THE PUBLIC ARCHIVES OF CANADA'S EXPERIENCE IN ESTABLISHING A MACHINE READABLE ARCHIVES

BY

MICHAEL E. CARROLL

PUBLIC ARCHIVES OF CANADA

Introduction

The major portion of this paper was prepared in 1974 for an International Council on Archives Conference on Archives and Automation. This paper is an updated version of the ICA presentation and is intended to highlight the experience of the Public Archives of Canada in establishing a machine readable archives for computer-oriented records. General references are provided to the decisions that have to be made, those that have been made, and the status of their implementation.

Background

The Public Archives is not the first in North America to establish a machine readable archives program. The machine readable archives of the Public Archives was established one and a half years ago. The first machine readable archives in North America was established over fifteen years ago in the university environment. The main impetus to this activity has remained in the university environment. The institutions which have emphasized the collection of paper and more traditional forms of archives have lagged considerably in embracing computer-oriented records as a field of interest.

The objective of the Public Archives is to acquire,

This paper was prepared for the Annual Conference of Senior Archivists, April, 1975.

organize and preserve historical material relating to the history of Canada, and to provide a records management and microfilming advisory service to government departments and agencies. The objectives of the Machine Readable Archives is to provide archival services for machine readable records of long term value produced by the Federal Government and those of national significance produced by the private sector. In terms of sub-objectives this means: 1) ensuring that machine readable records of long term value are acquired; 2) ensuring the preservation of machine readable records of long term value; and, 3) ensuring that reference services are provided to meet user demands. The balance of my presentation will cover the proposed or actual activities for the three sub-objectives.

It should be mentioned that embodied in the objective, sub-objectives and activities of the Machine Readable Archives is the policy assumption that the medium of the record is relevant only from the viewpoint of form and not substance. In effect, this means the application of existing archival principles and administrative practices with adaptations where necessary. This approach will be demonstrated in specific instances in the presentation.

Acquisition:

Criteria of Archival Value

Consistent with our policy assumption, the criteria of archival value have not changed because the medium has changed. The application of the traditional criteria, however, has created a number of new administrative practices which require a different perspective. The best way to demonstrate these points is to review our policy, Acquisition Criteria (see Appendix 1).

Selection Process:

Government Sector

The Public Archives has an existing system whereby files in departments are inventoried, schedules are established and approved by the Dominion Archivist, the schedules are applied by the Departments, and once the operational life of the files has ended, the dormant and historical files are transferred to the Public Archives. The Machine Readable Archives will use this system to acquire government records of archival value. (Appendix 2).

The only change in this system will be a telescoping of the time frame from the point that files of long term value are identified to their eventual transfer to the Historical Branch. In the past once files were scheduled the Public Archives would wait until they had no operational value to departments before the files were

transferred. This procedure has changed with machine readable records. Since tapes are so vulnerable, are so easily copied, and many problems can be encountered in interpreting and collecting the supporting documentation, once machine readable files are identified as archival we request, as soon as possible, that a copy be transferred. If this were not done it is possible that some files might be unusable once they were transferred.

There are a number of difficulties in implementing such a system. The difficulties are not in the system per se, but the fact that after fifteen years of neglect we are implementing it for machine readable records. We have to demonstrate that we know what we are doing, and that we should be doing it at all. It has also been discovered that without available archival services in this field the departments we service have established, in a few instances, their own archives-like organizations.

Our approach has been to use our past record of success in the paper and microfilm fields, and to work carefully toward establishing our credibility in the data processing environment and as a machine readable archives. A few key elements to this approach are the utilization of our established contacts in departments, the records managers, and the emphasis on contacting not the data processing people, the individuals who provide a service, but the users, the individuals who are the owners of the machine readable records. This approach has proven successful where it has been applied.

Private Sector

We do not and cannot impose a system on the private sector similar to the one we are using in the Federal Government. There are no legal requirements for the private sector to send a copy of their files to the Public Archives. From available information, however, we find that research work in this field requires considerable amounts of monies, and, generally, this comes from the Federal Government, through one means or another. Consequently, we have proposed to a few government agencies providing such sponsorship that one copy of the machine readable file and supporting documentation be sent to the Public Archives as a condition of the funding. We have argued that such an approach will ensure that the research produced will not be lost and that, if our recommended guidelines are followed, the quality of the data will be assured from a machine readable perspective. To date we have not had a negative or positive response to this proposal.

