ACCESS LEGISLATION AND THE MACHINE READABLE ARCHIVES

LA LEGISLATION SUR L'ACCESSIBILITE A L'INFORMATION ET LES ARCHIVES ORDINOLINGUES

Winston Gomes
Public Archives of Canada
395 Wellington Street
Ottawa, Ontario K1A 0N3

ABSTRACT

The introduction of the computer has altered, for all time, man's ability to generate, use and disseminate information. While contributing to better management, storage and retrieval of information, the computer is also responsible for a tremendous explosion in the amount and variety of data which is now available.

Federal government agencies and departments generate, use and disseminate large amounts of information as they pursue their mandate in providing services to both the Crown and the public. Such information possesses considerable value, the extent of which is a function both to the intended and final use of the data.

The new Access to Information and Privacy legislation will govern the proper and safe use of information in the custody of the Crown. The Public Archives of Canada, as an agency of the Crown will play a major role in ensuring the proper management and archival retention of such information.

The paper will present a detailed overview of the PAC/MRA and how their mandate ties in with the requirements of ATIP legislation in making machine readable data, appraised as having archival value, available for research purposes.

RESUME

L'utilisation de l'ordinateur a modifié, de façon drastique, la capacité de l'homme à produire, à utiliser et à diffuser l'information. Bien qu'il contribue à une gestion, à un stockage et à un repérage de l'information nettement améliorés, l'ordinateur est aussi à l'origine d'une fabuleuse explosion dans la quantité et la variété des données maintenant disponibles. Les diverses agences du gouvernement fédéral produisent, utilisent et diffusent une énorme quantité d'information dans l'accomplissement de leur mandat et dans la provision de services à la Couronne et au public. Cette information peut avoir une valeur considérable selon l'usage qu'on veut en faire et qu'on en fait.

La nouvelle législation sur l'accès à l'information (LSAI) et sur la protection des renseignements personnels régira l'usage judicieux et sûr de l'information sous la garde de l'Etat. Les Archives Publiques du Canada, en tant qu'agence de la Couronne, joueront un rôle fondamental en assurant la gestion adéquate et la conservation archivistique de cette information.

L'auteur présentera une vue d'ensemble exhaustive des APC/AO et montrera comment leur mandat rejoint les exigences de la nouvelle législation LSAI en rendant disponibles, pour fins de recherche, les données ordinolingues ayant une valeur archivistique.

MACHINE READABLE ARCHIVES

Although the term "Machine Readable Archives" might be less familiar to you than the terms "Data Archives" or "Data Library", we will essentially be referring to the same collection of activities which are to be found within such programmes.

Data archives and libraries are creatures of academia and as such exist mainly for obtaining and providing access to information in response to the needs of an immediate research community. Such information is therefore acquired to fill more specific demands and might be of internal generation or acquired from outside sources. By their very nature, that of responding to the needs of the particular academic community, data archives and libraries are imbued with special characteristics as befits, their particular environment.

Given that they suffer generally from lack of funds, a situation not entirely unique but perhaps more acute in the academic world, their collection habits are necessarily constrained, and their acquisition criteria restricted to the particular needs of the funding organization. Their collections, then, may tend not only to be narrow in scope but, over time, may suffer from a lack of proper care necessary for long-term preservation. Although one can call to mind a small number of specialized data libraries/data archives with extensive holdings within specific subject areas, the differences between them and the more traditional archival repositories are still significant.

The concept of a data archive or library as an integral part of the more traditional archives is relatively new, having begun in Canada in 1973. Such an addition spans the information retention media from printed paper to micrographics and now to the machine readable medium.

Machine readable is defined as a medium which is suitable for use as an entry or exit document in a computer environment and includes punched cards, punch paper tape, optical scanning devices, computer magnetic tapes and discs, floppy discs etc.

