Counterpoint

*Archival Theory and Electronic Records**

by **CATHERINE BAILEY**

Modern society has undergone many changes since the end of the Second World War, none more momentous than the development of the computer. From the first vast, room-sized machines dedicated to the simple manipulation of scientific and mathematical data, computers have become small enough to fit onto a desk or into a briefcase and powerful enough to do things that the original computer scientists had only dreamed of. Their influence on society in general and on private and government record keeping in particular is so pervasive that, today, it is almost impossible to avoid interacting with them on a daily basis.

Archivists too have been affected by the development of the computer. The popular image of archives as dark, dusty warrens filled with yellowing paper and staffed by old, hunched figures is slowly being modified by the computer in two ways: through the introduction of automation in the administration of archives and the intellectual control of its holdings, and through the appraisal, acquisition, preservation, and communication of computer or electronic records as archival documents.

Twenty-five years ago, the thought of actually acquiring electronic records as permanent archival records was considered by many to be an unacceptable course for an archivist to take. The prevailing attitude was to consider computers as tools designed for rapid processing of information, and electronic records as transitory records created in the processing stage between paper input and paper output. Because of the transitory nature of electronic records, only the paper input and output were accepted into archives.

In the late 1960s, as the use of computers in record-keeping systems increased, archivists began to change their minds. They realized that in many of those systems paper output records occupied a relatively minor position, and that the information stored within the computer or on magnetic storage tapes had largely replaced them. Electronic records could no longer be dismissed cavalierly because they were "transitory", as they often contained the only copy of archivally valuable information. Unfortunately, vast quantities of archival quality computer records vanished or were destroyed before the archival profession began to grapple with the issue. Confusion over the methods of handling computer records for archival purposes

predominated, fuelled by the strong divergent opinions expressed by authoritative representatives of the archival profession. Some archivists argued that the magnetic records were so different from traditional paper records that it would take either a complete overhaul of archival theory or the elaboration of a completely new theory to handle them properly. Some took this belief to such extremes as to insist that archivists must stop being archivists and concentrate on becoming information managers. Others expressed the belief that computer records were of the same generic nature as paper records and could and should therefore be treated according to the same theoretical principles.[1] Still others were so confused that they half-heartedly hoped that they could ignore the whole question.

Things have changed in the last ten to fifteen years. Various articles have appeared in archival journals such as *The American Archivist* and *Archivaria*, focusing on specific issues in the archival treatment of electronic records and discussing them in the light of theoretical principles and common sense. Yet the efforts were largely sporadic, resulting in the illumination of only a few areas of general concern, such as appraisal and the role of the archivist in the future "paperless society." There seems to have been no clear perception of the overall picture, that is, of the impact of electronic records on archival functions from appraisal to public service, with the result that many key issues remain unexplored and neglected.

The root of this confusion is in attempting to find solutions to questions or problems which are themselves unclear. Only when a specific question has been defined can an hypothetical solution be formulated, tested, and either accepted or rejected. The fundamental question to be asked is whether modern archival theory really requires extensive revisions before it can be applied fully to computer records. However, in order to answer such a general question, archivists need to consider answers to a number of more specific questions: how does the physical medium of the record affect appraisal, arrangement, description and public service? What role does records management play in the treatment of electronic information? In order to decide if archival theory as it now exists is capable of providing the principles which can guide the handling of computer records, that theory must be examined in detail and then applied to electronic records.

The object of this study is to identify some of the problems arising from the application of modern archival theory to the treatment of electronic records. It is not intended to be the definitive work; rather, it is designed to focus archivists' attention on those areas needing further discussion and revisions within the profession as a whole. These areas include various aspects of appraisal and scheduling, arrangement and description, and three aspects of public service, namely copyright, the use of computer records as evidence in court, and transborder data flow.[2] While the archival literature has dealt before with each of these subjects in varying degrees, it has not discussed them in enough depth to allow archivists to state with certainty how they will deal with computer records in the 1990s. Such discussion is vital if archivists are to deal successfully with these records before they are lost or destroyed.

