim dale. |w cn
>245 10  Wilderness visionaries / |c Jim dale Vickery.
>260 0   Merrillville, Ind. : |b ICS Books, |c c1986.
>300 __  xv, 263 p. : |b ports. ; |c 24 cm.
>504 __  Bibliography: p. 241-247.
>500 __  Includes index.
>650  0  Naturalists |z United States |x Biography.
>650  0  Conservationists |z United States |x Biography.
>960     CLM'D 31-NOV-88.  BM

rec 4 of 4  type in rec #    - for prev rec
```

Fig. 1. Acquisitions Record, with Bib Record Imported from a Database of MARC Records [NOT SHOWN: Labels Are in Reverse Video].

INDEXING, SEARCHING DUPLICATE RECORD DETENTION

Each acquisitions record is indexed as it is entered into the Bib-Base database, whether it was typed into a new-record template or loaded from a disk file.

The first twenty characters of authors and titles are indexed. The first ten characters of the call number (LC, Dewey, other), ISBN/ISSN, order number, accession number, and invoice number are indexed. For any of the above search keys, entering a shorter string of characters will retrieve all records that begin with those characters. The LCCN, MARC record number (i.e., OCLC or RLIN number), and Bib-Base record ID number are fully indexed, and the full string of characters must be entered.

The author search key includes all name entries, whether main or added name entries, excluding names which are subjects. The title search key includes all title entries, excluding untraced series titles (entries in a 490 field with a 0 first indicator) and titles that are subjects. To discard initial articles, Bib-Base uses either the nonfiling indicator of the MARC field (if present) or its own list of common articles (English, German, Spanish, French).

Subject searching online is available by adding a module called Bib-Base/Subject, which indexes up to fifty-five characters of each subject entry. Our library does not have this module, so I cannot report on its success. See the *Reports* section of this review for off-line subject lists.

Bib-Base does *not* offer AUTOMATIC duplicate record detection when a new record is added to the database. This is not a major fault, since staff will always be searching the Bib-Base database before typing in a new record or before making a new record loaded from a disk file into an order.

Performance: When working with an individual record, the time the system takes to index the record as it is saved is barely noticeable. Entering a search is simple and fast. Retrieving searches is essentially instantaneous, even with over 12,000 acquisitions records in the system. The twenty-character indexing of titles and authors is quite adequate for arranging the entries alphabetically on the screen, putting the same or similar titles next to each other. Being a partial depository, our library does not order many U.S. government publications on Bib-Base, so the combination of both the same author and the same beginning words of the title does not occur frequently.

ORDER CREATION, PRINTING, ELECTRONIC TRANSFER

Creating an order is a short process. Search the database, locate a suitable acquisitions record or type one into a new record form, fill in the acquisitions fields (typically Status, order price, Supplier, number of copies if different from one, and Fund), add notes in the bibliographic part of the record, and save the record. Duplicating a previous order is a two-key operation (Ctrl key + function key) once the record is on the screen.

Notes fields are repeatable and virtually unlimited in length. As is true of the bibliographic fields, notes fields can be configured to be printing or nonprinting. Supplier notes can be printed on the order, and library notes on a routing slip. On the order, all printed acquisitions notes are printed first, ahead of the author or title. There are specialized notes fields—searchable—for Requestor, Order number, and Invoice number.

The acquisitions fields do not include a place to indicate how many volumes each title represents, and Bib-Base does not read the 300 field (extent/physical description) of the bib record to obtain it. Alerting a supplier to the extent of a multivolume order requires adding a note to the supplier.

Printing the order—either to a blank form or in purchase-order format to a sheet of paper—can be done with the record on the screen OR can be batched with other orders, sorted by supplier, and then sorted by main entry or title before printing. Batch printing "backup orders" for filing in a manual on-order file, if desired, can be done by choosing the option in the batch order printing routine. The orders can also be written to an ASCII disk file.

Sending orders electronically to vendors implementing the BISAC order transmission format is soon to be available for Bib-Base. Library Technologies indicates that this capability will be offered as an extra-cost module, Bib-Base/BISAC, because of the programming required to implement it for each vendor. The first release will include Blackwell/North America and several other major vendors.

Performance: Creating orders is simple and fast, especially if there is already a bib record available. If there is not an existing record, the time depends on how much information one wants to have typed into the template for each new record. Bib-Base checks that valid values are typed into the acquisitions fields and checks that the bib field *tags* are valid, but the content of the bibliographic fields is dependent on the source. Using notes to alert suppliers to multivolume or multipart orders is not a major inconvenience, but unformatted notes cannot be counted for statistical reports.

