

The Possibility of Partnerships in Archiving Electronic Records: A Proposal for the University of Manitoba

By Mark Vajcner

Paper presented to the Association of Canadian Archivists
Halifax, May 1998

The emergence of the computer in the second half of the twentieth century has had a profound impact on modern society. From the creation of documents using word processing software, databases, and spreadsheets to the increasing use of electronic mail, more and more of the transactions of modern organizations pass at some point through an electronic form. The development of archival theory in the past two hundred years has focused on the record and its ability to serve as accurate evidence of past actions. To the archivist knowledge of the creator and the context of creation are as important as the information conveyed in any document. Only in this way does the record allow us to answer who did or knew what, when, where, and why in carrying out a specific program or function. Electronic recording media, however, pose specific challenges to these long-established archival goals. In response to these challenges archivists are beginning to establish partnerships with stakeholders that have an interest in the long-term preservation of electronic information. This paper will examine one such proposed partnership for the University of Manitoba.

The University of Manitoba is one of Canada's major teaching and research institutions. Located in the city of Winnipeg at the eastern edge of the Canadian prairies,

this university has about 25,000 students and 3,300 faculty members. Established in 1877 to confer degrees on students graduating from the three denominational colleges then in existence in Manitoba, the university began to teach in its own right in 1904 with instruction in the natural and physical sciences. Over the next 25 years the central faculties of Agriculture, Arts and Science, Medicine, and Law were added. Beginning in the 1950s the University Library began to collect the manuscripts of several notable western Canadians, but, it was not until 1978 that a formal archival program was established and a University Archivist appointed. In addition to the manuscript collections, the University Archives houses records relating to the governance and administration of the University. Records from the Office of the President, Senate, Board of Governors, General Faculty Council, the University Council, and from other administrative and support services along with those of numerous faculties and schools make up the bulk of the archives collection.

These offices are beginning to generate and maintain more and more of their vital documents in electronic form. The Archives, in order to continue to meet its mandate, needs to develop strategies and procedures to ensure that records existing in electronic form remain accessible over the long term. Recognizing this as a strategic challenge, the Archives and the Archival Studies program within the Department of History undertook a study of the situation. After several months of interviewing key people within the University Library administrative office, the University Archives, and the Information Services and Technology unit, or IST, an approach began to formulate for the long-term management of

electronic records which was neither entirely custodial in the traditional sense nor entirely decentralized, in the sense of leaving archival records with their creators.

The Archives, like most university archival programs, does not have the technical nor the financial resources to undertake the complex issues of acquiring, migrating, and preserving electronic records. However, other service units on campus perform many of the technical tasks that will be needed in the long-term preservation of electronic records. Partnerships with these service units is the approach which is recommended. The key to success is coordination and cooperation between the Archives, records creators, and IST which plays a central and vital role in the manner in which most university record creators manage their current electronic records.

Information Services and Technology at the University of Manitoba has numerous roles and responsibilities. Its mission is to support the university in its teaching, research and administrative functions. To meet this goal IST maintains the central mainframe computers and the Local Area Networks, or LANs, that support the academic and administrative needs of the University. Electronic records produced on the mainframe computers are well documented and maintained for unlike the world of personal computers, where the technology changes very rapidly, the evolution of mainframe computers is relatively slow. This is due to the complexity of the machines and the huge cost associated with their purchase. The mainframe system, which the University currently uses for administrative purposes, has not changed significantly since about 1976. Although there have been new versions of the software required to make the system

function, these new versions have always been "backward compatible", or capable of understanding data generated with older versions of software.

Nevertheless difficulties do arise with older records. For example, if the current software were, for the sake of argument, the eighth version, it would have no difficulty in reading and understanding records produced in the seventh, sixth or fifth versions. Attempting to read a record written in the original or second version, however, may prove impossible as the software has changed significantly enough to make the older versions unintelligible to the current program. If older versions of the program have not been maintained, or, if the record has not been migrated to new versions of software as these new versions have been introduced, then the record may be difficult or impossible to access.

These difficulties are not as serious for the database systems that the university maintains. Because records are part of an active database, they are usually migrated with each updating of software. This does not mean that software and hardware obsolescence does not occur in a mainframe environment. The early computing files of the late 1960s, although still in the vaults, were rendered unreadable when the university converted to its current mainframe system and are now probably useless.

The computer industry now recognizes the importance of electronic records with enduring value. Many current database packages have been designed to handle what is known as "operation access" to data. Operational access is access to the current state of

specific data. Organizations, however, require information not just for single specific tasks, but for higher level assessment, planning, and strategic decision-support activities. This usually implies having a knowledge of the change in information over time, or access to past data.

