VELMA VENEZIANO
Systems Analyst
Northwestern University Library

and

JAMES S. AAGAARD
Associate Professor of Computer Sciences and Electrical Engineering
Northwestern University
Evanston, Illinois

Cost Advantages of Total System Development

The question we will consider in this paper is whether, at the end of a decade of effort to harness computers to the needs of libraries, it is economically feasible and operationally practical for an individual library to design and operate its own in-house automated system. At Northwestern University Library, the answer is both yes and no, but more yes than no.

In his 1975 article, "Library Automation: The Second Decade," Richard DeGennaro says "no." To quote him:

> Many of the premises upon which research libraries based their decisions to build in-house library systems staffs to automate their internal operations in the late 1960's are no longer valid. Important advances in automation have been made, including the widespread acceptance and use of the MARC format and distribution service, the general success of the cooperative network concept, and the availability of package systems. The day of the one-man or small group library systems development effort is past. The jobs to be done and the equipment required have become complex and expensive, and it requires a team of highly qualified computer specialists to design and

> implement a viable system. With the increasing sophistication and success of computerized systems for libraries, the need for systems groups in individual libraries is diminishing.
>
> The era of localized library automations has effectively come to an end. Experience has shown that it is not economically feasible for any but the very largest libraries to afford the heavy costs of developing, maintaining, and operating complex localized computer-based systems. Many libraries are quietly abandoning this approach in favor of joining networks such as OCLC or its affiliates or purchasing turnkey mini-computer systems from commercial vendors for specific applications.[1]

DeGennaro correctly identifies a trend, but for the wrong reasons. In our judgment, library automation efforts have failed or been minimally effective because libraries have not approached automation realistically.

In the past it was often assumed that a librarian, by simply taking a few courses in data processing, could direct the library's automation effort. Even worse, it was often believed that people from the local computer facility, with no knowledge of libraries, could tackle the problem effectively. There were unrealistic expectations about the time required to do the job, with many administrators believing that a year or two would be sufficient for systems to be developed and to begin paying off in cost savings and improvements in efficiency.

Some libraries underestimated the complexity of the problem; others overestimated it. Using huge staffs paid with grant money, these libraries found that too much staff time was spent attempting to communicate with one another and in writing reports. Under such conditions, the amount of progress tended to be inversely proportional to the number of people involved and the amount of money being spent.

Fortunately, Northwestern has avoided many of these pitfalls. We began our development more than eight years ago with a full-time staff equivalent to one and two-thirds people, which has now grown to two and two-thirds people. With dogged determination and little money or fanfare, we have been making progress, although it sometimes seems to be very slow.

We believe that the Northwestern University Library is more highly automated than any of the large research libraries. Our system was developed with no outside grant money, and the development cost was a fraction of what has gone into the development of some of the other more highly publicized systems. Development costs for the first five-year period were about $300,000, including data conversion costs. For this expenditure we have operational on-line ordering, cataloging, serial check-in, and circulation systems, as well as a batch SDI (selective dissemination of information) system.

Our operational costs are moderate and well within the resources of any large library. Not including terminals, most of which we own, our computer costs are less than $120,000 a year. Considering that we have a book budget of about $1 million, this is not excessive.

We catalog about 40,000 books a year, producing about 500,000 catalog cards and 20,000 purchase orders. We prepare more than 100,000 worksheets and about 12,000 claims in a year. We circulate about 400,000 books, producing fine notices, book-needed and book-available notices, and overdue notices. We print pocket labels and produce punched circulation cards automatically. We check in about 90,000 periodicals yearly, in addition to monographic series.

We do not yet have a true on-line catalog; this is the module which is presently being developed. We do, however, have our entire serial collection—approximately 40,000 titles—on-line, and we have bibliographic data in machine-readable form for about 200,000 monographs.

Our success is due to a combination of factors. In part, we were lucky in having just the right set of conditions at the right time. There are literally hundreds of factors which can influence a project like this; we will try to outline a few of the most important ones.

Economy of Scope

One of the prime rules for an effective and economical in-house automation system is that it be comprehensive. It is essential to realize that there is no single operation performed by the library which, by itself, can be automated economically. Ordering systems, cataloging systems, circulation systems, or serial systems, if designed and operated in isolation from one another, tend to be costly and have minimal impact on overall library efficiency. This realization is behind the mass movement to the Ohio College Library Center (OCLC), which produces catalog cards in phenomenal quantities, thus achieving the objective of "economy of scale."

