

Automation in the Federal Archives Division of the Public Archives of Canada

by **ELDON FROST**

This paper has three purposes: to discuss the experience of the Federal Archives Division of the Public Archives of Canada in developing its computer system; to describe very briefly the current status of the system; and to comment on the applicability of the division's experience with automation in other archives.¹

EDP development in the division got underway in the early 1970s. The Public Archives played a large part in the development by federal government departments of the "Records Control and Index" or RECODEX computer programme. The RECODEX system was designed to control information at the file level. It sorted information by file number and date; and, by a system of Key Word Out of Context (or KWOC), the retrieval of subjects from file titles was possible. In the early 1970s, the RECODEX programme, including the indexing methodology which was enhanced within the division, was the basis of divisional EDP effort.

By the mid-1970s, the division held 17,400 linear metres of archival material. Eighty-five per cent of it had been identified for permanent retention, given a series-level inventory description, entered on a file list (usually handwritten), and placed in an archival container. It was estimated in 1975 that putting the files into the computer system would generate some 1.2 million computer records at the file level and that annual accretions of 60,000 records (1000 metres per year) could be expected. A master record in machine-readable form was to be created for each central registry archival file in the division, thus gaining bibliographic control over holdings. The emphasis was obviously on capturing descriptive file information for the purpose of information retrieval. The extension of this system, and its amalgamation with others, was to permit subject access to information within all of the PAC's Archives Branch records.

Reality set in rather quickly. For example, considerable resources were devoted to RECODEX for the records of the Indian Affairs Branch of the Department of Indian Affairs. However, the division has been barely able to keep even with records accessions for this branch. The Key Word Out of Context approach, which proved to be labour intensive, was used sparingly elsewhere, especially in recent years. By 1977 it was recognized that a different approach was needed. Clearly there were missing elements in planning the use of the computer in the early 1970s. Failure to recognize the extent of the enormous and complex task of keywording file titles — the work of inputting information, verification, and editing of output — was the main inadequacy. One other dimension of the problem which had not been accurately forecast was the growth of holdings. By 1978 the annual accession rate had tripled. In that year, it was estimated that the archival records documenting the spectacular growth of the federal government in World War II and the postwar years would soon begin to be transferred to the Public

1 The Federal Archives Division of the PAC is responsible for acquiring, controlling, and making available unpublished textual records of the federal government which have archival value.

Archives. At the same time, new directions in research necessitated the preservation of new types of records such as operational files, which documented not only policy but its implementation, and case files. Under these new demands, divisional holdings were expected to double within five years. As a matter of fact, holdings increased by 138 per cent in the ten-year period ending in March 1985; the extent under "preliminary control" fell in the same period from 85 per cent to 65 per cent. Given the revised growth expectation, continuing public service demands, and dim prospect of obtaining new resources for any of this work, it was concluded in 1976-77 that the division would be very busy establishing basic physical control and a rudimentary level of intellectual control over these new records. It could do little bibliographic work for the purpose of information retrieval.

Could automation help with these new realities? The staff of the division began to tackle this question by listing the functions it performed. How did the division acquire, control, preserve, and make available our records? What descriptions of procedures were available? What forms were produced? What reports were used? What purpose did they serve? Rather grandiosely, the staff not only wanted to improve productivity by examining workflow and procedures but also wanted to develop and implement an information system which would ensure that, as records were acquired, processed, and made available to the public, every element of information generated through these processes would be captured. Despite such optimism, help was needed to prepare a clear statement of requirements from which decisions could be made about whether to proceed and, if so, in what direction. Approval was forthcoming for a feasibility study to be undertaken by contract. It was to provide four reports: a description of the existing system; a description of alternative proposed systems and recommendation of an optimum system; the conceptual design of a recommended system with implementation plans; and, finally, a cost benefit analysis of a recommended system.

The recommended solution was an interactive system run by a service bureau (the Public Archives had no computer) at a staggering cost of \$1.5 million. There were limited developmental and no extra operational funds in the department to support EDP applications. Treasury Board Secretariat, for its part, wished to see proposed development in all eight divisions of the Archives Branch. It suggested a phased approach. The National Library advised that quoted in-house development costs were probably low. The Public Archives suggested a slower approach, gradually phasing in development. A proposal was finally approved which drastically scaled down the division's initial requirements to a listing of those priorities which had been developed eighteen months previously, namely, automating those processes which would indicate what was in the division's holdings, where those records were located, and their access status. The establishment of control over the initial processing of the records was also included. Other aspects — text processing and circulation control — could be added on later; file level information would continue to be addressed by RECODEX. The proposed system became a "document control" system known as "FEDDOCS" (Federal Archives Division Document Control System).

