

R.J. BRAITHWAITE
Assistant Director for Network Services
Library Automation Systems
University of Toronto
Toronto, Ontario

Automation of the Catalog: The Transition from Cards to Computers

I WONDERED IF MY being asked to speak at this conference on “Problems and Failures in Library Automation” was perhaps a 2-sided compliment, as the University of Toronto Library Automation System (UTLAS) is one of the more successful projects in library automation in North America. However, in any automation project, no matter how successful, there are always some problems which can serve as lessons to others.

For instance, in the early days of the UTLAS project, we attempted to produce an on-line circulation system with complete standby facilities at every terminal. Each station was to have a badge reader, a punched-card reader, a keyboard printer and a paper tape reader/punch. However, the project was abandoned after a pilot operating phase. Its requirements were far ahead of technology; it was a case of too much, too early.

Each component of this system was from a different manufacturer, and since this was before the advent of microtechnology, the entire assembly occupied a large desk. (Today most terminal requirements, along with a microcomputer, can be packaged in a desk calculator case.) The planned procedure was that the system would first read the patron badge and then the book card, and would produce a date due slip on the keyboard printer—much the same as today’s systems. If the system went down, or the communications line failed—a not-infrequent occurrence—the paper tape punch would take over and record

the transaction for feeding into the system later. Unfortunately, the effect on the library was at times devastating. The sound of the paper tape punch was more like a machine gun than a piece of library equipment, and it would start up without warning. Other problems related to the reliability of the central processor, disk storage, terminals and software. The decision was therefore made to abandon the project after the pilot phase rather than implement a system which had low reliability and a mean time between failures of about two hours.

With that out of the way, UTLAS pursued other developments which have been much more successful and serve as the foundation of current services. The UTLAS project began as part of the bibliographic system of the University of Toronto Library (UTL), and its other early endeavors included experiments with MARC and non-MARC holdings formats. Among these latter was the ONULP project, for which the University of Toronto Library prepared the initial collections for five new universities being established in Ontario. The cataloging data were converted to machine-readable form and a printed booklist was prepared. At least one of the five libraries is still using the data and format today; however, because the project had its own format which was not compatible with MARC, UTLAS did not develop it further.

Another non-MARC format was adopted for most of the holdings of UTL (about 1.25 million records). This presented some interesting problems when it became necessary to merge these records into a composite data base with records from several other sources, including some which were MARC-based. We kept in close touch with the Library of Congress (LC) system during the design of LC MARC and further developments were based on this format, including a service (which is now being terminated) for searching MARC tapes and producing unit cards or copies of records on magnetic tape.

At about this time UTLAS became a separate unit of the library and a new director with experience in the computer field was recruited. Work was begun on the development of an on-line system for inputting MARC-like records; this became known as LODES (Library On-Line Data Entry System). This system was further developed into LODES II where the entry process became an editing process, creating a system which could maintain a data base as well as enter in it. Further development of this system included revision of the format to bring it more in line with the LC and Canadian MARC formats, ultimately producing the system which is known today as CATSS (Catalogue Support System).

One of the biggest problems which affected UTL stemmed from these early pioneering efforts. The library held data in each of the

formats of the early on-line systems, and used a non-MARC format for a batch system to collect a large portion of the data base. When the time came to make use of these data, the various formats had to be correlated with the current standards, and conversion programs were written to salvage as much data as possible from the earlier input. The cost of conversion has to be carefully weighed against the benefits. We have found that only with large collections of data is it economically justifiable to attempt automatic conversion from one format to another. In some cases, we have advised libraries to discard the results of a previous project and to start again. The second time around they are much wiser in setting up objectives and much more realistic about what can be achieved, so all is not lost. It is very unlikely that such an abandoned project will be written up as a paper, since it would require a very enlightened administrator to recognize the merits of "washing the dirty linen in public," even if the staff concerned were masochistic enough to want to invite public comment on their apparent incompetence or ineptitude.

With the UTL data base, it took an entire year for a team of two to three programmers, with considerable support from the systems librarian, to integrate the data from the formats into a data base from which a microform index could be generated to serve as an alternative to the card catalog. Even with all that energy expended, a considerable amount of work had to be done to clean up some of the data and coding problems that had occurred over the years; approximately 50,000 records have been edited to remove errors or to improve entries, principally with regard to filing rules. Some attempts were made to correct by program the lack of coding in the earlier data. For instance, honorifics were not coded fully; to correct this, the filing key generators were programmed to look for honorifics in the names and automatically generate the proper form. Unfortunately, computers are rather blind and obedient slaves. They were programmed to look for "Sir," "Lord," "Lady," etc., and they took the instructions literally; hence, John Sirica of Watergate fame became "Ica, John, Sir," and was condemned to obscurity as a misfiled entry. Tests are unlikely to point out such problems, for if a programmer foresaw their occurrence, he or she could have avoided them in the first place.

A related problem requiring careful attention is enforcement of the use of coding standards by the programming technicians, since the results of their efforts are not obvious until they become part of the final product. Then they may be only too obvious to the public, or more devastatingly, to the library procedures. One microcatalog had a large number of musical scores filed under "uartets" and "uintets" as the

programmer was confused about which code indicated a traced entry. He wrongly chose the code which indicated the number of nonfiling characters. In a second case, a coder was mistaken about the use of hyphens and slashes in the holdings statements for serials. One indicated a continuous run of holdings in the serial while the other indicated a number of items bound together. When the data base was used to generate the item records for a circulation system, it was found that in some cases there was one record for a whole shelf of bound serials, while in other cases there were many records for only one item.