The University of Manitoba requires access to precisely this kind of information and IST has been investigating the implementation of a "data warehouse" at the University. A data warehouse is the computer industry's response to this need for current and past data. Regarded as the newest and most advanced method to store and access data, the concept behind the data warehouse is relatively simple. A data warehouse is little more than an integrated database. Designed for access to past information and not simply current records, detailed metadata, or information on the creation and amendment of the records, is kept. A computer industry trade journal recently described the data warehouse as a "subject-oriented, integrated, time-variant, non-volatile collection of data".¹ "Time-variant" and "non-volatile" are the key terms. Records are not modified or deleted in the data warehouse. Therefore they are not "volatile". If changes need to be made to a record then the old record is "closed" and a new one is made which contains the contents of the old record plus the new information. The date and time of the change are "stamped: on each record and this allows for a historical view of the database as it was at any point in the past.

This permits historical and trend analyses, but more important for the archivist, it maintains the integrity of the record and its status as evidence of past actions.

"Recordkeeping systems", according to David Bearman

are information systems which are distinguished by the fact that the information they contain is linked to transactions which they document. Records may be consulted for documentation of those transactions or because they contain information that is useful for some completely separate purpose, but recordkeeping systems do not just contain data to be reused; they maintain evidence over time.²

The data warehouse concept meets several of the criteria that Bearman sets out as being necessary in an electronic record keeping system. It is "responsible" because it is governed by documented policies, assigned responsibilities and formal methodologies for its management. Operational transactions would increasingly be conducted through the data warehouse, thus beginning to meet Bearman's criteria on exclusive implementation of record keeping systems. Finally the concept is "reliable" because it is shielded against loss of information from systems failure.

The records that would be created and maintained in the data warehouse also meet several of Bearman's criteria for electronic records. These criteria state that records be

- Comprehensive, or created for all transactions undertaken by the record keeping system. The data warehouse does this by generating a new record for each transaction undertaken.
- Identifiable. Each record must be discretely identified and separate from the next. This is accomplished in the data warehouse by the time and date "stamp" that each record receives.

- Complete and Inviolable. All records must contain the content, structure and context generated by the transaction that they document. Also no data within a record may be deleted or altered once the transaction which generated it has occurred.
- Auditable, and
- Accessible and redactable. It must be possible to retrieve a record of any transaction at any later date but still "mask" records when it is necessary to deliver censored copies of restricted information.³

While the data warehouse meets or exceeds the above criteria for archival records, it falls short in a number of areas. The most serious is in preservation. It immediately becomes clear that the record keeping process – in which records are never modified but closed and succeeded by new ones – will generate a large amount of records, most of which will contain duplicate information. No one in Information Services and Technology envisions that such records would be kept permanently. The amount of storage space would become too great. As the concept is now taking shape the system may hold, for example, 20 records of a particular activity. Then when the twentieth record is closed and a twenty-first opened, the system would automatically delete the first record in the series, thus there would never be more than a certain number of records per activity in the data warehouse. A second option is to use time as the determining factor. In this scenario a certain number of years worth of records would be kept.

The data warehouse concept could easily serve as the rudimentary beginnings of an archival system on which appraisal, arrangement and description, migration, and reference functions could then be built. As we have seen the data warehouse meets numerous criteria for an archival electronic record keeping system. Information Services

and Technology have been receptive to the concerns of the University Archives recognizing the preservation of electronic records of enduring value as a legitimate and inevitable issue. Partnership between the Archives and IST would allow both to draw on the unique expertise of the other and develop a response that would be viewed as innovative and groundbreaking in both their professional fields.

How would such a partnership proceed? From an archival point of view involvement at the beginning of the record life cycle, or the system design stage, is desirable. For the University of Manitoba it is recommended that the archivist have representation on any committee or planning group that develops the specification for any new computing systems that the university acquires. This will allow for archival issues to be considered up front by allowing the archivist to work with systems technology people and make them aware of archival concerns. Some of these concerns can then be addressed immediately by ensuring that records management tools are built into the design of the system or clearly specified in any tender information.

An added benefit is that when new systems are implemented, a determination on the value of the records the system is likely to keep and a retention schedule can immediately be made before the first record is even created. Scheduling procedures for traditional records have long been defined and inventory techniques for electronic records would resemble those utilized for traditional records. The distinctive characteristics of electronic records, namely their machine-readable content and dependence on specific configurations of hardware and software, would require special consideration.

