

## The Application of Computers to Library Technical Processing

A 1967 WHITE HOUSE report, *Computers in Higher Education*, begins with an arresting statement: "After growing wildly for years, the field of computing now appears to be approaching its infancy."<sup>1</sup> Library automation has passed through similar throes, and we may be at the beginning of a period of new and significant development.

Several important milestones have already been reached. Computer experts, now facing the problem of structuring and maintaining complex files, and dealing with a wide span of graphic output characters, have begun to appreciate the data management complexities inherent in bibliographic data. We no longer hear from computer people that our problems are trivial. We, in turn, have realized that it is no longer possible to speak of one component or subsystem such as an acquisition system, in isolation from other technical processing functions. Automation has confirmed the integrity and unity of technical processing.

The economics of applying computers to library data processing has come as a rude shock to many administrators. The old idea that an automated system could be operated at a new lower cost than a manual system is dead, indeed.

---

Mr. Veaner is Assistant Director for Bibliographic Operations, Stanford University Libraries. This paper was prepared for "New Dimensions in Acquisitions," an American Library Association preconference held in Atlantic City, New Jersey, June 19-20, 1969.

One now needs to plan future budgets in terms of cost avoidance or improved library services.

### *The Large System: Maker or Solver of Problems?*

The choice between stand-alone equipment and procurement of services from a central facility is the first major decision in any automation endeavor. The small or medium-sized stand-alone device is attractive because one can fully dedicate it to a specific application. But as the user's sophistication and system requirements increase, he outgrows the smaller machine and soon finds that he must cast his lot with a larger facility in order to enjoy certain technical benefits and operational features not available on smaller devices. It is at this point that one must be prepared to give up some freedom in exchange for more computer power, and where the complexities of scale begin to compete with the economies of scale.

Software in the large system carries with it unforeseen problems that seem to crop up endlessly and affect the scope of many operations in unknown and unpredictable ways. Hardware manufacturers and software developers have already learned about this, much to their chagrin, especially with time-sharing. W. F. Miller, Associate Provost for Computing at Stanford, characterizes this facet of software thus: "The reward, and at the same time the retribution, of software is self-change."<sup>2</sup> The reward is the enormous increase in our power to do things; the retribution is

the unforeseeable perturbations which come back to haunt operations thought to be fully debugged and dependable.

Fifteen major hazards in the development of large multi-use systems have been enumerated in a paper by F. J. Corbató, of Project MAC at MIT.<sup>3</sup> The dangers cited include lack or inadequacy of documentation, failure to implement designs, overstaffing of the design team with its attendant communication and supervision problems (Corbató conceives of ten as a maximum number), overextension in time, the attempt to undertake more than one significant advance at a time, the assumption that a finish date can be predetermined, lack of essential hardware, geographic scattering of resources (people and equipment), too many maintenance people in the systems programs.

Yet, once in the grasp of an automated system, there is no turning back. Entering upon an automated system in any enterprise is practically an irreversible step. This is why reliability in automated systems is a factor of overwhelming importance for library operations. The thing about library operations is simply that they *must be operational*. Our users and our management demand facilities that work during all normal service hours, and sometimes beyond that.

With this critical background, I would now like to describe what I believe are useful and profitable computer applications to acquisitions and technical processing. I also wish to report in some detail Stanford's development work in automated technical processing, an effort supported by the Office of Education's Bureau of Research. (Contract OEG-1-7-071145-4428)

#### *Candidates for Library Automation*

First, it is clear that a significant number of libraries do not require and should not embark upon library automation programs; they should instead

participate in regional technical processing centers, operated either by a jurisdiction or a commercial organization. Typically, these libraries order and process mainly current English language imprints marketed in the book trade, and they buy multiple copies of the same title for branch libraries. NILINET (New England Library Information Network), The Ohio College Library Center, and the Colorado Academic Library Book Processing Center are examples of service agencies for libraries which should not individually undertake automation, because their local operations are too small in scale. In the aggregate, the scale is sufficient to support the personnel and machine overhead demanded by computerized operations. These new centers may soon supplant in-house technical processing operations. While it is not clear at this time that technical processing will disappear altogether in the small and medium-sized library, it will certainly be radically altered in the near future. It is doubtful whether large university and research libraries can ever dispense with internal technical processing services, but even there, more widespread utilization of centrally produced data is likely to shrink the size of technical processing departments.

