

ARCHIVAL APPRAISAL: *Theory and Practice*

Proceedings of the Joint Meeting of the
Association of British Columbia Archivists and the
Northwest Archivists Association
Vancouver, April 26 - 28, 1990

EDITOR
CHRISTOPHER HIVES
UNIVERSITY OF BRITISH COLUMBIA ARCHIVES

ARCHIVES ASSOCIATION OF BRITISH COLUMBIA
VANCOUVER, BRITISH COLUMBIA
1990

Canadian Cataloguing in Publication Data

Main entry under title:

Archival Appraisal

1. Archival resources - Management - Congresses.
2. Archives - Administration - Congresses. I.
Hives, Christopher L. (Christopher Lawrence),
1957- II. Association of British Columbia
Archivists. III Northwest Archivists (Association)
CD950.A73 1990 o25.17'14 C90-091764-4

APPRAISAL AND ACQUISITION OF ELECTRONIC DATA PROCESSING RECORDS

by

Walter J. Meyer zu Erpen *

Integrated Approach to Records Management and Archival Appraisal

To begin, archival appraisal must include the values of records to the agency which created them. If your archives has responsibility for the overall management of your organization's informational resources, including their office use, off-site storage, and archival preservation, then you must first ensure that the records you are appraising are retained for the appropriate retention period.

The question you should ask is: Has the electronic data which you are appraising been retained a sufficient period of time to meet the administrative, operational, audit, fiscal, and legal requirements for which it was created?

I pose this question as a records manager, based upon the fact that the majority of the information created in today's society will have more value to the creating agency in the days, months or years immediately following its creation than it will have to any other individual or organization at any subsequent time.

By intergrated approach, I mean also that archivists must know the context in which records were created and must be able to appraise all formats of a given record series at the same

* NOTE: Opinions expressed are personal and do not represent the view of the British Columbia Government.

time. When electronic data is involved, records scheduling and archival appraisal is best conducted during systems development.

As we are all aware, the volume of information being generated today in all formats is so great that society can only afford to preserve for posterity a very small percentage of the total. Archival literature suggests that five percent or less should be retained from a large corporation or government. Given the excessive duplication, I feel that the amount which society can afford to preserve, process and make available to the research public is more likely one to two percent.

The task of the archivist is to identify those records which have ongoing archival value to one's corporate sponsor and/or to society generally. Regardless of the media or format being appraised, the task when dealing with modern archives is the equivalent of searching for a needle in a haystack or a somewhat less polite expression, picking fly droppings out of pepper.

Increasingly, our focus must be on those archival records which have sufficient value to justify their preservation. Sufficient value must often be defined in terms of the national or provincial significance of the records. For the majority of archives, the days are gone when we could afford to preserve everything that related to a given subject area or everything which survived in a special media format.

Today, at the British Columbia Archives and Records Service, we are already passing judgement on records created in 1989! The process of natural selection which used to leave the provincial archivists picking over what had survived is no longer. Now we

are able to pass judgement on virtually 100% of the information created by the provincial government!!!

If you are an archivist working for a large organization, you know that the task of archival appraisal is enormous. Virtually all records have value to someone, but very little of what has potential value can be saved. As someone who has done a considerable amount of genealogical and local history research, I realize that some of the most boring modern administrative records will one day be of value to the genealogist in attempting to piece together the fabric of the life of his or her ancestors. That does not, however, mean that society can afford to preserve those records.

Before I turn to the appraisal issues specific to electronic data, which you are here to hear about, I must state that I am no expert on the appraisal of electronic records and that the institution I work for has in fact had little experience in the appraisal of such records. My most concentrated experience with electronic records was in 1983 at the National Archives of Canada when I did my archival studies practicum in what was then called the Machine Readable Archives Division. My practicum dealt primarily with processing and description. Secondly, what I have to say relates chiefly to the appraisal of electronic information stored on magnetic media such as tapes, cassettes, and disk packs, rather than the newer laser technology.

APPRAISAL AND ACQUISITION CONSIDERATIONS:

What, then, are some of the considerations that impact upon the appraisal and acquisition of electronic records?

What type of data?

There are many types of preliminary and transactional processing data which are the equivalent of the drafts of a memo. Obviously, such data should be destroyed as soon as its usefulness has ceased.

Does Better or More Complete Data Exist?

Does the data which you are appraising exist in a better or more complete copy or location? Obviously, archives cannot afford to acquire electronic data where it is reasonable to expect that a more desirable copy will be available at some time in the future.

Often, one information system feeds data into another. Consequently, it is important to understand the interrelationships between information systems. For instance, in one appraisal, we established that the most important archival data contained in the Corrections Administrative Records Entry (CARE) System and the Probation Records System (PRS) is regularly down-loaded to the Provincial Case File (PCF) System. As a result, we recommended that all of the data in the first two systems be destroyed and that the data in the Provincial Case File System be preserved.

Does Data Exist in Summarized Format?