One problem which affects archivists even before they begin to grapple with the theoretical issues relating to electronic records is that of terminology. This is not a new issue. Since the practice of archives spans several centuries, continents, and societies, its terminology greatly reflects the social context of the nation in which a particular

term was first coined. The effects of social, political, and judicial changes are often compounded by the difficulty in translating specialized vocabulary into other languages. While this has long been true of traditional archival terms, archivists today also need to incorporate new terms into archival terminology from other disciplines formerly completely foreign to archives, such as computer science, in order to handle the newer forms of modern records. When archivists and computer scientists take part in discussions in which the terminology is unclear to them, they tend to fall back on their own discipline or personal knowledge to provide meanings for unfamiliar terms; unless they are directed to do otherwise, the potential for lasting communication problems is quite high.[3]

It is very important to ensure that archivists and computer people can communicate well, because archivists well versed in the archival theory used for textual records appraisal often need the help of computer scientists to appraise the more technically complex computer records. Even more important than clear communication to solve technical problems, however, is the application of the theory of archival appraisal to computer records and its subsequent adaptation, where necessary. Appraisal of electronic records is one of the few areas of archival theory which has been illuminated in the literature, from Charles Dollar's seminal article in 1978[4], to John McDonald and Sue Gavrel's "Appraisal Guidelines in the Machine Readable Archives Division" in 1981[5] and Harold Naugler's 1983 RAMP study[6], to mention only the better known works on the subject. After much discussion, it is generally accepted today that the appraisal of computer records consists of two parts: technical analysis and content analysis. Technical analysis, at its most basic level, comprises a detailed examination of the record's medium with a view to determining the readability of the data, ensuring the availability of proper documentation, and assessing cost factors. It also involves analyzing the logical data model used to construct the information flow throughout the computer system, examining the connections between the various parts of the system and other computers, and understanding both the routines used to create the information (master files, transaction files, "archived" copies) and the data export capabilities which must be used in order to acquire the data for archival use. Technical analysis is analogous to the examination of conservation issues considered during the appraisal of textual records. When technical analysis is combined with content analysis, which is essentially the "traditional" process of appraisal usually performed on paper records (although different "value" factors are added, such as manipulability, and linkage possibilities), the appraisal techniques for computer records look no different, conceptually, from those used for paper. Their apparent intricacy is merely a reflection of the increased complexity of the electronic medium, not the nature of the information it holds. Consequently, there is no need to redefine archival theory on appraisal to handle computer records; the theory merely needs to be reapplied in practice.

The electromagnetic form of computer records also raises many supplementary questions which have an impact on the appraisal process for all electronic records. The simple question, "When do we appraise electronic records?", can cause difficulties for archivists and records managers alike. Traditionally, records management practices have designated records as being active, semi-active, or inactive. Using the life cycle model, paper records are deemed to be active from creation to classification and maintenance when they are used frequently in the carrying out of daily business.

Later, when they are consulted less frequently, they become semi-active. Finally, the records are designated inactive when they are no longer used by the organization for its activity. At this stage, the records are usually transferred to an archives or destroyed. Inactive records may be "reactivated", but this is a rare occurrence.

If archivists were to apply the life cycle model as it was designed for paper records directly to electronic records, the initial impression would be that the computer records could not be classified so neatly into active, semi-active, and inactive stages. Electronic records are clearly active if they reside on a computer system which is used daily; they are inactive if they and their locational indexes are removed from the system and frozen in a certain form. It is the definition of a semi-active record, however, which seems to cause the real problem. If similar records in paper form have their status altered from active to semi-active and inactive based on a reduction in the number of times they are consulted, how can computer records in the same situation be classified? Are they active because they continue to reside, unconsulted, on an active system? Are they semi-active or inactive simply because they have ceased to be used in the course of everyday business? Can the two kinds of records exist at the same time if proper backups of the system are taken at regular intervals?

The answer to these questions lies in treating the life cycle model on a more conceptual level. If archivists consider the life cycle as an abstract expression of the legal authority over a record rather than a designation of its physical state or activity, then the differences between a paper record and an electronic record disappear. It does not matter whether a record is located on a disc pack in an organization or department, on storage tapes in a records centre tape library, or on tapes or discs in an archives; its administrative and, especially, legal status is still determined by the amount of use it gets and the jurisdiction that controls it. Consider as an example the records of the database the federal government uses to produce Family Allowance cheques every month. Active records are those which are stored on the system's disc drives and can therefore be called up to the screen at any moment for modification. Records before a certain date, 1979 for example, which are presently not on the system but on storage tapes in a tape library, are semi-active since they can be restored to an active administrative status at any time simply by uploading them to the database. Finally, those records which are removed or copied from the backup tapes and either transferred to the archives or destroyed as having no lasting value, are considered inactive records because at this point, they are outside the control of the creating department.

