

DEBORA SHAW

Associate Professor
School of Library and Information Science
Indiana University
Bloomington, Indiana

Libraries and Mainframe Computers, or When Do You Need a 747?

INTRODUCTION

Historical Notes on Databases

In consideration of the long-standing title of these meetings as “clinics on library applications of data processing,” we should remind ourselves that data processing is a means of improving the work of libraries as information-handling systems. Information has been defined as “data placed in context” (Loomis, 1987, p. 3) with the database as one part of the context, and the library another. We are also concerned with data from the system’s viewpoint, noting that one goal of database management has been to “create more independence of the data from the programs that access them” (Lucas, 1986, p. 220).

These quotations highlight important aspects of how databases and their associated software have evolved, and how they are viewed by current developers and knowledgeable users. Data are an essential component of information, and hence of information systems, including libraries. Because of their enormous processing power compared with manual filing and retrieval systems, computers can be used to create a revolution in library services. It is therefore incumbent on librarians and information specialists to understand and make the best possible use of computer power in information handling.

Early computer systems used files of data, but did not treat these files as a coherent whole, or database. One advantage of a database system is the ability to make a collection of data available to many users, not unlike the goal of a library to provide a pool of resources for a variety of patrons. IMS, the first database system developed by

IBM and North American Aviation in the 1960s (Loomis, 1987, p. 177), represents a hierarchical database structure, in which a record for a specified type of information can have dependent records. For example, a BOOK record might have COPY records providing details on each copy of a book in the library. In 1971, the Data Base Task Group of the Conference on Data Systems Languages (CODASYL) published its description of the network data model (Loomis, 1987, pp. 131-32). This structure emphasizes one-to-many relationships (networks) among records of different types. For example, an ORDERS record could be related to REQUESTORS and SUPPLIERS.

The third database model is the relational model proposed by E. F. Codd which found commercial applications in the 1980s (Loomis, 1987, p. 78). In a relational system, the data are seen as flat tables, with all information on an item presented in one row or *tuple*. Tables (also called relations) can be joined, for example, to pull together an list of orders placed with foreign sources from an ORDERS table and a SUPPLIERS table. The relational approach has the advantage of allowing any question to be asked of the database—the user is not limited to the retrieval approaches anticipated by those who designed the hierarchy or network. However, a relational system requires considerably more computer power, especially to support a large database.

Traditionally, data have been considered distinct from programs, and common wisdom holds that the more distinct, the better. This attitude has allowed development to proceed on record format (e.g., the MARC record), record content (e.g., cataloging rules), and user interfaces (e.g., online catalogs) without requiring that those involved understand in detail the procedures that will be used to store or retrieve information in the database. This division of responsibility has been helpful, but may have created some artificial distinctions between data and programs. Research in information science has revealed clues as to how people use information; for example, Richard Trueswell's (1965) finding that 80 percent of the questions to a system can be answered with 20 percent of the system's resources, and recent studies of how cognitive processes affect the search for information (Borgman, 1986). When systems designers take such findings into account, improvements in both human-computer and program-database interfaces can be developed.

LIBRARY USE OF MAINFRAME COMPUTERS

Libraries use mainframe computers in two ways. At times, libraries are customers or clients who "reach out and touch" large databases,

housed on large computers and serving a large number of customers (which may be libraries, other institutions, or individuals). Bibliographic utilities and search services are examples of this kind of use.

A library may also use a mainframe which supports a local automated system for circulation control, online catalog access, and other applications. In this case the computer may be owned by the library, its parent institution, or a library group. The computer is likely to be housed in the same city or state as the library. The library is likely to have more to say about planning for a local system, and generally deals with a systems staff and/or computer center directly responsible for development and maintenance.

In June 1988, a special section in *Information Technology and Libraries* described various experiences in measuring system performance, including capacity modeling at RLG, response time and performance analysis with MELVYL, and measuring system performance with Carlyle. Describing the problems of performance analysis and improvement, Clifford A. Lynch (1988) notes that, "in most real systems performance is limited by a small number of bottlenecks at any given time; however, when one is eliminated, a new one will limit system performance" (p. 178). The efforts reported in this special section are uncommon, but all systems need detailed information on current system performance in projecting future needs. Speaking from the library's perspective, Julie Brown (1988) advises other libraries to "consider how to get the necessary [performance] information in another way" (pp. 184-85).