Medium:

As archivists we are concerned with the preservation of the medium retaining the information. When the medium deteriorates, so presumably does part or all

of the information of archival value. The principal medium at this time for storing machine readable records is magnetic tape. Magnetic tape has a number of characteristics which are desirable and undesirable from an archival viewpoint. It is very vulnerable to destruction if improperly handled or stored. It is reusable, or in negative terms, it can be easily erased. Magnetic tape files can be easily and cheaply duplicated relative to files on other media. Information on magnetic tape is machine readable; this also means that it is machine dependent.

We conducted a study to find an alternative medium to tape that was less vulnerable to destruction, machine readable and if possible less machine dependent than tape. The conclusion reached was that computer input microfilm (CIM) was a possible alternative, but there were few financially viable companies to provide the product at relatively inexpensive rates. As a consequence we have chosen to use magnetic tape as the medium of storage for the next five years.

The choice of magentic tape as a retention medium has resulted in the development of what we call an archiving system for machine readable archives. The first element of this system is the retention of at least two copies of any file in separate physical locations and a computer output microfilm (COM) copy of the file. Thus, if problems arise with one file, we have a backup; if these problems are duplicated on the backup files, we have the COM backup, and can manually input the lost data.

The second element of this system is to store and handle our tape under ideal conditions to maximise the possibility of data retention. To this end, and also with the hope that our procedures will be used by agencies, we are working with one of our standards agencies, the Canadian Government Specifications Board, to develop standards for the care, handling and storage of magnetic tape. This is being developed and will cover the following areas: 1) quality of medium; 2) preparation of tape for storage; 3) recording on tape; 4) tape delivery; 5) personnel security; 6) storage environment; 7) preventive maintenance; and 8) physical control of tape.

The third element of this system is to acquire our own computer facilities to ensure proper handling and processing of our files. The only alternative is to use commercial or government service bureaus. We feel, however, that because of the stringent quality control procedures required to ensure that data is not lost, our needs cannot be met by a service bureau geared to normal data processing operations. It is interesting to note that with a projected average annual volume of 280 machine readable reels that an in-house facility is more economical over a ten year period than any of the alternatives

considered (see Appendix 3).

The fourth element of this system is to convert all incoming files to one standard archives format. This approach will reduce our costs, make our files relatively hardware and software independent, and facilitate the conversion of our entire holdings in about ten years time to remain current with the technological developments and, thus, maximize the possibility that the files can be used.

This archiving system offers no guarantee that 100% of the data will be retained over a ten year period. We can only talk in terms of minimizing the possibility of losing data permanently and maximizing the possibility of recovery should loss of data occur.

This system is being studied at this time to obtain the necessary approval and funding.

Accessioning:

When files are transferred we require that a back-up copy be retained at the source until we can produce a working copy. With this latter copy we verify that the appropriate file and necessary documentation has been transferred, and, if there are any problems, followup is done. All files received are eventually converted to a standard archives format. If we have any problems with these conversions, we use computer consultants on staff with us for this purpose.

Supporting documentation is critical to machine readable files for without it the files can be virtually useless. To increase the possibility that all the necessary documentation is transferred we have prepared a checklist or guideline of documentation requirements that we use in contacting the transferring agency. Initially we thought the transferring agency should complete a documentation package using this guideline. Subsequent evaluation of this approach indicated that this was too onerous a demand, and now a staff archivist works in the agency contacting all the necessary people to ensure the documentation is completed.

Processing

There is little comparison between the processing that must be done for machine readable files and other more traditional archival media. Since machine readable archives are generally statistical files, the processing is so oriented. The main purpose of the processing is to ensure that the machine readable files correspond to the original data capturing document which was made machine readable, and if there are variations, where these variations are, and can they be corrected.