Perhaps contrary to the initial impression, it is not impossible to have textual records in active and inactive forms at the same time. Suppose that a government department considers some of its records to be administratively important (active) for several decades, and therefore unavailable for transfer to the archives, yet those records have also been appraised by the archivist as having high archival value. Ideally, in such a situation, the archives could make a copy on microfilm or fiche and treat them as an inactive, archival record in order to ensure the preservation of the information, but at the same time allow the department to retain the copy for administrative use. While the active and inactive textual record can exist at the same time, it must be pointed out that it is usually the cost of making such film, fiche, or photocopies, not conceptual problems, which prevents this course of action; the computer has simply made it easier to duplicate administratively and archivally valuable information and make it available in multiple settings. Proper application

of the life cycle model at a conceptual level, therefore, shows that there is no difference between traditional textual records and electronic records, hence no need to make major revisions to archival theory to deal with the latter.

Because it is too easy to lose valuable information in the early active and semi-active stages of computer records' existence, stages which are traditionally the domain of the records manager, archivists can no longer accept the role of passive recipient. They cannot wait until inactive electronic records are offered to them for appraisal, as they might have for paper records; too many computer records have vanished by then, and the documentation necessary for their proper appraisal has been lost, destroyed, or is hopelessly outdated. The sheer volatility of electronic records should be a powerful inducement for archivists to accept an increased involvement in the scheduling process, beginning at the system design stage. Again, however, this is not an issue of new or revised theory or principle, but merely one of timing and strategy.

A more serious problem that archivists face in appraising electronic records, even at the design stage, comes not from the application of current archival theory on the scheduling of electronic records but from its application in the near future. In the past few years, it has been generally assumed that each individual computer system or application is an expression of a single function and is the responsibility of a single agency. Under these circumstances, it is relatively straightforward to complete, for every system, a system overview, which is a general description of the informational content of a computer application from a "top-down" perspective. A system overview identifies the corporate data which is associated with either a single application supporting a single function, or for a portion of a larger system, and provides details of the physical form in which the data is to be found. It also records the application's key input, processing, and output steps.[7] While the "system overview approach" does give valuable specific details about the information contained in a single application, however, it has a serious flaw as an appraisal and scheduling tool. Instead of treating the information in the computer system as a part of the organization's entire record-keeping system, it treats each computer application or sub-system as a relatively isolated entity. If in fact the system overview was describing a relatively straightforward, single function application, it might be possible to add the kind of contextual analysis usually done in appraising paper records as a supplement. The trend in many large corporations and governments, however, is towards data resource management, that is, towards a situation in which many ministries, agencies, departments, or organizations or, more often, many of their constituent parts combine their resources to create and maintain a single, large system or database which can serve all of their diverse but related needs at once. Electronic information then becomes so fluid that not only does it become difficult to determine the active, semi-active, and inactive stages of records, but it also becomes next to impossible to determine the provenance of the records. There is no longer a single application upon which to focus attention, so that the system overview approach becomes complex and difficult. Where do archivists begin to schedule the contents of these shared databases? Can they legitimately break them down into smaller units fit for individual schedules or overviews, or will this act destroy the true nature of the system? Or will such a system require a scheduling technique completely different from that of the system overview?

Until archivists master the most basic procedures for scheduling electronic data which is linked to a single application within a larger system, they will not be able to develop the techniques necessary to deal with the more complex data management computer systems of the future. The system overview method does not seem to be what archivists need to appraise and schedule the electronic records of the 1990s. In order to develop the procedures necessary to replace the system overview, cooperation between records managers, archivists, and other related professionals is not only encouraged but demanded; without such cooperation, the preservation of electronic records will suffer greatly.