CLAIMS, CANCELLATIONS, ETC.

Bib-Base recognizes only one acquisitions Status between "ORD'D" and "RCV'D," and "CLM'D" is that status. Retrieving unfilled orders for claiming review is the type of job that the generic Scan function can perform superbly. One does not specify a claim period for each order in Bib-Base, although special claims periods could be indicated in a note field. Rather, one configures the Scan function to pull up all the orders between a range of dates or older than a specified date. From the review, a decision can be made on each aging order whether to claim. Then the generic List function is used to print batched lists of the claims for each supplier.

Reports from suppliers concerning the status of an order must be entered as notes. To keep reported orders from being retrieved by date for possible claiming, the status date on each such order must be changed to the report date.

Cancellations can be handled by a note on the acquisitions record, if the title is to be reordered from another source. However, if the title is not being reordered, the best way to remove the record from the encumbered statistics is to define a special Status (perhaps "CNCLD").

Bib-Base allows the library to define its own statuses (up to eighty) beyond the seventeen defined for the system. Such statuses do not affect the acquisitions statistics, yet allow records to continue to exist in the database. In our library, additional record statuses define

- low-priority order requests being held,
- order requests for out-of-print titles (printed for sending to an O.P. dealer using the generic List function),
- receipts that we returned to the supplier,
- personal book orders, and
- acquisitions records that are unused duplicates (deleted at regular intervals using the generic Scan function).

Performance: Without a specifiable claims period attached to each order, ALL orders initiated within a certain range of dates must be reviewed. A labor-intensive task, the impact is increased if the frequency of claim review must be changed to the shortest claim period to assure catching them promptly. However, in our library's experience, the system capability for reviewing orders and printing batched claims is so much faster than our manual system, the lack of the ideal claims set-up goes unnoticed.

Reports from vendors and cancelled orders are not recognized with their own Statuses by the Bib-Base system. Notes and special actions allow the system to accommodate them.

RECEIVING & INVOICING

Invoice check-in is done by searching the item and displaying an acquisitions record on the screen. For individual items, pushing a function key AUTOMATICALLY edits the acquisitions Status field to a configured choice of "RCV'D" or "INV'D," puts in the date invoiced, and parks the cursor in the received price field. By the time the cursor gets there, the order price has been replicated in the received price field, ready to be edited or accepted.

Batch-mode invoice check-in operates the same way. It offers running counts of titles invoiced and money expended, a final sum of money expended, the ability to amortize additional costs like shipping and handling over all of the invoiced titles, and will add the (searchable) Invoice number to each of the titles in one step.

However, there exists no BATCH-mode physical receiving function—one which is divorced from the invoicing function—in Bib-Base. Library Technologies may still be anticipating that libraries will pull a backup copy from a manual on-order file to receive an item, and that entering the invoiced price on the system will effectively mean that both receiving and invoicing have occurred. This may be the case with most Bib-Base libraries, or they may be willing to wait for the invoice to arrive before releasing already-received materials for cataloging and processing.

Printing routing slips for received titles must be done as an individual operation, as each record is on the screen, so that the slips are printed in the order the books are shelved on the book truck, another need which could be part of a batched physical receiving function.

What to do about receiving only part of an order? The Bib-Base manual does not mention what to do. Receiving the number in hand on the existing on-order record, entering a note to that effect, then replicating the record, and finally editing the new record to reflect the items still outstanding (i.e., still on order) is one possibility. We now do this. It works, and does not take as much time to do as to tell. Even when the supplier invoices for the total order, one can change the new acquisitions record to "pre-paid," and note the Invoice number.

Performance: The batch-mode invoice check-in function works slickly. Some time lag is experienced at the end of a long invoice of titles as the computer adds and indexes the Invoice number to each invoiced record. We do not use the individual invoice check-in function much. The lack of a batched physical receiving function hampers libraries which separate receiving from invoicing, and wish to dispense with manual on-order files.

Bib-Base uses the same acquisitions field for "received date" and "invoiced date," so inputting an invoice which arrives before the materials arrive would be confusing to the normal understanding of the acquisitions cycle (first "RCV'D" and then "INV'D"). A note would have to be added to indicate that the items had not yet been received.

Partial shipments can be accommodated, though not elegantly.

GIFTS, APPROVAL PLAN RECEIPTS

To successfully handle these "orders," an automated acquisitions system must make it easy to receive each title without having to order each title first. Bib-Base can do this succinctly, provided the bibliographic elements can be limited and brief. Defining one pattern record for gifts and another for approval plan receipts will cut out typing elements common to each title.