In late 1988, ten providers of mainframe services to libraries were surveyed on the hardware and software that support their database applications, particularly the retrieval/access aspects librarians use. Questions addressed the types of computer(s) and associated software, size and annual growth of database(s), and number of users. The institutions surveyed represent the variety of sources through which a library might use a mainframe computer. Online search services providing information were Chemical Abstracts, Dialog, National Library of Medicine, ORBIT, and Wilsonline. Bibliographic utilities were OCLC, RLIN, and WLN. The University of Illinois' LCS/FBR and NOTIS's installation at the Florida Center for Library Automation represented online catalog and circulation systems. The results of this survey are given in Table 1.

The respondents use a variety of equipment, five with various types of IBM mainframes, two with National Advanced Systems computers, two with Amdahl machines, one with Unisys in addition to IBM, and one with Xerox and Tandem equipment. The programs for ongoing

operations are written in one or more of the following languages: Assembly language (different for each machine)—eight systems; PL/I—eight systems; C—two systems; and Pascal—two systems.

TABLE I
EXAMPLES OF MAINFRAME COMPUTER SYSTEMS USED BY LIBRARIES

	<i>Computer</i>	<i>Programming Languages</i>	<i>Number of Databases</i>	<i>Millions of Records</i>	<i>Database Growth Rate</i>	<i>Simultaneous Users</i>
CAS Registry File	IBM 3090 Unisys	Assembler C, PL/I	1	8	10%	n.a.
Dialog	National Advanced Systems XL80, XL60, 9080	Assembler PL/I, Pascal	311	193.8	40%	several hundred
NLM	IBM 3081 IBM 3084	Assembler, PL/I	26	11.6	3%	250
NOTIS (Florida)	IBM 370 systems	Assembler	1	5.4	2% (typical)	n.a.
OCLC	Xerox Sigma 9, Tandem TXP	Assembler, C	1	18	40%	9200
ORBIT	National Advanced Systems XL70	PL/I	100	100	40%	n.a.
RLIN	Amdahl 5890	SPIRES, PL/I, Pascal	2	36	14%	1170
Univ. of Illinois	IBM 3081	Assembler, PL/I	31	15	6%	1000
WLN	Amdahl 470	Assembler, ADABAS, PL/I	3	14.7	5%	270 terminals, avg. 110 users
Wilson	IBM 3081	Assembler, PL/I	22	4.6	11%	30 external 175 internal

A rough measure of size is possible from the number of records contained in each system's files. The largest system, Dialog, reported 311 databases, containing a total of approximately 194 million records. The smallest number of records was reported by Wilsonline, with 22 databases and 4.6 million records. The average (mean) number of records

in all ten systems is 40.7 million; excluding Dialog, which is unusually large, the average is 23.7 million records per system. There is a great range in number of characters per record, from MARC cataloging records, to bibliographic records including abstracts, to full text of articles in some files supported by the search services. In addition, different systems support differing levels of detail and flexibility in accessing the databases. For these reasons it is not possible to use number of records to make absolute comparisons of size. However, the number of records does reflect the complexity involved in database navigation and record retrieval.

The systems grow at widely varying rates, with NOTIS reporting an average annual growth rate of 2 percent for a "typical system" to the 40 percent growth reported by Dialog. The bibliographic utilities averaged growth of 11 percent per year, while the search services averaged 21 percent annual growth.

The number of simultaneous users each system supports ranges from 205 for Wilsonline (of which thirty can be outside, i.e., non-Wilson searchers) to 9200 for OCLC. It is interesting to note that the highest number of simultaneous users of a search service was "several hundred" for Dialog, which is presumably lower than the 1170 reported by RLIN, the bibliographic utility supporting the fewest simultaneous users. Furthermore, bibliographic utilities users need dynamic access—they are adding or changing records even as they and other users search and retrieve from the database.