Once archivists have determined how and when to appraise computer records, they are faced with another conflict if they follow Schellenberg's view that archives are only those few records carefully selected to make the best use of available space.[8] Computers have made it quite possible to store vast quantities of information and data in smaller and smaller spaces; in fact, the suggestion has been made that routine administrative or housekeeping records in electronic form could be kept, whereas their paper equivalents would be deemed too bulky to be useful.[9]

It is all too easy, however, for archivists dealing with electronic records to find themselves in the awkward position of having concentrated their appraisal on the single datafile or group of datafiles. Because computer records take up very little space in relation to paper, an archives might find it reasonable to accept a datafile of which only a portion has archival value. After all, it may be only one tape. This situation leads to a relaxation of appraisal standards just at the time when archivists are learning that they must decide early in the life of an electronic record system exactly what data elements will have archival value and must therefore be selected for preservation. On the other hand, the more practical concerns of conservation costs associated with rewinding, reformatting, and copying tapes to ensure their long-term preservation means that acquiring mountains of redundant data burdens archives with unnecessary expenditures. Similarly, the costs to researchers to "run" or process the data, if millions rather than thousands of records in a data file are kept, greatly reduces the archival value of the material; if no one can afford to pay the processing costs to use the records, are they actually worth preserving? In that respect, appraisal standards for computer records need to be tightened up dramatically to ensure that what is preserved is indeed archivally valuable. Finally, if archivists once again take their analysis to a more conceptual level and look at the place that the specific datafile occupies within the entire record-keeping system, they might discover that the datafiles which looked so promising on their own are really only a minor part of the entire system; to preserve the best archival picture of that system, they would have done better to take fewer tapes from a higher functional level in the system and/or organization. This reinforces the fact that if archivists look closely at archival theory, there is no difference between paper and electronic records; if they simply continue to apply to computer records the existing textual appraisal techniques, including their regard for the records' proper context within the whole record-keeping system, they should have no need to make drastic conceptual changes to the theory.[10]

What is necessary for the proper preservation of electronic records is an improved appraisal process that would combine the best elements of the archival appraisal of conventional records with the technical considerations linked to the

magnetic medium. While this may sound like what McDonald, Gavrel, and Naugler have outlined, what this author proposes is a more integrated program divided into three stages.

First, there should be a greater emphasis on the appraisal of computerized information as soon after its creation as possible. Clearly, this breaks with traditional archival practices in that it requires the archivist to become involved at the active stage of the record's life cycle. Unfortunately, in the early stages of the life of electronic records, only computer specialists responsible for the systems have any sort of control over the information, and they are much more concerned with processing up-to-date information than with preserving outdated information for archives. The result is that valuable information can easily be lost. If there were provision for early appraisal of electronic records, however, less of this loss might occur.

The second stage of the appraisal should encompass the combination of technical and content analysis. Once a record has fulfilled its purpose and become valueless to its creator, it should be considered for permanent archival preservation. If a machine readable record has already been assessed as being valuable in the first stage of appraisal, then it will be easier to separate it from the non-essential records around it and much time and energy will be saved. Careful application of the appraisal criteria is then necessary to ensure that only the most important records are preserved, no matter how easy the medium makes the storage of information of dubious archival value.

The final stage of the appraisal process must occur after the datafile is taken into archival custody. Since there is an overabundance of information available in modern public records, both paper and electronic, and in view of the fact that records can conceivably lose their value, datafiles should be reappraised occasionally to ensure that their archival values have not been overemphasized. If a file which was accessioned because it seemed to fit in with a certain research trend, or seemed to be the key information in a computer system, no longer appears to be as useful as it seemed during the initial appraisal, consideration must be given to having it removed from the holdings. This is not to suggest that archives should make a concerted effort to cull their holdings; rather, this process should be considered a means of reviewing past appraisal decisions in which the "benefit of the doubt" was given in favour of retaining the records. There must be every effort made to ensure that the burden of proof lies on the side of deaccessioning, and that reckless destruction is not practised, but it still should be done.

Of the three stages, this author believes that the first is the most vital because without an identification of the importance of information soon after its creation, electronic records may never even reach the second stage of a full archival appraisal. Already countless valuable electronic records have vanished because of the transitory nature of the medium. If the trend continues, there could be a very large gap in the records of modern society.