Reference:

Description of Holdings

In Canada, there are few, if any archival institutions, which use librarians and their techniques to produce an intellectual control of their archives. In particular, no archives uses cataloguing, but do employ cataloguing-like procedures. It is the intention of the Machine Readable Archives, however, to use cataloguing procedures. These procedures are in the process of being developed by a sub-committee of the American Library Association (ALA). From my understanding of the processes of the ALA it will be a few years before the necessary cataloguing rules and formats are developed. We have an immediate need and, consequently, have created a task force consisting of librarians, computer specialists and archivists to study whatever has been developed by the ALA, and produce a cataloguing system for machine readable archives. We are one year from achieving this objective.

We have given no consideration to the other means of describing our holdings since we feel that most questions concerning a file could be answered from the accompanying documentation. This is the supporting documentation necessary to read the file and to interpret it in a meaningful fashion.

Access to Machine Readable Archives

Probably one of the more interesting aspects of machine readable archives is how researchers will be given access to them. In Canada, the main orientation of providing access to archives is to have researchers come to where they are retained. Over the past number of years this policy has changed by providing microfilm copies on a loan basis to researchers through established institutions. Although growing, this policy applies to a very small, though important, segment of archives.

With machine readable archives, researchers require access to computers, and, in some instances, data processing staff. Archives should not be expected to provide either except on a cost recovery basis. To avoid this issue, at least for the moment, the Machine Readable Archives, like all others in the field, intends to provide a tape copying service, and for larger files, a data element extracting service. The latter refers to providing a copy of only selected elements of a large file. Tape copies will be tailored to meet most researchers' needs in the computer area. With these copies we will provide a copy of the supporting documentation in a published form. With the data element extracting service this means more information than is required, but we will use this approach as a stopgap measure. Thus, the nature of machine readable archives for the foreseeable future

almost determines that archives will be going out to the researcher rather than the reverse as in the past.

This is not to imply the democratization of archives. The users of machine readable archives must have considerable funding and access to computer systems. This marks such a researcher as belonging to a small elite group.

One area that we are convinced is inevitable, but we have yet to investigate is machine readable documentation and control. Our catalogue entries could be made machine readable from the start. This would allow researchers to use this entry for citation purposes. It would also allow us at some future point in time to have a machine readable catalogue that can be easily disseminated or published. Machine readable documentation, on the other hand, will require considerably more time and resources. We see as inevitable the transmission of machine readable data over telecommunications systems. It would follow that the supporting documentation must also be machine readable.

The Machine Readable Archives would like to make available as many of its files as possible for research. The question of restricting public access is critical, however, given the general character of its holdings and those of similar institutions. A great number of machine readable files are based on an individual's or an organization's response to a data gathering instrument, such as a questionnaire. Some may have been collected in confidence; others may contain confidential information. The question of confidentiality has been the major concern when considering potential access restrictions. The policy that has been decided on can be briefly reviewed. (See Appendix 4).

Staffing

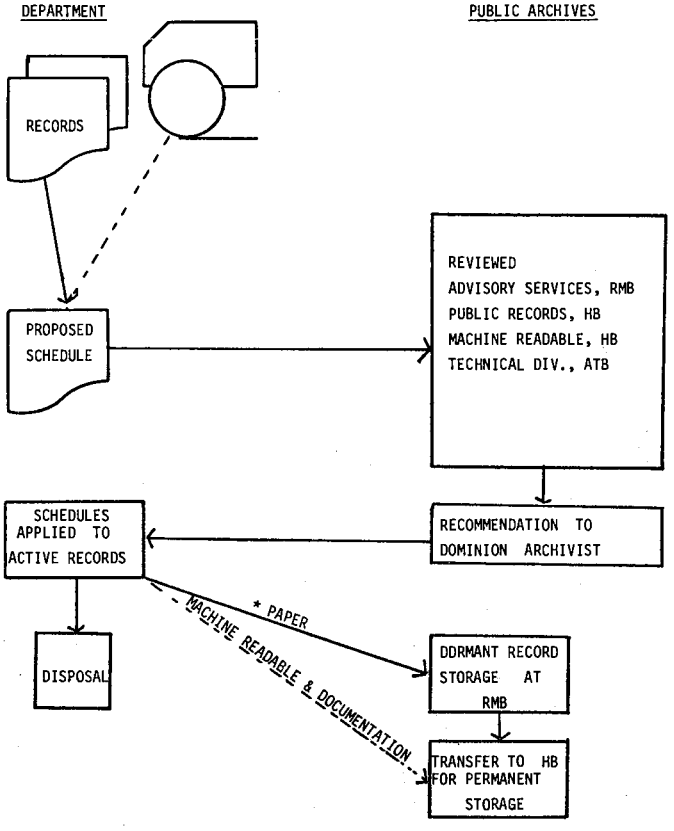
The question of staffing is critical to ensuring that our objective is met. Ideally, we would want a computer-archival expert but such a combination is rare, if not unavailable. Until such expertise is developed we have divided the two disciplines. On the archival side we are looking for an individual who has a statistical background, some training in one of the social sciences, and with some programming experience in one of the social science programming packages. The computer expert, on the other hand, should have a broad experience in most of the data processing areas, but with special emphasis on software packages and documentation. We hope that by having a team with this expertise our objective will be achieved.