Because an in-house system cannot take advantage of the "economy of scale" concept, it must be designed to permit "economy of scope." By spreading the costs over a broad base of applications, the cost of any one application can be minimized. From the very first, Northwestern's system was conceived as a "total integrated system," with the objective of eliminating all manual files, including the card catalog. However, it is a very complex task to design and implement a total system. No matter how much money is available and how many people are assigned to the task, there is an irreducible amount of time required to design and implement a system. Unfortunately, library and university administrations, like their counterparts in business and industry, take a dim view of projects which drag on for years without visible results. Next year's appropriations are often dependent on the demonstrated results of the current year. For this reason

it is essential that the total system be designed so that it can be implemented in modules.

In addition to using the modular approach, and because of the necessity to establish credibility as a basis for continued funding, it is often necessary to design a first-generation system which does not have all of the "bells and whistles" which we have come to expect of computers. This is a perfectly valid approach, provided one is aware that sooner or later such a "stripped down" system will have to be enhanced. Substantial investments in time and money may be required to make such enhancements.

The cost-effective in-house library system, in addition to being "total" and "modular," must take advantage of every possible money-saving device. The concept of "multiple-use" data is essential to an economical system. For this reason, the separation of systems—circulation from cataloging, for example—is to be avoided. Thus, in an ideal system, a file of call numbers can be used both as circulation control and as a shelflisting tool. This same file becomes a means by which an ongoing inventory of the collection can be maintained, reducing the amount of lost time and annoyance created by lost or misshelved books. The file is valuable for special studies; for instance, we have used our inventory file at Northwestern to help us to evaluate our book loss problem and to project space requirements.

An effective and economical system must also avoid redundant data entry. With an integrated order/catalog system, author, title, and imprint information entered for purposes of ordering can be modified or used "as is" for the cataloging function. However, the avoidance of redundant data entry does not necessarily mean that data redundancy in files is undesirable. We have had to be constantly on guard against designing a system which is wasteful of computer processing time. Our philosophy has been that with computer storage costs declining so much faster than processing costs, reduction of the latter has the greater priority. For example, we have not yet been convinced that complete inversion of bibliographic records is practical. Although we strive to take advantage of data compression techniques wherever possible, we think that some data redundancy is a small price to pay for processing efficiency.

The Northwestern system is not yet a total system, although it more nearly approaches that goal than any we know about in other large research libraries. In its present configuration we cannot claim to have achieved any cost savings; the best we can say is that it costs no more than it would cost to do the job manually. However, it does the job better and faster. We have no way of determining the value of this.

The important thing to stress is that we have established the foundation for moving on to the next module—the on-line catalog. This is the area in which we anticipate the greatest payoff. It is our conclusion that the

cost of library operations can be affected substantially by automation only when it enables us to cut off our slavery to the card catalog. As long as technical and public service personnel are tied to this monster, and as long as catalog copy, whether from MARC or not, must be integrated and reconciled with it, then there will be little increase in efficiency.

This brings up something which has become a sore point with the catalogers at our library. They have heard stories of libraries which claim to have greatly increased their cataloging productivity as a result of using the OCLC system. Because we ourselves use MARC copy, acquired automatically from the MARC tapes, we wonder how they have been able to do it. Although we have virtually eliminated the typing and reproduction of catalog cards, thus cutting our clerical costs, we have not been able to cut the cost of the cataloging operation itself. Frankly, we are skeptical of claims of large cost savings. We suspect that the transition to OCLC has given administrators an opportunity to make changes—in organization, in the level of personnel assigned to cataloging, and in work-flow and procedures—which by themselves would have increased productivity. Such improvements did not accompany the changeover to automation at Northwestern because we had already streamlined our cataloging operation.

Our catalogers also wonder if some of these reported cost savings are not being achieved at the cost of an impaired catalog in which the user and public service personnel will pay the price of lost access to materials. Even though we have had to compromise the quality of our cataloging to some degree, we still take reasonable precautions to avoid conflicts with earlier cataloging.