Once archivists have successfully determined the proper appraisal techniques for electronic records, they must turn their collective attention to intellectual control, the arrangement and description of the records that will assist researchers in gaining access to the material. Archivists have written a great deal on the arrangement and description of paper records, yet there is a distinct paucity of literature on the

arrangement and description of their modern counterparts, computer records. Why has there been such a gap in the archival literature on electronic records? Could it be because archivists dislike writing about one of the most time-consuming or technical parts of their function? Arrangement, after all, is often considered tedious, while description, in the absence of standards, is recognized as a highly individualistic task, differing from institution to institution. Is it because the large volume of electronic records requires swift action on the part of archivists and concentration on appraisal, acquisition, and conservation? Both of these hypotheses could account in part for the lack of writing on arrangement and description of electronic records. The real reason is much more profound. The principles of archival arrangement do not seem to be as relevant to electronic records as they are to paper records; therefore, it is even more difficult to write about the attempts. Furthermore, description, although it is vital to both paper and electronic records, suffers from arrangement's apparent lack of importance. Archival finding aids, such as the inventory, are designed to show the fonds' arrangement; if arrangement in the traditional sense is irrelevant, the accompanying description collapses.

While the accepted archival principles of *respect des fonds* and respect of original order function extremely well for paper records, they are somewhat more difficult to apply to the archival arrangement of electronic records. The principle of provenance has become so ingrained in archival practice that its application to computer records should be an accepted fact. No matter how many individual tapes, files, or records are received by an archival institution, if they all come from the same administrative body or official they must be treated as a single fonds. Unfortunately, in many of the data-library-type systems adopted by archives from the library world, such is not the case, an observation which will be examined more closely later.

If the principle of provenance for electronic records is neglected, it seems to be even more difficult to apply the principle of respect of original order and all of its various corollaries to electronic records because of the nature of the medium. Physical arrangement is of little concern for electronic records because the flexibility of the medium allows anyone using the information, either the creator or a secondary user, to impose his own order on the records according to his particular needs. It is therefore almost impossible to determine exactly what the "original order" of the datafile was. Even if the archivist knew which locational index had been in place when the datafile was first made operational, he or she would need documented proof that the index had not been altered since that initial operation because changes are not immediately obvious to the observer and the documentation of them is not usually kept.

If it is not possible to determine the original order of a datafile, it must also be said that there is really no need for archivists to try and determine it. Emphasis on the preservation of original order is essential for paper records because the physical arrangement of the files themselves gives fundamental information about the records and the functions and structure of the administrative body which created them. With electronic records, information about the records derived from their arrangement comes from examining the locational indexes which the computer uses to find requested records. As long as those indexes are removed from the computer system, verified as accurate, and remain frozen and unaltered, the information the archivist seeks from the file's arrangement will remain intact.

Furthermore, if archivists stop considering arrangement as a purely physical concept, they can see that electronic records do have something entirely analogous to the traditional archivist's arrangement. When a new application is designed for any computer system, the designers and users give much careful consideration to the actual layout of the individual data elements within each record in order to make the most efficient use of the system. They consider how the records can be linked to each other and to other systems, and what role or function each type of record fulfils. If archivists were to preserve information about these design decisions, would they not gain as much information from electronic records as we could from the arrangement of paper records? This author believes that they would.

A corollary of these observations on the arrangement of computer records is the question of the role of software in an archives. Like the indexes which the computer uses to keep track of the physical location of the data, the commands of a particular computer program, such as DBase III or WordPerfect, which are used to manipulate the data, can leave as great an impression on the resultant electronic information as the physical arrangement of paper files. If archives insist on combatting problems of hardware and software incompatibility or obsolescence by accepting only flat files in ASCII or EBCDIC (IBM) format, are they not losing some valuable evidential information about the records? As persons responsible for documenting all aspects of society, should archivists become, in a limited fashion, software collectors by accepting some of the more popular and widely used programs along with the data they manipulate? How would this further affect the application of copyright? Would it be beneficial to future users to have these programs available to see how the original creators used the information? In the rapidly changing world of computers, where software comes and goes quickly, these questions need to be addressed by the profession in the near future.

If arrangement as a status and as a process seems to have less significance for the archivist dealing with electronic records than for the archivist dealing with paper records, the same cannot be said about description. Archival arrangement and archival description are aspects of the same function, namely preparing records for use by any interested party; they are traditionally connected (or put into relation) by three main assumptions based on archival theory. While archivists follow the principle of provenance and consider the fonds to be the fundamental unit of archival organization and arrangement, they also assume that the arrangement of archival material can be viewed in terms of levels. The "five levels of arrangement" are: repository, record group (or fonds), series, filing unit, and item.[11] Furthermore, archivists assume that, since archival description is designed to reflect the arrangement of the records, there can be a differentiation of levels of archival description into inter-institutional, repository, thematic group, records group (or fonds), series, filing unit, and item.[12] Finally, archivists assume that specific finding aids are usually more appropriate to certain levels of arrangement than to others.[13] An inventory, for example, usually describes a fonds, but in larger institutions, such as a national or provincial archives, it could describe many fonds within a record group. A guide is generally done at the institutional level and describes several fonds containing information on related subjects; it is also well suited to the inter-institutional level. In any case, the logical progression of a finding aid is from the general to the specific.[14]