Another major decision remained to be made. Following revised requirements and during the conceptual design of the recommended system, it was learned that an off-the-shelf package existed which could meet the revised requirements and had the potential to meet other delayed requirements. The contractor was directed to investigate the costs and benefits of this software package, called MINISIS. (MINISIS is a relational database management system developed by the International Development Research Centre to

run on mini-computers.) Lower development costs, on-line entry, modify and report sequences, along with proven information retrieval capabilities made MINISIS an attractive option. It appeared that MINISIS would also allow us to do the bibliographic control work at the file level which had been done with RECODEX, but this remained to be proven. On the negative side, processing costs with a service bureau would be very expensive. Following considerable discussion and deliberation, approval was given for a one-year pilot project using MINISIS. One year following completion of acceptance testing, approval was given to continue with the system on an operational basis. The division moved subsequently to deal with high service bureau costs by obtaining the use of a dedicated computer shared by other Public Archives users through a facilities management arrangement.

The aim of the FEDDOCS design is to provide a comprehensive and integrated system for managing the division's archival holdings. New accessions of records are registered upon entry into the division and monitored by the system through the selection stage to their final disposition as archival records or their destruction, if they have no archival value. Volumes of records which have been "processed" (in other words, selected and described at the series level) may be stored within any of the five storage areas occupied by the division in the National Capital area or in regional records centres in Halifax, Winnipeg, Edmonton, and Vancouver. (Some thirty-two kilometres of shelf space are occupied.) The records' exact location by floor, bay, and shelf is identified within FEDDOCS, and the availability of each volume under the *Access to Information Act* or other access authority is also tabulated. Access and location information is provided for unprocessed accessions of records; in addition to this information, the size, type, and polarity of the division's microfilm holdings are captured in the system. Following early testing, it was learned that FEDDOCS is capable of producing the finding aid reports previously generated by RECODEX, and significant improvements in the present system have been developed. File level information in the computer, which is of course linked to its archival container, is now integrated and manipulated by standard routines rather than a variety of programmes as was the case with RECODEX; the file record structure is sufficiently flexible to accommodate nineteenth-century file docket, early twentieth-century numerical file series, modern block numeric systems, and such oddities as land patent data for Western Canadian homestead grants; finally, a special file number sorting scheme developed in the division permits the sorting of files and their parts or volumes for virtually any filing scheme ever used by the federal government, regardless of alpha-numeric or special characters. Some 40,000 file level records have been successfully converted from RECODEX into FEDDOCS. The comparative ease with which this conversion has been accomplished makes it apparent that such conversions could also be made from information in records management support systems in government departments to FEDDOCS as the records described within those systems are transferred to the division. The system produces reports and provides on-line verification and query for the 1750 unprocessed accessions, 132,000 volumes of processed records, 80,000 microfilm reels (16,000 unique reels), and 225,000 item or file-level entries contained within it. Some one million records will be contained in FEDDOCS by the end of March 1986. Not only is FEDDOCS one of the largest MINISIS databases in the world, but the number of fields and variety of record types make it one of the most complex.

Much remains to be done and, because of limited resources, continued development will be gradual. Circulation control over the records within the Public Archives and externally for microfilm through the interlibrary loan system are requirements that were

identified at an earlier stage but left in abeyance. It should be possible to add these enhancements to FEDDOCS without much difficulty, and this is currently underway. Inclusion in the system of textual information from the division's 130 record group inventories remains to be undertaken, as does investigation and full use of the information retrieval capabilities of MINISIS whether at the level of the inventory or the file. The prospect of extending the use of the computer in archival description is at once exciting and challenging. Can archivists, and perhaps researchers, depart from the traditional paper inventory and finding aid to use an index to information in archival records which they themselves build on-line? The construction of a file-level index has been a laborious process which, in order to achieve consistency in terminology, was limited to particular record groups and even to series within record groups. Can the computer itself make meaningful linkages between subjects upon instructions from knowledgeable users? What degree of standardization in terminology is required in order to make possible an on-line search for information across record groups? Limited resources will require that choices be made in determining what descriptive information about archival records should first be put into the system. Is the researcher's interest best served by permitting him/her to retrieve information on archival holdings at the "overview" level provided by the record group inventory or at the file level? Is the archivist's interest inimical with that of the researcher? With whom, both within and outside of the Public Archives, must the division be prepared to share information in the computer? What kind of information should be shared, and how? MINISIS, the software currently running FEDDOCS, does not permit the full MARC format, although it is possible to convert data from MINISIS into a database containing records in MARC format (or the reverse) by means of an International Standards Organization conversion process. Sharing computer information with users or other clients by thus being able to load it into selected bibliographic databases is viewed by the division as being an acceptable mode of communicating information about our holdings of textual archival records of the federal government. It is possible that as the division continues to explore other potential uses of the computer, whether in monitoring work, describing records, or in general administration, that other software and hardware will be required to supplement or even replace that currently running FEDDOCS.

In the federal government, Treasury Board defines standards for the management and control of EDP projects which are, in conjunction with departmental standards, to be used as criteria in judging project performance. These standards are described within the steps of project initiation, feasibility study, general design, detailed design, implementation, and post-installation. In the division's case, following this process for the first major development was time well spent. It is useful at the initial stages when the rigorous thought required by the process forces one to choose carefully. An archives — like any other institution contemplating automation — needs at the "project initiation" stage to identify clearly what the project is and to establish appropriate controls for it. What problems *now* need to be addressed? If it is a problem now, will it be so in five years? What was identified as a major issue for the division in the mid-1970s, dealing with information at the file level, was not a priority several years later. One should be prepared to recognize that not all the problems can be identified at the project initiation stage and to acknowledge that none of the solutions are known. One must be prepared to rethink issues at this stage and during the feasibility stage when solutions are being devised. It is important to sort out *clearly* what the problems are even if it appears to take an extraordinary amount of time and effort. Some experience can undoubtedly be gained from other archival applications.