Advantages of Mainframe Computers for Library Applications

Every week seems to bring another breakthrough in computer technology, making unit sizes smaller, faster, and less expensive (measuring instructions per second per dollar). Still, many library operations rely on big mainframe computers. Why? The obvious reasons are processing speed and the ability to handle large files and many transactions. With the large numbers of users mentioned above, many doing complex search operations, it is critical that the system be able to receive, process, and respond to commands rapidly. Simply retrieving information from large numbers of disk drives requires a powerful machine; moreover, mainframe architecture supports many independent "channels," or disk controllers, providing greater throughput. To date, only mainframe computers have been able to provide the speed of execution needed to support many users with real-time access to large files. Database machines, hardware designed to optimize database functions, have been proposed as one way to improve system performance (Salmon, 1984). To date, library systems have most often chosen instead

to stay with general-purpose computers; and the development of computer power seems to more than keep pace with the design improvements offered in the database machines.

Some more subtle considerations also speak in favor of mainframes for library applications. While the size and complexity of databases is impressive, even more staggering is the investment of storage and processing power required to develop and maintain the indexes that support the sophisticated retrieval capabilities to which libraries have become accustomed. After years of experience with the search services, Boolean logic is no longer enough. Now we want field restrictions, string searching, proximity operations, and other refinements. The problems of textual databases, especially full-text databases, require more discrimination in the retrieval process. The addition of artificial near-intelligence will place additional demands on systems, as knowledge bases and search heuristics are developed.

As databases become larger and more complex, and as non-specialists use these systems, we see the need for more helpful or "interventionist" search software. Lynch (1987) has described problems with very large bibliographic databases, where the traditional search keys (e.g., subject headings) produce too many hits. Advanced retrieval software can specify the complex search statements for beyond-Boolean searching and provide assistance to new or infrequent users. However, this sophisticated software will likely require more computer speed and power.

As systems become more powerful they must also become less hostile (more friendly) and easier to use. John Scully's "knowledge navigator" envisions such an interface, applying computer processing power to traverse a large file and find useful information (Apple, 1988). It would seem logical to house the navigator's capabilities in a personal workstation rather than a central mainframe, but mainframes will need to be consistent enough to interact with a variety of navigators and powerful enough to support several sophisticated users at a time.

Another reason for library reliance on mainframes may be, as Dennis Reynolds (1985) has observed, that "libraries are generally adapters rather than innovators of technology" (p. 159). This desire to use proven hardware and software has been a prudent response to rapid changes in technology, notably when computer generations were succeeding each other with great rapidity in the 1960s and 1970s. Given libraries' responsibilities to preserve information, a cautious or skeptical approach to new developments may be in order. However, one consequence of this caution is that libraries are not (or are not often) on the cutting edge of technical developments, but rather adapt existing technology for library applications. Thus breakthroughs in computer hardware or

database management are likely to be fairly well-established by the time they are adopted by libraries. The trick is to adopt a breakthrough before it has become a dead end.

Human Issues in Use of Mainframes

One of the attractions of mini- and micro-computers is the sense of participation and autonomy possible for the computer's users. The ultimate feeling of individual control is embodied in the notion of a "personal computer." In fact, one handy way of distinguishing between micros, minis, and mainframes is based not on speed or power but on cost, the ultimate determinant of ownership. A micro can fit into a personal budget, a mini into a departmental budget, and a mainframe into an institutional budget. When a library uses a big, sophisticated computer system, it is unlikely to own the computer, and in fact is seldom the only user. Rather than dealing simply with one's own designs and implementation problems (as with a personal computer) or with one or two systems librarians (as is common with many turnkey systems), the library as an institution may be working with (or against) a large number of systems staff whose allegiance is to another part of a university, corporation, city, or even to an outside vendor. Regardless of the good will and charitable emotions professed at the start of such a venture, there are times when differences of opinion are unavoidable.