Another crucial question is whether the data has been adequately summarized in report format, whether electronic, paper or microform? Given the choice, my preference would be to preserve information in miniaturized human readable format such as computer output microfiche or a published statistical report because of its greater accessibility to researchers. By preserving a summary of the information, the long term storage and retrieval costs will be much smaller. If an adequate summary does exist, the archivist must still determine whether or not the electronic data has significant statistical value based upon the quantitative information which might be calculated from it. Obviously, a summary report does not cover everything.

Does Data Exist in Hardcopy Format?

While the benefits of automated retrieval and manipulation in electronic format are tremendous while data is being used very frequently, it may be more cost effective to store information for long-term research values on computer output microfiche and delete the electronic copy. For instance, the provincial land title registration process is being automated so that the titles to the property and homes we own is stored electronically. Once the title to a parcel of property changes hands, the old title is cancelled. After cancelled titles have been stored electronically for a number of years, it may be feasible to copy them to microfiche and have the computer system generate an electronic cross-reference index between the title and the

corresponding microfiche. The cancelled electronic titles could then be destroyed without loss of information. Given the choice, I would acquire the microfiche titles, assuming that adequate statistical reports about land transactions already existed in summary form. I would much rather deal with the problem of changing microfiche reader technology than software and hardware compatibility.

Is Data Readable?

Once you get to the point of appraising data located upon a specific set of magnetic tapes or disks, the first technical question you should ask is whether the data is still readable. If data is five or more years old and has not been stored in proper environmental conditions and regularly conserved, it is possible that the data on the tape or disk you have in front of you is no longer readable. The electronic imprint on a magnetic tape fades over time, but can be recopied to 100% its original intensity if no data has been completely lost.

Are Documentation Manuals and Code Books Available?

Are the code books and the system and user documentation originally used by the system operators and users still available? If not, the electronic data on the tapes you want to appraise may be useless. The code book is the manual which indicates the format of the tape, the length and storage format of each electronic record, and the address of each data element within each record. Without the code book, it is possible to reassemble some information about the data, but not all. Take,

for example, data relating to unemployment insurance claims. Assuming that you are able to determine the format and length of each record, you could produce a partial hardcopy printout, which is called a "dump." You could then determine the beginning and end of recognizable fields such as names, addresses, postal codes, phone numbers, gender, etc. However, you could do nothing to determine the beginning and end of the most valuable data elements which are stored in indecipherable codes. A series of Y's for YES and N's for NO are meaningless unless you know the question they are given in response to. Likewise, numeric codes from 0 through 9 assigned to employment history status are useless unless you know what they correspond to.

Is Special Software Required?

In appraising the data manipulated by a database or other software package, existence of the original software used with the data is an important consideration. If the data is software dependent and the software does not exist, the data is useless. Where the data is not software dependent, the data will probably be accessible through use of statistical software, such as the Statistical Package for the Social Sciences (SPSS), provided that the data is acquired as a flat file. Access by this means may, however, greatly diminish the original usefulness of the data.

Each time that a database system is significantly enhanced, the related purge or history files retained on external storage devices must be upgraded to the new format if they are to be usable with the current software. The other alternative is to

retain a copy of each version of the software package and carefully document the history files with which each can be used. This issue is generally referred to as data portability.

Is Special Hardware Required?

Related to the issue of software dependency is hardware compatibility. As we all know and have probably experienced, most software will only run on specific hardware. To avoid the problems of hardware compatibility, archives usually acquire data in a standard format called a flat file, which is accessible by statistical analysis software packages.

Does Your Archives Have Access to an Environmentally-Controlled Data Storage Library?

Once you have appraised and acquired electronic data, what do you do with it?

This question must really be addressed before you begin an acquisition program. We know that electronic data is adversely affected by magnetic fields and electric currents. I am sure that some of us have had the frustrating experience of finding our floppy diskette wiped clean because it came too near to the magnetic field in our desk telephone. In addition, some of the devices on which data is stored are affected by changes in humidity and temperature.

The standard recommended for the storage of magnetic tapes is an environmentally controlled tape library surrounded by a lead data shield. The National Archives of Canada has constructed a tape library (or Bally Data Shield) in Hull,

Quebec. In addition, a backup copy of each tape is stored in an off-site location.

What are the Costs to Store Data?

Most archives will not be able to afford their own storage facility and some form of shared cooperative venture between archives or with the private sector will be required. I have often thought that archivists might approach universities which have a data library to arrange to deposit their archival data on a charge-back basis. This seems logical since a data library already has the storage facility, access to computer hardware, and a public research area to make the data available.

How much does it cost to store one cartridge/cassette or 2400 foot tape per day? The current rate of charge-back used by the British Columbia Systems Corporation, a Crown Corporation which provides computing services to the British Columbia Government, is 22 1/2 cents per day, which is about \$82 per year. Currently, the archival standard is to store one datafile per tape or cassette in order to facilitate retrieval. Often, this represents only a very small percentage of the storage capacity of the tape or cassette.