#### *Standardization*

Second, it is abundantly clear that the major impact of library automation will be felt in the area of bibliographic standardization. Page 1 of the final report, *The MARC Pilot Project*, contains a crucial observation: "The single most significant result of MARC has been the impetus to set standards."<sup>4</sup> Standardization efforts will be greatly aided by budgetary considerations. In every enterprise there is keen competition for the dollars needed to run every operation in the organization, and the dollars can be very determining. The increasing trend towards measuring perform-

ance effectiveness is already being felt in libraries. For example, Booze, Allen, and Hamilton is conducting a major management study for the Association of Research Libraries, a study whose aim is management improvement and increased adequacy of budget justification.

Two-thirds of a century ago, Herbert Putnam, then the Librarian of Congress, outlined the Library's proposed card distribution service. The purposes of distributing centrally produced bibliographic data are stated in clear and simple language:

to supply libraries with information of books which they do not possess . . . to enable them to avoid expense in the preparation for use of those which they do possess.

He goes on to quote the contemporary library press, pointing out that the two most costly factors of getting a book recorded in the catalog are the work of the cataloguer, the expert, and the work of the compositor or transcriber. It is worth the time and space to quote *in extenso* from this 1901 report:

Now, the interesting thing is that until now libraries have been, in effect, duplicating this entire expense—multiplying it, in fact, by each one undertaking to do the whole work individually for itself. There are thousands of books which are acquired by hundreds of libraries—exactly the same books, having the same titles, the same authors and contents, and subject to the same processes. But each library has been doing individually the whole work of cataloguing the copies received by it, putting out the whole expense. . . .

American instinct and habit revolt against multiplication of brain effort and outlay where a multiplication of results can be achieved by machinery. This appears to be a case where it may. Not every result, but results so great as to effect a prodigious saving to the libraries of this country. The Library of Congress cannot ignore the opportunity and the appeal. It is, as I have said, an opportunity unique,

presented to no other national library. For in the United States alone are the library interests active in cooperative effort, urgent to "standardize" forms, methods, and processes, and willing to make concession of individual preference and convenience in order to secure results of the greatest general benefit. . . .

A centralization of cataloguing work, with a corresponding centralization of bibliographic apparatus, has been for a quarter of a century an ambition of the librarians of the United States. It was a main purpose in the formation of the American Library Association in 1876. . . . The economies effected to the libraries of the country might alone justify the maintenance expenses of the Library of Congress even without a single direct service to scholarship. The country at large might indeed save great expense by purchasing a copy of a book merely to be catalogued at Washington, even if that copy should never go outside of the walls of the Library nor find a reader within it.

There are many difficulties of detail, and the whole project will fail unless there can be built up within the Library a comprehensive collection of books, and a corps of cataloguers and bibliographers adequate in number and representing in the highest degree (not merely in a usual degree, but in the highest degree) expert training and authoritative judgment. But the possible utilities are so great; they suggest so obvious, so concrete a return to the people of the United States for the money expended in the maintenance of this Library; and the service which they involve is so obviously appropriate a service for the National Library of the United States, that I communicate the project of this report as the most significant of our undertakings of this first year of the new century.<sup>5</sup>

Is it not time to realize Putnam's dream? Is not the day long gone when we can justify a host of alternatives to centrally produced bibliographic data? It is my conviction that *there will be no justifiable computer operations in libraries until we realize that the computer is an instrument of standardization, not a device whereby we perpetu-*



ate the alteration of bibliographic data produced by a central source. The idea of a local cataloger examining LC prepared data on a CRT terminal for editorial modification is economically unsupported and managerially unwise. Yet there are still libraries which, even in their manual systems, alter 100 percent of the card sets they receive from the Library of Congress. The cost of performing such chores of questionable necessity is likely to be intolerable in a computer environment. The aggregate of system resources spent on data management, Central Processing Unit cycles, Input/Output, channel time, and so forth, will be too great, and the computer's ability to do its own bookkeeping is relentless. Hence, it will be impossible to bury the cost of changing bibliographic data.

#### *Perfectionism: Friend or Enemy?*

Perfectionism and permanence are two interdependent fallacies of modern bibliography. Perfectionism is based upon the idea that the librarian is creating a permanent record. Unfortunately, even in the manual system this has never been true. Even the Library of Congress' Official Catalog changes substantially, the amount varying according to the age of the record and ranging from an estimated rate of about 5 percent in the first year of a record's life to an aggregate of about 40 percent of all records after thirty years.<sup>6</sup> To prepare for future network applications it is essential that changes in the nation's bibliographic records be kept as consistent as possible, and this is achievable only by rigorous adherence to data centrally produced at a national bibliographic center, *even if those data contain errors when issued*. At least in this way, errors will be consistent, and they can be corrected later in a consistent way by the central distribution service.