Conclusion:

As stated in the introduction this paper was intended to sketch the development of a machine readable archives. We have tried to learn from the experience of other machine readable archives and adapt what we considered desirable to our environment. This presentation is not a list of our achievements but more a status report on an organization that is at least a year from being fully operational.

APPENDIX 2

SCHEDULING SYSTEM OF
CANADIAN FEDERAL GOVERNMENT



* DIRECT TRANSFER FROM DEPARTMENT TO HB IS NOT THE RULE BUT IS POSSIBLE

Appendix 1

ACQUISITION CRITERIA FOR THE MACHINE READABLE ARCHIVES

1. An objective of the Machine Readable Archives is to collect machine readable records of long term value of the Federal Government and those of national significance produced by the private sector.

2. The Machine Readable Archives will also accept offers of machine readable records of long term value.

3. Criteria for Long Term Value

Records (whether magnetic tape, questionnaires, etc.) being appraised for inclusion in the Machine Readable Archives shall be considered to have long term value if one or more of the following conditions are satisfied:

- -they were or may be used to support the formulation of policy
- -they were created for a study which might be considered seminal either because of the nature of the study or the type of analysis used
- -they were created for a study conducted by an individual or group of individuals renowned in that field
- -they contain information of a non-housekeeping nature which is not of limited value for further analysis or reanalysis.

Where it is difficult to determine whether or not one or more of these criteria are met, the advice of outside consultants and/or the originator(s) of the file(s) should be sought.

4. Selection of Files

(a) Retention of Questionnaires

In general it is desirable but not always possible to dispose of questionnaires. When there is doubt, they should be kept.

Questionnaires for which the corresponding machine readable file is not held by the Machine Readable Archives should be retained.

If the file is held by the Archives the questionnaires will be destroyed if all of the following conditions are met:

-all the usable questions or variables are on the

original machine readable file or merged with the original file to create a new one (excluded from this provision is the instance when the questions or variables permit unique identification of individuals, such as name, address, or S.I. number); and

- -there are no open-ended questions; and
- -the usefulness of the questionnaires in making corrections in the corresponding machine readable file(s) is exhausted

(b) Retention of Tape Files

The computer environment produces a considerable number of different files which revolve around a study. There are generally four categories: master, summary, transactional and working files. As a rule, we want that file which is the product of all files creating it. This usually means the master file.

(c) Retention of Extract Files

From a master file a system (SPSS, Data-Text, etc.) file may have been created which not only may contain less information by virtue of its being an extract or sample, but also may contain recoded variables. This latter file may well have been the one with which the researcher actually worked, and therefore be of long term value. In this case it will be advisable to keep both the master and system file.

(d) Retention of Ongoing Master Files

Because ongoing files may be added to or updated hourly, monthly, etc., the question arises as to what file copy of an ongoing file of long term value should be retained by the Machine Readable Archives. The decision will depend almost entirely on the subject matter. The two types of ongoing files likely to be encountered are:

- -a file in which none of the previous records or data elements has been destroyed or modified but which has new records added to it continually e.g. the meteorological file of the Atmospheric Environment Services
- -a file in which some or all of the previous records or data elements are being destroyed or modified continually, while new records may or may not be added, e.g. Unemployment Insurance Commission file; this case presents the most difficulty as records are being destroyed or altered.