The Objective

We have said that we expect a "total system" to improve greatly our operating efficiency, and if it does not allow us to cut costs, we expect that at least the per-unit cost of processing a book will level off. We also expect that the user will benefit greatly from the system. However, we have not yet defined exactly what this total system will be. We expect it to be an on-line system with a file of bibliographic records for all items either held by the library, on order, or in process. Linked to these bibliographic records are files of local processing and control information—order records, holdings records, circulation records, invoice records, fund records, and patron records. Patrons as well as librarians use the system directly, searching the files by means of any of several access points. The card catalog as a means of access to materials acquired since 1971 is gone; also gone are manual files of orders, bindery records, serial check-in records, and so forth. In their places are terminals, in public areas as well as in the processing areas.

After searching to see if a potential purchase is already held, the de-

cision to order is made. Bibliographic and order records are created at this time, and purchase orders are produced by the system. A commitment is automatically entered against the appropriate fund. When an item is received, whether it is a monograph, a multivolume set, or a single issue of a journal, its receipt is recorded in the order record. A record for each vendor invoice is created and updated automatically as the items on it are checked in. When the sum of the line items balances the total, the invoice is approved for payment and the check is written automatically. Claims for overdue items (books as well as journal issues) are also generated automatically. Using a terminal, catalogers review and update the bibliographic record at the appropriate time and request book materials (labels and punched circulation cards).

Patrons use terminals to search for wanted items; having found them, they can interrogate the circulation file to see if the items are available. Using self-service terminals, patrons charge out their own books. The circulation system takes care of the production of overdue notices, call-in notices, and notices of books available. Although fines are assessed in cases of gross delinquency, for the most part the system is self-regulating, blocking a user from taking out books after he has accumulated a certain number of "demerits."

When we make the transition to a true on-line catalog, enabling searchers and catalogers to search and modify records from a terminal, we estimate that we can achieve about a 30 percent increase in productivity on the part of our technical services staff, both professional and clerical. The savings should be more than enough to offset the additional costs of terminals and computer time.

It is important to point out, however, that there are two aspects of library operations which are not good candidates for an in-house system. The first is the maintenance of a large data base such as the MARC file in an on-line mode. This is completely impractical for a single institution; it must be done on a regional or national basis. At present we maintain and search the MARC file off-line. With the file at its present size (about 600,000 records), we can do this more cheaply than we could if we participated in OCLC. This will become difficult by the time the number of records reaches one million, however, and we sincerely hope that there will be a method of acquiring catalog copy for direct transfer into our files at a reasonable cost.

The other area which makes sense only at the regional or national level is the maintenance of a union file of holdings, one which can be searched in order to locate items for interlibrary loan. However, because the volume of materials which we borrow is relatively low, this has not been an area of major concern. With the steady increase in the cost of purchasing materials, this might become of more interest to us in the future.

Hardware

We will now examine some of the details of the system design, both hardware and software, which have enabled us, with a modest investment, to accomplish what we have.

At the time we started our development effort, we were not aware that the comprehensive system we envisioned could not be implemented on the small IBM/360 Model 30 computer that the university was using for administrative purposes, so we proceeded to do it. (Actually we had little choice.) For data storage we used part of a "data cell" which the university had obtained primarily for storage of alumni records, and we located an inexpensive source of special type balls for our terminals. As we were able to build the administration's confidence in our abilities, we managed to get the computer storage upgraded to 96K and encouraged the replacement of the data cell with disc storage. During the entire project we have been required to justify each increment in computer capability.

Fortunately, we have been aided by advances in computer technology. The Model 30 was replaced by an IBM/370 Model 135, at essentially the same cost but with a substantially increased CPU speed. This helped us to accommodate the steady increase in work load. Further increases in storage were needed to accommodate the teleprocessing monitor program (CICS) which we later elected to use. At the present time, this computer has 192K of storage, and we estimate that one more small increase will be enough for our growth in the next two years.

We emphasize that this is not a dedicated library computer; it is used for all university administrative data processing, and this is during the first shift hours when library usage is heaviest. For this reason our response time sometimes slips below what we would like it to be, but we know that we are not alone with this problem. We conducted a study a few years ago to compare our charges with the costs of a dedicated computer. The university charges our account approximately $120,000 per year for data processing services. This amount includes batch processing for catalog cards and purchase orders, batch processing for SDI services, batch processing for program development, and teleprocessing charges (see Table 1). By doing all library batch processing on the second shift, when the teleprocessing load is much lighter, we felt that an IBM/370 Model 115 with the same amount of storage could handle the load of the library system alone, and we found that the cost of such a system was almost identical to what we were being charged by the university. The university was understandably not willing to let us make this change, because there was no way the cost of the administrative computer could be reduced by $120,000 if the library pulled out. However, we think this does illustrate that it is possible for a library to have a system like ours, even if a computer which can be shared is not available.