DEPOSIT ACCOUNT ORDERS

A specific acquisitions field in every record indicates whether the order has been pre-paid. The key to keeping track of deposit account orders on Bib-Base is to define a DIFFERENT supplier code for EACH account, and to monitor them by checking the online statistical reports by vendor.

STANDING ORDERS

Bib-Base does not receive items on standing order well. New receipts might be added to a single record for the title, by documenting the additions in notes fields and title-added entry fields. Or each receipt could use a new acquisitions record, and be linked by series to the standing order title. It is difficult to decide between such awkward choices. We use the latter.

Performance: Gifts, approval plan orders, and deposit account orders can be done successfully with Bib-Base. Standing order receipts (continuations) are problematic.

ACCOUNTING STATISTICS AND REPORTS

Accounting statistics are updated as the Status of each acquisitions record changes. This allows instantaneous access to an up-to-the-minute snapshot of the expenditure/encumbrance picture. It is possible to allocate one's acquisitions budget by Fund, by Supplier, and/or by Type. Collection management people may utilize the fund-based statistics most, but the supplier-based statistics are useful for monitoring deposit accounts.

Five hundred Fund codes, 200 Type codes, and 300 Supplier codes may be defined. For the Funds, Types, and Suppliers, three hierarchical levels of subtotals may be defined. This is not the same as hierarchical levels of funds, for example, because the subtotal funds cannot be assigned to an acquisitions record. Our library has defined just over forty fund codes with two levels of subtotals, four type codes, and twenty-some supplier codes.

Standard statistical displays or printed reports show budgeted amount, expended amount, unexpended balance, encumbered amount, and "free" balance. They are available by Fund, by Supplier, and by Type and for the current fiscal year and the previous fiscal year. Also shown are counts labelled "titles ordered," "titles received," "volumes ordered," and "volumes received." These are available by Fund, by Supplier, and by Type and also for the two fiscal years.

However, "volumes" is a misnomer. Bib-Base cannot give a true count of the number of volumes either ordered or received because the *system only keeps count of the number of copies*! The acquisitions fields do not include a place to indicate how many volumes each title represents, and Bib-Base does not read the 300 field (extent/physical description) to obtain it. So one copy of a three-volume work ordered on a single title record is counted as "1 title ordered" and "1 volume ordered" in the statistical counts.

Additional statistical reports can be generated by two more "generic" functions, which allow the compiling of selected and specially-defined statistics. These functions are used regularly in our library for reports by fund for the past month: the number and value of the firm orders placed and received, the number and price of standing order receipts, and the expenditures and encumbrances. The report parameters cannot be saved. But since dates are important parameters in these reports, and dates change, the inability to store the report parameters on the system is not vital.

Bib-Base keeps two fiscal years' statistics up-to-date. At the end of the year, the unfilled encumbered (and not pre-paid) orders can be transferred to either the new current year or the previous year by the touch of a function key. Budgeted allocations can be transferred as well. If a mistake is found while making any transfer, or soon afterward, the old arrangement can be restored. However, no transaction trail is kept by the system when budgeted amounts are changed.

Performance: Bib-Base can replicate all but one of the accounting statistics reports we kept manually in our old system. That report—total number of volumes ordered and received each month—cannot be deduced from the Bib-Base database.

REPORTS

The generic List function and the statistical report facility have been described. A limitation of the List function is that it cannot select or sort acquisitions records by BIBLIOGRAPHIC fields, only by ACQUISITIONS fields.

The print facility for acquisitions records in Bib-Base/Acq is another "generic" function. For printing parts of acquisitions records, three different tables of bibliographic fields can be set up and saved for printing duties. These tables are somewhat revealingly called "Order," "Bibliography," and "Wrapped bibliography." A final option is the "Full" bibliographic record, complete with MARC tags and indicators. One can combine one of these four print options with one of two printer choices, and use that combination to print a single record from the screen or to print a group of records using the List function.

For collection development purposes, subject specialists in our library have used the generic List function in Bib-Base/Acq to compile and print an alphabetized listing by subject of titles received and/or on-order. It should be noted that the Bib-Base/Subject module offers a much handier, online search and retrieval.

Compiled vendor performance reports are nonexistent in Bib-Base. The data—order price, received price, order date, and received date—is available within each acquisitions record. There just has been no program written to take advantage of it.

Performance: The report capability is far from perfect, but it is flexible enough to fill a wide variety of needs. There are no compiled vendor performance reports available.