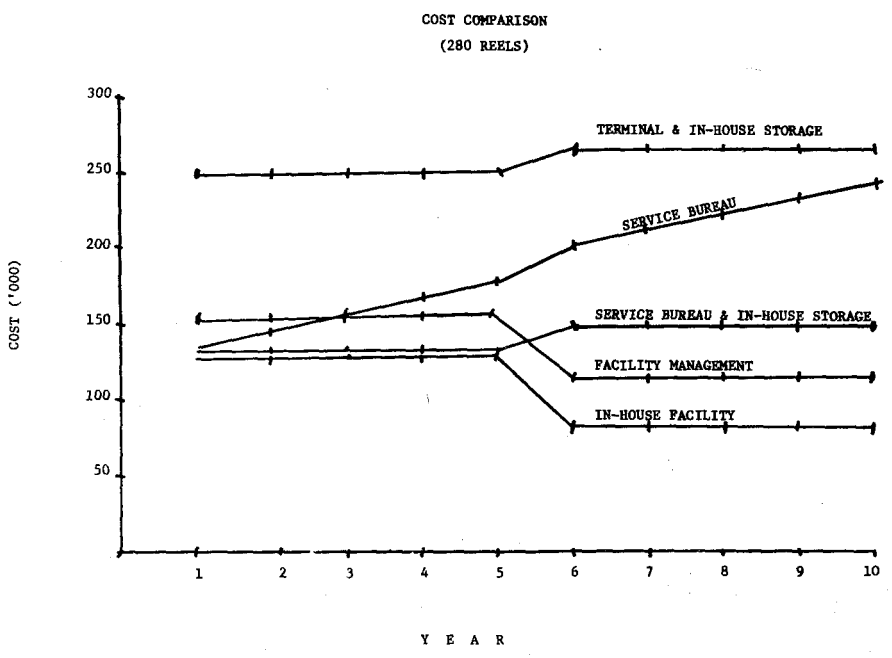
(e) Retention of Original Files

All machine readable files of long term value transferred shall be retained in the condition in which they were received. This is to satisfy any future requests for the original file and to support the new file created from the original. In the latter instance, the creation of the new file may have been improperly handled and the retention of the original will permit recovery.

5. Documentation

A diligent effort should be made to ensure that all hard copy and/or machine readable documentation necessary to read the machine readable files of long term value and interpret meaningfully their contents is/are transferred. All the documentation obtained from the transferring agency will be retained to support policy 4 (e).

APPENDIX 3



Appendix 4

ACCESS TO FILES IN THE MACHINE READABLE ARCHIVES

It is the general policy of the Machine Readable Archives to provide as quickly as possible, at a nominal cost, copies or extracts of its files together with supporting documentation, subject to the following conditions:

1. Under no circumstances will files (regardless of medium) originally transferred to the Machine Readable Archives be sold, loaned or given out; only copies or extracts of such files will be released.

2. Copies or extracts of files (regardless of medium) will not be made available if such access contravenes the conditions or regulations determined by transferring agency or individual.

3. Copies or extracts of files (regardless of medium) containing information gathered through questionnaires, interviews etc. in which respondents were assured of confidentiality, when they are released by the Archives, will contain as many data elements/variables as possible (or a subset thereof) without jeopardizing the anonymity of individual respondents, interviewees etc. whether these be companies, organizations or individuals.

4. The Archives reserves the right not to release any file copy or extract if it considers such an action to be or possibly to be, a breach of public trust, detrimental to public welfare or injurious to one or more individuals.

5. The purchaser of file copies or extracts sold to him by the Machine Readable Archives will agree that these are for his exclusive use and that the information contained therein must not be sold or otherwise transferred.

6. The purchaser will also agree to credit the Machine Readable Archives and the principal investigators in publications which use the files and to indicate that neither bears any responsibility for the analyses or interpretations which appear in such publications.