The abandonment of perfectionism in bibliography needs to be established as

a goal. (It need not be employed as an excuse for deliberate carelessness.) The future of a computerized update mechanism for bibliographic records should encourage libraries to make rapid inroads on arrearages now, without waiting until every bibliographic problem is solved with a score of 100. We may be approaching the first time in history when we can afford a few errors.

Another facet of the technical processing problem has been a traditional view, fortunately not shared by everyone, that all books are equal and must receive equal technical processing. Just as we need to establish time priorities for processing, we need to make intellectual judgments concerning the quality, amount, and depth of bibliographic treatment to be given publications. Because such decisions are no longer irreversible, there is an opportunity for expedited processing and the preparation for public use of more books.

The idea of self-sufficiency in resources, *i.e.*, exhaustive collection building, is dead. Self-sufficiency is a laudable heritage of the protestant ethic, needed in eras of slower communication. High "budget visibility" of book funds has aided in the development of a variety of cooperative acquisition programs, based on the idea of building national rather than purely local resources. The costs of technical processing have not been so visible, but they are coming into sharper focus all the time. Costs now hidden in personnel and overhead are likely to be surfaced by the application of computer technology.

#### *Applications to Technical Processing*

There are two categories of work which can be substantially aided by computer applications.

First, we have a great mix of data-management activities: keyboarding, updating, deleting, sorting, printing, distributing, calculating, merging, filing—

dull and boring activities. It is difficult to recruit and train, and almost impossible to retain staff for this kind of work. Rapid staff growth needed to accommodate recent large increases in publication output makes for very difficult management problems: supervision troubles, lack of employee satisfaction, high turnover, poor communication within the organization, and difficulty of following standard procedures.

Searching is the second category of technical-processing work which can be materially aided by computer applications. Stanford has applied substantial effort to develop a capability for on-line searching, because we believe that in this area there can be a future payoff in public service when computer costs come down to the point where public terminals can be justified. Meantime, the paucity and rigidity of access points for searching card catalogs and in process files makes searching for technical processing frustrating and much less productive than it should be.

#### *Development in On-Line Search and Retrieval*

Stanford has developed a search facility by which many users can search the same or different files simultaneously, just as one can do with the card catalog, but with these additional features which no card catalog can offer: (1) users can interact or negotiate with the files expanding or contracting searches at will, even saving them for future reference if desired (saved searches can be run against new MARC tapes); (2) users can carry out coordinate searches; and (3) users can access any of several central files anywhere that there is a terminal. System response time can be kept reasonably short—a few seconds—because an inverted file structure searches index files which point to data base entries. In other words, no serial searching is employed.

With the aid of a grant from the Li-

brary and Information Science Branch of the Office of Education's Bureau of Research, Stanford is developing an on-line bibliographic control system dubbed SPIRES: Stanford Public Information REtrieval System. Acquisition and cataloging are the two chief areas of current research and application. However, it is well to mention that interactive searching is practical only on fairly large computer systems.<sup>7</sup>

#### *Requirements for On-Line Retrieval*

An on-line search facility requires several things: (1) a large computer facility (Stanford's system uses a partition of an IBM 360/67); (2) software with built-in feedback features to facilitate system modification; (3) a large data base; (4) very large storage facilities; (5) a means of rapidly displaying search results, preferably by visual terminals; (6) a wideband communication network to transmit processed data to remote stations.

SPIRES software already provides its users with the capability of communicating their satisfaction or dissatisfaction to the system's designers. A large data base is obtainable through MARC, and an even larger one will be available through RECON (RETrospective CONversion), if the full RECON Project materializes. Really large storage facilities—enough to store even a million records locally—must await future, more economical devices, perhaps photodigital stores or laser beam recorders, such as the UNICON (Unidensity Coherent Light Recording). In terms of screen capacity, character set, and writing speed, visual displays are still quite costly and not yet truly satisfactory for bibliographic data. A wideband communication network means coaxial cable, which costs about \$1.50 per installed foot.

#### *Need for Collaborative Development*

One of the first automation lessons librarians learned was the astronomical

communications gap between computer people and librarians. We conclude that this gap must be reduced nearly to zero—if the automation of library technical processing is to succeed. Three groups need to be brought together: the librarian, the computer expert, and the information scientist. The library can't do this job alone; in fact, none of these people acting alone is likely to succeed.