In fact, library applications tend to require a relatively large amount of storage, both in the computer and externally, and relatively low computation speeds. For this reason it is quite possible that a library which shares a large, high-speed computer encumbered by a complicated operating system may be paying for computing power which it does not really need. We cannot offer any data to support this thought, because the opportunity has never been available to us, but we would caution you, especially in this day of what one manufacturer calls the "megaminicomputer," not to overlook what can be done with a relatively small computer.

We have tried very hard to avoid efforts to develop special hardware. This has caused inconveniences in several areas, particularly in the entering and display of special characters in bibliographic records, and in providing a reliable output device for our self-service circulation terminals. Development of hardware of this type can be very expensive and time-consuming, for it requires expertise in all areas of computer science, from electronic circuits to data communications and operating systems. The problems with the circulation terminal finally became sufficiently annoying that we entered into an agreement with a group within the university to develop a new terminal. This development has been in process for more than a year, and we have yet to see an operational prototype. We have not given up hope, but this experience has reinforced our determination that hardware development is to be done only as an absolute last resort.

Another recommendation is to deal with as few vendors as possible. The computer trade press likes to headline the dollars which can be saved by "shopping around" for computer hardware, but for a small installation these dollars may not be worth the annoyance. A large computing center, with dozens of tape and disc drives, may be able to save more than enough to pay the salary of the hardware specialist who can draw up specifications, help with selection, and then pinpoint which vendor's equipment is the cause of a particular system failure. Again, we have not been able to follow our own advice completely; our hardware comes from three different vendors (if the telephone company is included). Our experience, however, reinforces our recommendation that this is a situation to be avoided if possible; the savings of $5,000 or $10,000 a year is not sufficient reason to add another vendor.

It must also be decided whether to buy or lease the equipment. There is no question that purchase or a long-term lease can save a considerable amount of money if the equipment will be used for a period of five or more years. Hindsight indicates that we, and our university, have not always made the best decision in this respect. Although the CPU represents the largest single cost in the total hardware budget, a short-term lease—or a long-term lease with appropriate contractual provisions—will facilitate the gradual expansion of the system as the various modules are implemented. The other

Function	*Processing Charge*		*Terminal Cost*
	Batch	*Teleprocessing*	
Circulation and Public Services	$16,000	$17,700	$21,500
Technical Services	27,000	41,500	18,000
SDI	13,500		
Development	1,500		4,000
TOTALS	$58,000	$59,200	$43,500

Table 1. Computer costs for Northwestern University Library, 1974-75

alternative is to pay for more initial capability than is needed and hope that someone else doesn't make use of it before the library needs it.

On the other hand, the peripheral devices seem to be more suitable for a long-term commitment. Tape and disc drives and terminals can be used with different computer models, and more units can be added without affecting the status of those already in operation. In our case our biggest mistake was in not purchasing our circulation terminals, which we have had for over six years; the primary reason for leasing was that we kept hoping that something more suitable would appear.

There is one other criterion which can be applied: as a general rule it is safer to make a long-term commitment to equipment which has just been introduced than to devices which have been in the field for several years. Special consideration is necessary if the new equipment comes from a new vendor—will the vendor still be in business in five years to provide parts and service?

Software

It is, of course, the software which can make or break a system, and it is here that the system designer is faced with a bewildering array of choices. Should it be written in the library, in the computing center, or contracted with an outside vendor? Should you try to incorporate existing data management or communication management packages into the system? Should the programs be written in assembly language, COBOL, or PL/I?

To answer the last question first, the combination of the complexities of

bibliographic data and an on-line environment almost dictates programming in assembly language. There is no doubt that this requires a more highly skilled staff and raises the initial cost of the programs, but this money is recovered over and over again in the daily use of these programs. A well-designed assembly language program will occupy much less storage itself, require less execution time, and often permit compression of data, as compared with a program written in some other language.

The decision whether to purchase generalized system components, such as data base management systems, or telecommunications monitors is less clear. Available packages must be examined to determine how closely they fit the library's requirements, whether they will operate on available hardware, and their cost. For example, we found that the capability to process variable length fields usually was lacking in the data base management systems we investigated. On the other hand, capabilities which are included in the package but which are not required by the library can substantially increase the hardware needed to operate them.