The scheduling of electronic records should proceed in the same manner as that employed for traditional paper documents. Regardless of format, a retention schedule will indicate the period of time that each series is to be retained, the physical storage medium, the location where the records are to be stored, the date and method of final disposition or other instructions if disposition is not authorized. The hardware and software manuals should also be included on the schedule as they are important in ensuring access to the records. Finally, the fragility of the recording medium necessitate that a period be set after which record series not stored within the data warehouse servers must be recopied onto new media to prevent loss. The schedule now becomes a tool for retention, recopying, and final disposition.

Electronic records and documents with continuing value need to be maintained, either within their originating database or transferred to a special portion of the data warehouse designated as archival. The unique circumstances of electronic recording media make traditional archival transfer difficult. So much so that numerous archival theorists have advanced the idea that archives should give up their traditional custodial role and concentrate energies on monitoring record creators and assisting researchers. Known as the “distributed custody” or “post-custodial” approach, in this model only the functions of providing record keeping advice, appraisal, and possible description are carried out by the archives. All other archival operations, including custody, preservation, and reference service, are performed by the creating agency. The archival institution assumes no responsibility for the actual preservation of electronic records of continuing

value, other than as a monitoring agency ensuring compliance with archival standards for preservation. Ultimate responsibility is devolved to the creating agency, its successors, or its delegates. Supporters of this approach argue that it makes technological sense to leave software-dependent records in the hands of the creators, because they are most likely to be able to maintain access to the records through forward migration.

Supporters of the traditional custodial approach believe that taking electronic records out of the context of the creating agency and into the control of the archives is the only way to imbue the records with true archival value. The archival environment is one in which there are adequate controls to guarantee that records will be preserved and that they will not be altered. No less an archival authority than Sir Hilary Jenkinson, writing in the first quarter of this century, provided the moral defense of archives as repositories of impartial evidence and his vision of the archivist as guardian of that evidence has become the clarion call of the archival profession.

Being separate from the creation and use of the document, and thus not having any stake in its interpretation, allows the archives to serve in this role as protector of the evidence. The response of post-custodialists has always been to raise the issue of technology and the perceived inability of archivist's to preserve the record in its shadow. There is, however, the possibility of a third option for the long term preservation of electronic records. For institutions such as the University of Manitoba, which manage their computing systems along the lines of a corporately centralized model, this approach is particularly well suited.

Custody of electronic records in this third model remains with the Archives and with the Information Services and Technology unit. Like the Archives, IST is relatively impartial in that it does not create or alter the records. IST's sole role is to maintain the systems on which the records reside. Thus once an electronic record keeping system is deemed to contain records of long-term evidential value, through the process of committee/design representation and scheduling examined earlier, copies of the system's records and its audit logs would be periodically transferred to the archival sector of the data warehouse or downloaded to some other electronic storage media that would be in the custody of, although not necessarily physically in, the archives. Once transferred the creating agency and IST would maintain some responsibility toward the record. This responsibility would be to ensure that the record remains accessible over time. Thus, when computing systems are upgraded or replaced, there would be a formal obligation to ensure that records previously transferred to the custody of the archives are either, compatible, converted to be made compatible, or shown to no longer have evidential or historic value. In this manner, the issue of technology, which so concerns "post-custodialists" is somewhat tamed. At one and the same time the concept of preserving evidence, the fundamental principle behind all modern archives, is upheld.

This is the strategy which is proposed for the University of Manitoba. Perhaps many electronic archives will develop along these lines. The rather haphazard and decentralized communications and computer system implementation which has been undertaken by major government institutions, and which has prompted the development of the post-

custodial option, is not necessarily the pattern for many smaller, but still sizable institutions such as the University of Manitoba. Nor is it certain that government will continue along the decentralized path as they too might reconfigure their computer systems management along the lines of a centralized service model. The approach developed in this paper may provide guidelines for their electronic archiving strategies as well as the strategies of numerous other institutions. But above all it shows the importance of establishing partnerships with other information management professionals. Working together the challenges that we face will not seem as daunting.

Endnotes

1. Ray Ferrara and Philip Naecker, "The Data Warehouse: A Giant Step Forward" in DEC Professional (November 1993), p. 28.
2. David Bearman, Electronic Evidence: Strategies for Managing Records in Contemporary Organizations (Pittsburgh: Archives and Museum Informatics, 1994), p. 295.
3. Ibid., pp. 300-04.