### *Expendable Equipment?*

For many years we have been in an era of expendable software. In fact, software investment commonly runs two to three times the cost of hardware. It is not unreasonable to expect that the future is likely to bring us quickly to an era of expendable hardware. The American economy already provides an outstanding precedent: the automobile is a piece of expendable hardware. Basically, hardware and software are no different. Some hardware—terminals in particular—may have a useful lifetime of only one or two years.

### *The Future of Books and Bibliographic Files*

About ten years ago, the book began to come under some concerted attack as an inefficient means of storing and transmitting information. Despite the controversy surrounding this issue, one fact stands out: the book is still the cheapest to produce, the simplest and easiest-to-use device for information storage and retrieval. A 1969 article on the impact of the computer on publishing begins: "The most efficient information storage medium, by far, is the least sophisticated to produce—the printed page."<sup>8</sup> In 1968, consumers spent \$4 billion for broadcasting services and another \$4 billion for consumer electronic products. Yet in the same year, the value of printed and published goods totaled \$22 billion, of which \$12 billion was for newspapers, books and periodi-

cals—substantially more than the sum spent for nonprint communication media.

Looking ahead some distance in the future, I see a long life for the book. I see the retention of paper as a major medium of communicating data for acquisition processing; booksellers in developing countries (and even in some advanced ones) will continue to issue paper invoices, some written in a familiar illegible scrawl. I foresee continued lack of rationalization of the processing unit in book procurement (invoices, purchase orders, checks, etc.), the factor responsible for the great amount of effort we face in distributing and redistributing data over media in reconciling our budget accounts and invoice documents. I do not see vast on-line bibliographic files in our major research libraries, except possibly at the Library of Congress and maybe at a few regional bibliographic service centers. Rather I see the possibility that our entire concept of file organization will be restructured. A highly simplified model, which I hasten to add I have not costed, might look something like this: closest to the library user might be on-line access to current items in process and to those permanently held items known to be heavily in demand. Somewhat further away—in terms of ease of search and retrieval—might be book catalogs with relatively brief and simple entries supplemented by full bibliographic data in microfilm cartridges permanently arranged by sequence members in the form of a register. Such a master file could be centrally produced by computer output microfilm printers as a by-product of the MARC and RECON projects. This register would require virtually no updating—all the organization and maintenance would be confined to the book catalogs or on-line files, which would act as indexes to it. Even the book catalogs might be organized far differently from our present ones; some



might be topical, others chronological. A microform register would be extremely cheap to duplicate and distribute. Hard-copy of full bibliographic data could be easily obtained by conventional reader/printers.

Before any idealized file structure or service like this can be implemented, we need to know much more about our users than we now do. It is unlikely that we will reach this future by postulating

great, all embracing "total system designs," either conceived in ignorance of user requirements, or representing someone's pet idea. The necessary research, experimentation and implementation should be dominated by two principles: (1) construction and testing of development models capable of self-change through user feedback, and (2) implementation of major functional modules one step at a time.

## REFERENCES

1. *Computers in Higher Education*. Report of the President's Science Advisory Committee. Washington, The White House, February 1967, p. 1.
2. W. F. Miller, "Economic and Operating Realities of Present-Day Hardware and Software in Library Applications," *Proceedings of the Stanford Conference on Collaborative Library Systems Development* (Stanford: Stanford University Libraries, 1969), p. 145. (I am indebted to Professor Miller's paper for the section on large systems.)
3. F. J. Corbató, *Sensitive Issues in the Design of Multi-use Systems* (Cambridge: Massachusetts Institute of Technology, Project MAC, 1968), p. 17.
4. *The MARC Pilot Project, Final Report* (Washington: Library of Congress, 1968), p. 1.
5. *Report of the Librarian of Congress for the Fiscal Year Ending June 30, 1901* (Washington: Government Printing Office, 1901), p. 28-37.
6. *Conversion of Retrospective Catalog Records to Machine Readable Form* (Washington: Library of Congress, 1968), p. 144-147.
7. A searching guide, the *SPIRES Reference Manual*, has been issued to explain the search facility; this publication is available on request. Further information on Stanford's automation program is contained in the *Proceedings of the Stanford Conference on Collaborative Library Systems Development* and in the first issue of the *Project BALLOTS Quarterly Newsletter*, which was distributed at the Atlantic City ALA Conference. The *Proceedings* are available for \$7.00 (prepaid) from the Office of the Financial Manager, Stanford University Libraries, Stanford, Calif. 94305. Overseas orders are priced at \$8.00 and must be prepaid in U.S. dollars.
8. Ron Schneiderman, "Printers Seek Electronic Image," *Electronic News*, vol. 14 (April 21, 1969), 5.