A few years ago we elected to modify our system to operate under the IBM telecommunications monitor system called CICS. We are still uncertain whether this was a wise decision. The package is widely used, its price is reasonable, and it appears efficient. It is also difficult to learn and uses much of the computer storage. We have not acquired any data base management system; most of them are poorly suited to bibliographic data and we feel that their price far outweighs any benefits. We are, however, using IBM's new VSAM (virtual storage access method) for file management. This is undoubtedly an improvement over the old indexed sequential system, but still appears to be overly general, much less efficient than it could be, and grossly wasteful of storage. Fortunately, file management is a well-defined part of the whole system, so that if something better comes along, or if we have the time to develop something ourselves, it should be possible to incorporate it without a major impact on the overall system.

The idea of acquiring a complete library system seems attractive, but there are few, if any, to be found. A system like ours, which is economical to operate because it is designed to do exactly what we want it to do and no more, would undoubtedly require extensive modification by another library. A very generalized system, designed to provide for the needs of a variety of libraries, would be costly to buy (because it would be costly to design) and costly to operate. A good compromise would be a modular system for which the purchaser could select certain modules and easily modify others. A truly modular system is also expensive to design, but as we have mentioned earlier, the modular approach is desirable for other reasons. We have tried to follow this approach as much as possible, and we suspect that if another library were to acquire our programs, it could use the bibliographic modules almost unchanged; the holdings modules might require minor changes, and the

ordering and accounting modules would need major changes. This situation, of course, reflects the influence of the MARC project on the library world.

The idea of contracting with some other organization to design and develop the system might be attractive—if the money were available and there were no other way to obtain the staff with the necessary qualifications. We have never seriously considered this alternative. It tends to remove the designers from the close contact with the eventual users which we feel is important. It requires that every last detail of the system specifications be put in writing to avoid misunderstandings, and this in turn inhibits the implementers from making minor modifications to the specifications, modifications which may result in a substantial saving in development or operational cost.

Staff and Organization

One of the reasons that the Northwestern automation effort has been relatively successful is that we have had good access to a small group of people who would have to use the system—we could get agreement from them about what the system should include and what it should not. It is much easier to please 5 users than it is to please 500, or even 50.

For the staff of such a project, it is necessary to find either librarians with an interest in computers and a willingness to learn more, or computer experts with an interest in libraries or text processing and a willingness to learn more. It does not work, and we can say this from experience, to assign a program to print catalog cards to a person who has spent his whole career writing COBOL programs to do payroll tasks. The frustrations encountered in trying to locate the right people are only exceeded by the frustrations in trying to do the project without them.

Our design and development staff has been small and cohesive, and able to make and implement decisions quickly. It has had full administrative support. It is organized in a staff, rather than line, capacity, with all administrative, managerial, and operational tasks assigned to other departments in the library. We have made special efforts to maintain good relations with other members of the library staff and to secure their participation in the design. They have been encouraged to take full responsibility for the operation of modules of the system at the earliest possible time, and have done so.

Cost Advantages of the Total System

So far, we have really not provided much justification for the possible cost advantages of a single-institution in-house system as compared with participation in a network. The comparison with network operation is difficult to make. There is only one network for which cost data are available,

and that is OCLC. The real problem, however, is that the services are not comparable, so we must either attempt to isolate the part of our costs related only to the production of catalog cards, or else estimate what OCLC might charge for producing purchase orders, serial check-in, and so forth. Using the first basis, we find that our costs are approximately equal to those of OCLC. We feel that the balance will shift in our favor when additional services are included.

Why should this be true? Primarily because there is much less redundancy in the data associated with the additional services than there is in bibliographic information. Libraries are willing to accept bibliographic records created elsewhere, but they are not likely to be able to use order or circulation records of another institution. By centralizing such records, there is little saving in storage costs and a considerable increase in communication costs. The possible savings in using a large computer rather than several smaller ones are often lost in higher overhead costs, both in the computer software and in personnel to operate it. When a library decides to make an on-line catalog available to its patrons, the communication costs will increase several times, and it may then prove more economical to maintain even bibliographic files locally.

As we said earlier, we feel that there is definitely a place for large networks in providing access to bibliographic information and for interlibrary loan purposes. However, we firmly believe that other types of information should be maintained on an individual basis by large libraries, or perhaps by small groups of small libraries.

REFERENCE

1. De Gennaro, Richard. "Library Automation: The Second Decade," *Journal of Library Automation* 8:3-4, March 1975.