At such a high daily storage charge, storage cost is obviously a factor for consideration during appraisal. We should not believe that since data is so much more condensed than paper that we can always afford to keep it. This may be true when deciding between keeping the paper or data copy of an identical record series, but does not apply to the overall question of

whether or not we can afford to keep the record series at all. The storage of one electronic tape per year in an environmentally-controlled facility costs a minimum of eight times the storage cost of one box of paper (at a high storage cost per box).

What are the Costs to Process the Data?

Does your institution have on staff the technical expertise to process and make available data, which may, if it contains personal identifiers, require what is called anonymization of the data? Data is anonymized by copying all data elements which do not contain personal identifiers to a new data file which can then be used for research purposes. Processing also includes data verification and preparation or enhancement of documentation and code books so that the data is accessible and usable by the researcher.

What are the Conservation Costs?

Magnetic tapes deteriorate with time as do all other storage media. Whereas you may discover a hundred year old photograph or book tucked away in a desk drawer which has been accidentally preserved in pristine condition, you will not find a magnetic tape in such condition. The emulsion on the surface of magnetic tapes is known to become brittle and "flake off" and, as already mentioned, the electronic imprint diminishes. Distortion is also caused by the impregnation of the electronic imprint from one part of the tape to the parts of the tape with which it is in contact on front and back. To avoid such distortion, the

archival standard recommends that each archival tape be precision rewound and cleaned annually and recopied (i.e., to a new archival quality tape) every five to ten years in order to restore the electronic imprint to its original intensity.

Of course, some of these problems will be alleviated as the technology is refined and new storage media are developed. Nevertheless, the costs of data conservation must be calculated into your archival appraisal. There is no use investing resources in acquiring data if your institution does not have the resources to properly conserve it and the imprint is left to fade away while it sits unused on a shelf.

Can Your Archives Provide Access for Reference Use?

Some data is completely without access restrictions and the archivist can simply have a service bureau make a user copy for a patron to take away for research purposes. However, in many instances, archives must make provision for researchers to use the electronic records it has acquired under more controlled conditions, especially where the data has not been anonymized. Most archival repositories will not have the resources to maintain on-site the equipment required to mount tapes in order to retrieve data for the user.

I feel that this function could be most economically provided through contractual agreement with a university data library, the information systems division of the organization which created the data, or a computer service bureau. Depending upon the nature of the agreement, installation of a user terminal

in the archival repository might still be required together with staff able to give instruction in the use and manipulation of electronic data.

Once provision has been made for reference use, it is imperative that the archival repository establish a policy regarding research costs. To what extent is the repository able to subsidize the cost of manipulating, retrieving, and sorting data and printing reports? I suspect that the entire cost of these research activities must be borne by the researcher.

What are the Computer Costs to Use Data?

The chief benefit of data stored in electronic format is that it can be sorted and manipulated to summarize quantitative information in statistical reports which may not have been previously produced. However, unless you own your own computer, the computer time required to manipulate data on a mainframe computer is most often expensive. The larger the data file you work with the more computer time required.

Most private researchers cannot afford the computer time required to manipulate the large data files which can provide the broad demographic views which are meaningful. Most require a research grant from a funding agency such as Canada Council or the Social Sciences and Humanities Research Council of Canada. Corporate researchers are better able to afford such research. How much electronic data can society afford to preserve if only a very small number of researchers can afford to use it for statistical research?

REAPPRAISAL: SHOULD WE RETAIN THE DATA INDEFINITELY?

The National Archives of Canada, the only Canadian archival repository to initiate a substantial electronic records program, has recognized the need to review and reappraise electronic records "to determine that there is still sufficient evidence of archival, historical and/or research value to warrant their continued retention." In fact, the National Archives' policy requires that the reappraisal date be established at the time of acquisition. The questions suggested for consideration during reappraisal are:

- 1a) Would we accession this machine readable data file if it was offered today?
- 1b) If we would not, why should we continue to keep it?
- 2) Is there a "reasonable" expectation that anyone, with a serious purpose, will request this file?
- 3) Will "scholarship" or Canada's history, in general, be adversely affected because a particular record no longer exists?
- 4) Would you get the consensus of other archivists to agree with your decision?

CONCLUSION

I have raised a lot of questions, but offered few, if any, answers. A lot of my questions relate to cost, which is a factor we don't like to think plays a role in archival appraisal. However, even the largest archival institutions are faced with shrinking resources in face of the explosion of electronic data and other forms of information being offered to them.

Obviously, the issue of machine readable information is one which the archival profession must address and the sooner the better. I fear that there will be no easy answers in this quickly changing field. The one thing I do know is that we as archivists will have to be ruthless in our appraisal of electronic data in determining which has sufficient value to warrant acquisition and preservation. If we are not, we are fooling ourselves because society simply cannot afford to retain anything but a very small fraction of one percent of the machine readable data being created today.