This instance raises another issue involved in planning for library automation. Many projects have been conceived to streamline a particular aspect of the manual system without first looking at the fundamentals of the problem. For instance, the card catalog should not be an end in itself. Its complicated structure was developed around the restrictions of a manual guide to a library collection. It was a vast improvement over the book or sheaf catalog; it is much easier to add cards to a drawer which is full by redistributing some to adjacent drawers than to soak pages overnight and repaste all the entries when a sheaf catalog page becomes congested. However, this does not mean that a card catalog is instantly up to date.

Data processing experts may encounter considerable problems with library data. They do not conform to nice, fixed record layouts but can vary in content and size as much as the books themselves. A title can vary from a single character to an entire essay. We had one "short title" for a pamphlet that filled twenty-two lines of a catalog card. This caused the formatting program to loop after completing the title just when it had to start a continuation card. It had formatted 22,000 cards before the operator killed the job because the computer was asking for the fifth output tape and only 4 tapes had been assigned to this type of run.

Another library had a contents note which exceeded 8000 characters and caused all sorts of problems. The largest record we have handled was over 44,000 bytes and prompted a reevaluation of all the size restrictions in our handling programs. They are now set to the system limit of 64,000. With this point in mind, careful consideration should be given to the claims made by some systems designers that allotments of twenty-five characters for author or thirty characters for title provide adequate clarification.

Another aspect of libraries that worries computer analysts is the size of the files. Many schemes have been tested on 1000 records or 10,000 records which would nevertheless collapse under 1 million records or more. Pauline Atherton's very interesting subject access project

at Syracuse¹ generated index terms for 2000 monograph records and produced one of the largest inverted files that SDC had ever seen. This indicates that if we are to break out of the straitjacket of LC subject headings, system designers will have some complex problems to solve.

So far discussion has focused on the data base in the migration from cards to computers. This is the essential first step. The machine-readable data base is the basis of any alternative to the card catalog, whether it be printed, microform or on-line. Several users are currently looking at all three forms.

The printed book catalog has been used by several libraries as an alternative to cards, but as the booklists got larger, the printing and binding costs became excessive and this format had to be replaced. One of the most interesting of these book catalogs has been a project undertaken by a school library to use PRECIS indexing; this library is now experimenting with microform. A number of colleges, universities and public libraries are now regularly receiving microform catalogs. We have produced the first "infant" provincial union catalog for British Columbia. This is intended to become a complete union catalog of all holdings for the province within the next few years.

There is one other major problem associated with data bases concerning our authority facility, which will become operational this year: the availability of machine-readable source files. The subject authority file will be available soon (assuming LC resolves its problems with the issuance of the eighth edition), but there are no plans to provide a conversion of existing names in the foreseeable future. We see this as a major obstacle to the implementation of AACR II, which will have its greatest impact on the form of names. Therefore, we are proposing as an interim solution that some of our major user groups enter the cross-references from their card authority files; we would then link these automatically to the bibliographic records and generate skeletal heading records for all the names without cross-references. This plan is still in the discussion phase.

Another area fraught with problems might be termed "expectations of the users." This refers to the strongly held belief that the computer is a god and that its priests can do everything in no time at all. In reality, computer systems have considerable weaknesses that can be catastrophic if not compensated for in the system design. There are scores of examples of the fallibility of computer systems, such as astronomical electricity bills, and the unresponsiveness of charge account systems in correcting an error. These are the fault of the system designers who have overlooked the checking which in a manual operation would be done automatically by a clerk. However, it is often con-

venient to blame the computer. A blatant example of this was provided by a senior airlines official in defending the company's policies with regard to their charter class air fares. The system was set up so that there was a maximum of four charter class seats on any plane from Victoria to Vancouver; this was being challenged by a family of five who wished to travel charter class on the first leg of their planned holiday. The official stated that "the computer would not let them." However, not all problems can be blamed on the computer. One householder in England was so worried about his electricity bill that he turned off the power at the main switch, causing all the street lights on his road to go out.

Computer systems are not gods. They are very fallible and must be designed to survive all sorts of terrible events. They may not continue running when the lights go out, but they must not lose the data recorded up to the time of the power failure. So it is a good idea to ask the system designer (if you have your own system) or the supplier (if you buy service or a turnkey system): What happens if I have a head crash or am unable to read part of the disk? May I continue or are the transactions after that lost? Do I have to return to the last security save?

Computer systems are expensive and time-consuming to develop. When preparing a manual system, all the unusual events can be omitted and the technicians can simply ask about exceptions later. In a computer program, however, every possible eventuality must be covered in advance or the whole process may fail and have to be repeated. Systems must be designed with the computer in mind rather than as an attempt to mechanize the manual process. The limitations of the card catalog may disappear, but new, computer-dependent limitations may replace them.

On the road from cards to computers, problems undoubtedly will be encountered, but many libraries have started and are making progress. Some libraries will go only part of the way to find that printed or microform catalogs will be adequate for their needs. However, many will go all the way and provide a full on-line catalog, such as the University of California which is planning a 600-terminal inquiry system to replace the card catalog.

Finally, I must congratulate the UTL staff in their forward approach to automation. They may recognize many of the problems I relate here as theirs. This is not because they are worse than others, but because they have made more progress than others. If they had no problems, it would be because they weren't progressing.

REFERENCE

1. Atherton, Pauline. *Books are for Use: Final Report of the Subject Access Project to the Council on Library Resources* (Research Study No. 4). Syracuse, N.Y., School of Information Studies, Syracuse University, 1978.