However, archives are not all the same, nor are their priorities the same. Automation is expensive. It is easy to divert resources into automating functions which are relatively simple but which do not address those urgent needs. It is worth remembering that the development stage involves only the initial cost of automation. The institution must be prepared for additional expenditures. After all, why develop something if it is not important enough to maintain, or expend resources inputting data that is not used because it does not address important problems?

The experience of the Federal Archives Division indicates that if the initial planning is done thoroughly for the major work, subsequent, more limited development can proceed in a less rigid manner. It is possible to buy or develop solutions to meet specialized situations if one has already, by means of thorough planning, identified what all of these components are. Since it is unlikely that most archives will have the resources to implement initially the "complete solution," identification of components that can be integrated as modules at later stages is important. Automating archival processes and information about archival records is in its infancy. Because of this, and the difficulty of initially identifying the myriad of records already within a large archives or those which will be received in future, a flexible, generalized software which can be adapted to new situations and types of records is recommended. This choice will also enable those within the archives who know how the organization functions to be trained more easily and made responsible for running the application. External expertise, which is always expensive, can be used more sparingly for major modifications. Further development ought, of course, to be undertaken under the umbrella of a comprehensive plan. The plan should be updated regularly because of the phenomenal rate of change in automated options for both hardware and software.

The cost of automation is easily underestimated, and the benefits tend to be oversold by contractors, particularly those from outside the organization who have a vested interest in continued development. Automating archival processes and retrieving information from archival records are not simply accomplished. It is sometimes easy to forget that one must maintain information that one has captured, and that the more complex the information captured the more time-consuming and elaborate is the maintenance required. Is the expertise available to do this work? It is wise to force the consultant to address these matters by maintaining throughout the development process a questioning, critical attitude. The archives concerned must, after all, live with and pay for the results.

One ought to identify clear start-up dates from the manual to the computer system. It is difficult, and not very useful, to attempt to maintain both systems over a period of time. Ideally, new procedures reflecting changing work flows ought to be ready for issue at the same time as system implementation; one should be prepared to revise these after installation as new methods of streamlining operations become apparent. It is important to remember that staff requirements change considerably with the installation of large database systems. The archives will need access to programme analysts and probably have a data administrator on staff. It is safe to state that all computer systems, large or small, need to be tended and nourished, and continuing operational funds will be required to do this. Moreover, productivity benefits of automation in the sense of more work being done by fewer people may take time to materialize, or never appear at all as they are swallowed up by ever-increasing workloads. Rather than being surprised by this eventuality, one ought to expect it and, in more pessimistic moments, recall that a good automation application *does* enable one to do certain tasks better.

Finally, it is to be noted that while archives cannot ignore automation, automation is not an end in itself. Automation will not correct problems that are imbedded in an institution's structure or management process. A good automation development process will force an institution to be critical of itself, both as to what it is doing and why it is doing it in a certain manner, just as it is forcing archivists at national and international levels to come to grips with standards for describing what they do and how they do it. The more critical and efficient archivists are, the more they will accomplish and the better they will do it. Enabling archivists to do more with less will be the most visible and lasting effect of automation in archives.

Abstracts of Theses in Archival Studies at the University of British Columbia

by **TERRY EASTWOOD**

The following is a list, with abstracts, of the authors and titles of the first ten theses completed by graduates of the Master of Archival Studies Program at the University of British Columbia. The MAS Program is jointly administered by the School of Library, Archival and Information Studies and the Department of History. All the theses listed below were completed in partial fulfilment of the requirements for the degree of Master of Archival Studies. Theses are listed in the order in which they were presented for oral examination.

All MAS theses will eventually be available on microfiche through the National Library of Canada. Currently, it is about two years from the time a thesis is submitted to the National Library until it appears on microfiche. Future reports in *Archivaria* of completed MAS theses may include a list of titles on microfiche. It is hoped that interested parties may be kept up to date by regular reports on theses which have been completed and reproduced each year.

Laura Coles, "The Decline of Documentary Publishing in Canadian Archives, 1865-1984." April 1984.

Documentary publishing by Canadian archives has declined noticeably. Today, few archives actively publish diaries, journals, or other primary documents. An analysis of the history of English language documentary publishing by Canadian archives and historical societies suggests some reasons for this decline.

The rise and decline of documentary publishing by archives is traced for each region of the country in part by compiling a select list of publications of the sort being investigated. For comparative purposes, documentary publications by historical societies, prime among them, the Champlain Society and the Hudson's Bay Record Society, are also investigated.

It is suggested that the reasons for the decline of documentary publishing are varied. The changing nature of the economics of publishing and the technology available to