The Machine Readable Archives (MRA) is a Division of the Archives Branch of the Public Archives of Canada. It was created in the fall of 1973 in response to a realization that no programmes existed at the federal level of government concerned with the proper archival retention of records being created through the use of electronic data processing. The MRA was formed with a mandate" ... to appraise, acquire, control, conserve, and make available for research those historically valuable machine readable records produced in the public sector and those of national significance produced in the private sector, and to assist in the development and promotion of data archival techniques in other organizations". The "total archives" concept was now complete.

In pursuing its mandate over the years, the Division has had to come to grips with many problems related to the archival retention of machine readable records. In order to give you an appreciation of some of those problems, I shall briefly discuss the functions of the MRA.

(i) ACQUISITION

The acquisition function culminates in the successful transfer of an archival record from a donor agency to the PAC. Included in and as a preliminary to the acquisition of a record, is the process of appraising the record to establish the existence and extent of archival value. (Archival value here is defined as the historical value and/or national significance mentioned in the mandate statement). Herein lies the first hurdle called (appraisal) criteria.

Given that the MRA is an integral part of an agency of the Crown, its first priority is the consideration of all records of the Crown that might have archival value. Traditionally, such records as might be considered for archival retention must in some way possess informational, legal and/or evidential value. Machine readable records go beyond the traditional in that the records are quite often of value without in any way shedding light on the organization that created it, its mandate, or even its functions. But where the value of the content of the record might not present a problem as such, the form in which the record is retained most assuredly could. Witness the differences between the simple cross-sectional survey data file, the large complex data base which requires a data base management system for proper maintenance and control, the textual record data base, the mapping system and co-ordinate data base, and the bibliographic data base, each in turn a very different machine readable record, managed and retained in a very different environment, and each in turn requiring a different approach to its acquisition and documentation. I will briefly discuss three of the types mentioned above.

Survey data are the simplest form that a machine readable data file can take. The records which comprise the file are complete in themselves in that they usually start with identifying information, followed by data usually in a fixed format mode. The data, generally stored in card image form of eighty or more characters, and the variables in each record characteristically in the same place throughout the file, lend themselves to manipulation and analysis with the use of generalized statistical packages.

Data stored in a data base environment and managed through a data base management system require an entirely different approach to documentation and access. The heirarchical structure

characteristic of data bases precludes the fixed format of simpler files, requiring instead location codes with their propensity for repeating structures.

Textual and bibliographic data bases are familiar to most of us as their existence and use are predominant in library and similar environments. Their dependence on precise terminology and keywords for access and retrieval can, in the main, be a bane or boon to the user. The movement towards the use of "phrase searching", "segmentation", "truncation", "ranking methods" and "dictionary" words as utilized in the QL system, however, enhances the search capabilities though making special demands on the system documentation requirements.

The three broad types enunciated above are sufficiently different in character to illustrate the complexities of documentation in the first place, and of applying some uniform approach to appraisal of the information for its archival value. It might be said, and correctly so, that the appraisal process should not be hampered by the form in which the data are kept, as it is the value of the information which should be assessed. But access and retrieval are dependent upon the structure of the data, and its retention medium therefore must be considered as part of the acquisition process.

II. CONTROL

"Control" is the substitute term for "arrangement", a more familiar word in the world of archives, but appropriate nonetheless. For it is "arrangement", more precisely the description of the information in the machine readable record and their inter-relationship, that is of vital concern in this function. Where documentation tells you what to find and where, it is the verification of that documentation which ultimately determines the final disposition of the record. As in the acquisition function, the process is made complex and diverse given the many types of data records and the formats and retention media utilized.

Cataloguing of machine readable records is done utilizing the Anglo American Cataloguing Rules (AACR II) as revised for machine readable data files by the American Library Association. The exceptions to be made represent the physical differences between this medium and that of printed material since the intellectuel content would be identical in either form.

III. CONSERVATION

Conservation is the watch-word in all archival repositories and rivals in importance the initial generation of the data themselves. Present technology still relies on the magnetic tape as the medium for long-term storage of information.