The following observations on how people react to the stresses of library automation are based on my own experiences and on discussions with colleagues who have worked as systems staff members:

"Us vs. Them"

Library staff members have widely different expectations of an automated system and different degrees of willingness to make sacrifices for its implementation (Fine, 1985; Shaw, 1986). For various reasons, they may need or want a scapegoat when things go wrong. Similarly, computer operators, trainers, sales representatives, network librarians, and others feel frustrated by the reluctance of library people to "give the system a break." When these two groups work in different buildings (or even in different states) report to different supervisors, and may even have different overall goals for their work, it is not surprising to see differences develop.

The Project Mentality

Library automation is often undertaken as a special project. There are many milestones to full implementation of the system, and significant amounts of energy are expended and goodwill due bills called in. However, there is generally little attention to the ongoing management

of the automation system and the demands this will place on library staff and users. The initial emphasis on implementation is dwarfed after five years' ongoing effort and costs.

Limited Reward for Preventing A Catastrophe

Systems people spend a great deal of time doing things that everyone hopes the users will not see. Developing, debugging, and testing new applications is demanding. Since each large bibliographic system is unique, many even minor changes involve developing and testing new programs.

CONCLUSION

Libraries need mainframe computers and powerful database management systems to put data in context: to create information. The notion of a "databank" or "database" has been around since at least the 1960s. In 1973, Charles W. Bachman urged a revolution in database management, from a computer-centered to a database-centered viewpoint. It may now be time for the next revolution, from database-centered to a view that encompasses the information system of which the database is a part. This integrated information outlook must stress the seamless integration of hardware, software, databases, and intelligence to provide the information each user needs.

The 747 airplane mentioned in the title was chosen as a symbol of size and power. As we consider the impressive power of the computers and systems available to libraries, we should understand the importance of the environment in which a powerful system works. Recent experiences with airliners have alerted us to the need for ongoing attention to the human and mechanical aspects of maintenance and planning.

The demands of information processing have brought together technology ranging from micros to supercomputers. Library automation is generating communication among people who previously had little in common, and in some cases is reducing the differences between them. It may be time to reconsider distinctions, such as that between data and programs, which have traditionally been considered reasonable. It is not clear that the next generation of systems will place the librarian in the pilot's seat of a Concorde or a 747, but it is time to realize some of the great expectations librarians have long had for automation.

REFERENCES

- Apple Computer, Inc. (1988). *The knowledge navigator* [Videorecording]. Columbia, MD: Television Arts Production.
- Bachman, C. W. (1973). The programmer as navigator. *Communications of the ACM*, 16(11), 653-658.
- Borgman, C. L. (1986). Why are online catalogs hard to use? Lessons from information retrieval studies. *Journal of the American Society for Information Science*, 37(6), 387-400.
- Brown, J. (1988). Measuring system performance: The library's perspective. *Information Technology and Libraries*, 7(2), 184-185.
- Fine, S. (1985). Terminal paralysis, or, slowdown at the interface. In D. Shaw (Ed.), *Human aspects of library automation: Helping staff and patrons cope* (Papers presented at the 22nd Annual Clinic on Library Applications of Data Processing, 14-16 April 1985) (pp. 3-15). Urbana-Champaign: University of Illinois, Graduate School of Library and Information Science.
- Loomis, M. E. S. (1987). *The database book*. New York: Macmillan.
- Lucas, H. C., Jr. (1986). *Introduction to computers and information systems*. New York: Macmillan.
- Lynch, C. A. (1987). *Extending relational database management systems for information retrieval applications*. Unpublished doctoral dissertation, University of California at Berkeley.
- Lynch, C. A. (1988). Response time measurement and performance analysis in public access information retrieval systems. *Information Technology and Libraries*, 7(2), 177-183.
- Reynolds, D. (1985). *Library automation: Issues and applications*. New York: Bowker.
- Salmon, S. R. (1984). Applications of database machines in library systems. *Library Hi Tech*, 2(4), 69-72.
- Shaw, D. (1986). Staff opinions in library automation planning: A case study. *Special Libraries*, 7(3), 140-150.
- Special section: Measuring system performance. (1988). *Information Technology and Libraries*, 7(2), 173-197.
- Trueswell, R. (1965). A quantitative measure of user circulation requirements and its possible effects on stack thinning and multiple copy determination. *American Documentation*, 16(1), 20-25.