SECURITY

Security for Bib-Base consists of a password system which allows defining which functions each password holder may use. It will deter casual unauthorized use of the database by outsiders, but any individual authorized to work at a functional level above "search only" can change some records in the database.

DOCUMENTATION

The Bib-Base/Acq operating manual is arranged with introductory material and acquisitions record information first, followed by topics in the order of the main menu of the program, utility programs, appendices, and an index at the end. The prose is readable, and quite clear. The illustrations are helpful, and most do double duty in explaining several parts of the text which follow. Cross references within the text are not as frequent as I would like, but do occur. The most useful appendix for me has been the table of valid MARC tags, for all MARC formats. The index is long, but not exhaustive, especially when trying to locate a particular function.

What I miss most in the operating manual are workflow examples. There are occasional suggestions, but not of the scope needed by the first-time purchaser of an automated acquisitions system!

EASE OF LEARNING

It is relatively easy to learn to use Bib-Base/Acq, once one has conquered the fear of starting. Perhaps that is the best way. It is in fact the ONLY way—there are no tutorials. One "eases into" the program using one's own fund codes, budget, and suppliers.

EASE OF USE

The screen displays are clear, with judicious use of reverse video to highlight acquisitions field names in the acquisitions record display. The choices of functional operations are always surrounded by a box at the very top of the screen. There is also a Lotus-like moving bar to allow one to use the keyboard arrows to choose a function, or to choose a particular record from a list of records. I do not like the moving bar; it distracts the eye, and would be more at home in a system for the occasional user. I prefer using the function keys or type the entry numbers instead.

With the acquisitions record at the cen-

CURR Year by Fund -- 07-APR-89

item	budget	expen	balance	encum	balance	--ordered-- title	--ordered-- vol.	--received- title	--received- vol.
A-1	1100.00	40.00	1060.00	10.00	1050.00	1	1	3	3
A-2	1200.00	50.00	1150.00	30.00	1120.00	1	1	2	2
A-3	1100.00	0.00	1100.00	40.00	1060.00	2	2	0	0
A-4	760.00	30.00	730.00	0.00	730.00	0	0	1	1
A-5	400.00	70.00	330.00	60.00	270.00	2	2	1	1
+HUMANTES	4560.00	190.00	4370.00	140.00	4230.00	6	6	7	7
C-1	800.00	20.00	780.00	30.00	750.00	2	2	1	1
C-2	1900.00	90.00	1810.00	140.00	1670.00	6	7	2	2
C-3	1100.00	0.00	1100.00	110.00	990.00	4	4	0	0
C-4	700.00	30.00	670.00	40.00	630.00	2	2	4	4
C-5	1100.00	60.00	1040.00	80.00	960.00	2	2	1	1
C-6	500.00	0.00	500.00	0.00	500.00	0	0	0	0
+SOC SCIS	6100.00	200.00	5900.00	400.00	5500.00	16	17	8	8
++MONOGRP	10660.00	390.00	10270.00	540.00	9730.00	22	23	15	15
GIFT	0.00	0.00	0.00	0.00	0.00	0	0	4	4
TOTALS	10660.00	390.00	10270.00	540.00	9730.00	22	23	19	19

Fig. 2. Statistics Report: Current Year, by Fund

ter of most transactions, rather than some nest of menus leading one toward a decision, the capability for doing many different operations from that one display is very powerful. It takes more learning, but at the end there is more a sense of knowing acquisitions rather than knowing just the routes through the menus.

ERROR HANDLING

Most functions can be undone after they have been done. All record-processing functions can be aborted. The most finicky part of the software is its response to a non-powered printer: the program exits to the DOS "Abort, retry, ignore?" error message, and occasionally someone chooses the wrong response. The datafiles must then be rebuilt, using the BBREBLD utility program. The program works well, but with upwards of 12,000 records, it takes time.

SUPPORT

Bob Kepple and Jim Schoenung offer great telephone support. They are candid about the shortcomings of Bib-Base, but can often find a workable solution. Rarely are they at a loss.

Six months of free support comes with the purchase; after that support costs 10 percent of purchase price per year.

We have found it to be a good value! In its price range, there is no automated acquisitions system that has as much horsepower and versatility. Bib-Base is designed for the broad mainstream of acquisitions possibilities, and relies on its generic functions, record Status field, and free-text notes to handle uncommon events. It is a good design philosophy, and one we have come to appreciate as we push the system's capabilities more and more.—*David Buxton, Software Reviews Editor.*

INDEX TO ADVERTISERS