However, magnetic tape is not a particularly stable medium, being susceptible to a number of factors such as magnetic fields, fluctuations in temperature, mishandling etc., even when the most stringent precautions are taken in its storage, handling and conservation. The MRA follows a strict conservation procedure designed to prevent the loss of any data. A tape and file management system keeps track of all archival acquisition prompting for annual rewinding and cleaning of each magnetic tape and the recopying of the information to a newer retention medium every five years. But increased acquisition will gradually make an already expensive conservation programme prohibitive, necessitating the need for a more reliable and permanent medium for the long-term storage of information. Videodisc promises to be the answer in that the laser encoding technique will lessen the chances of information loss over time while providing a retention medium of potentially greater life-span, and considerably cheaper to recopy.

IV. PUBLIC SERVICE

It has often been said that providing service to the public is the raison d'etre of archives, and providing access to its holdings is a very important function of the MRA. The Division offers a range of services which include the provision of a complete copy of a record with a copy of its documentation, the provision of an extract of one or more records with appropriate documentation, and the provision of limited data analysis. All of these services are provided at a small cost to the researcher.

Divisional staff participate in information sessions, formal and informal presentations and activities, and contribute to the promotion of data archival techniques in other organizations.

The Division also has a publication programme designed to promote the maximum use possible of the services it offers. There is a brochure which describes the Division and its activities; a publication which gives an overview of its operations and procedures; a publication on drug related data files - the first in what is expected to be a series on special collections of interest; and a catalogue of holdings which will be updated periodically to reflect new acquisitions.

The preceding was a brief description of the MRA, its mandate, functions and procedures and how it approaches some of the problems it faces in pursuing its programmes. The following will give you some appreciation of what the Division is doing with respect to current legislation and the effect on its holdings.

INVENTORYING

MACHINE READABLE ARCHIVES

To determine the extent of holdings of machine readable records in the federal government, an inventory was undertaken in 1976 by the Public Archives of Canada. Results of the inventory indicated a conservative figure of about three thousand (3000) data files in sixty-seven departments. With that information an acquisition strategy was formulated. This saw archivists going into departments and, with the assistance of records management and EDP personnel, attempting to identify and apply retention schedules to all machine readable records.

More recently, the Division has been working closely with the Treasury Board Task Force on the implementation of Access to Information and Privacy legislation. Staff of the Division, as members of that Task Force, have been instrumental in drafting policy guidelines and directives for the management of EDP records.

The Task Force is also responsible for the preparation of a register of all classes of information held by federal departments and agencies, one of the requirements of new ATI legislation. To assemble this data, divisional staff have been responsible for drafting a new form which is being used to inventory the EDP records held by those departments and agencies to which the new legislation applies. As integral members of Task Force teams, they brief EDP, records management, and information management personnel on the requirements of the legislation, particularly as they pertain to EDP records, and assist those departments and agencies in identifying and reporting their EDP systems, records, etc.

The involvement of members of the MRA on such teams will enable the Divison to assess properly the degree of activity within the federal government and to identify those activities which have a high potential for producing valuable EDP records. The information which will be gathered to satisfy the requirements of the register will be an important step in supplementing the initial inventory of 1976 and forming the basis of an acquisition strategy which will be of benefit to the research community and general public at large while complementing the holdings of the other areas of the PAC.

CONCLUSION

The work of the Machine Readable Archives is just beginning. Being in existence for such a short time, staff are still in the process of developing liaison with the research community and coming to grips with its various needs. As a way of properly defining that community and determining the needs, the Division is in the process of planning a user survey. Such a survey will be directed to current and potential users of our services and will try to establish the need for such services,

the extent of activity which might justify an acquisiton strategy outside the federal government, and the need of the research community to have a forum to exchange ideas and information in the primary and secondary use of information acquired, arranged and disseminated by data archives and data libraries alike. It is our hope to sponsor such a forum in the form of a conference, possibly during the year 1983, but the results of the survey will better define the format and timing of such an event.

As users of such information, you yourselves might be included in the user survey. It is our hope that with your assistance we can better serve you in the future by learning what your needs are today.