A closer look at archival description as it relates to electronic records should show that in theory, planned, standardized and accurate description is equally important for all forms of archival materials, whether they be paper records, photographs, maps, or magnetic tapes. A proper set of finding aids allows users of archives to become more familiar with the organization and contents of a fonds without having to examine the materials themselves. An interconnected finding aid system also allows users to view the larger picture, to see how a single fonds or series of files fits into the context of an administrative body's total records holdings.

In practice, however, the fact that the magnetic medium is so physically different from paper makes accurate archival description a vital need for electronic records. Improper or sloppy description of paper records does not necessarily preclude a user from finding the information he or she wants by calling for the entire fonds and physically sifting through it. It would be a laborious task, but it can be done; indeed, it often is, because the lack of personnel and funding forces many institutions to forgo extensive description. That kind of exercise is much more difficult to do with computer records. Because the records are created and held in a form unviewable by human eyes without the assistance of a computer, there is a certain level of technical description necessary before a user can begin to access a computer file. Given only a name and a physical location, any user can begin searching a fonds of paper records and process the information seen on the pages; the user of electronic records would need to know not only the title and physical location of the tape, but also the kind of hardware (specifically, the tape or disc drive) which supports the file, and possibly the name of the software package as well before he or she can reach the same stage. This leads archivists to believe that there is a minimum level of description necessary for electronic records that is more detailed than the minimum requirements for paper records. That minimum level is also variable, depending upon the initial level of technical expertise of the particular user of the datafile. If a user is quite familiar with the computer system which runs the file, he or she may need only the name of the computer to proceed. If, however, a user is not familiar with either computers in general or the system in particular, he or she may require a set of documentation, ranging from a simple codebook to a series of user manuals outlining all the software commands, in order to begin extracting any information from the datafile.

With electronic records, then, archivists must devote much time and effort to seeking to improve their techniques of description. The limitations of the physical medium must be overcome through the use of a more organized, better integrated and standardized system of finding aids. At present, archivists are in a good position to develop proper archival standards for describing computer records; not only is the whole profession keenly aware of and eager to implement standards for all forms of archival materials, but so few institutions have magnetic material that it will be possible to develop and implement specific standards before they are required by the majority of archivists. A few archivists have suggested that when describing electronic records, the archival world can benefit greatly by adapting the machine readable data file (MRDF) descriptive systems as practised in the library world. Works such as Sue Dodd's *Cataloguing Machine Readable Data Files*[15] and *Anglo American Cataloguing Rules 2nd Edition* (AACR2) have been used as models for several archives setting up systems to handle computer records.

But do archivists really want to adopt a system which, although it captures much of the same basic information, concentrates almost exclusively on the item in hand, either the magnetic tape or the disk? Because libraries tend to receive single datafiles with survey and other research data, rather than series of operational or administrative records, they generally do not place the datafile within the context of the rest of the records produced by the same administrative body. This concentration on the technical details of the individual datafile detracts from more archival concerns, such as their function and their place in the context of the administrative body's organization. It reflects the 1970s' focus on one-shot, survey forms of electronic records and does not serve well the quite different electronic records to come in the 1990s.

What a system of archival description for electronic records needs to do is to provide the proper context for the datafile being described within the entire record-keeping system of the creating agency. The descriptive tools cannot be limited to catalogue entries or data abstracts focused on the individual datafile; there must also be a method of providing links between the electronic records and their related paper and other media records, which often include the input and output documents of the computer system. An archival system of description needs to put more emphasis on the inventory series description that gives the records their proper context, as well treat the documentation manual, the equivalent of a finding aid for paper records, as another focus of attention instead of concentrating on the card catalogue entry as librarians do, or as present MRDF description systems have done.

There are several things that archivists can do to improve current descriptive practices for electronic records. First, they can accept the basic elements of the library system, focusing their energy on adapting various elements to reflect a proper archival viewpoint. Secondly, they must decide what kinds of finding aids would best suit their purposes, how those tools should be linked, and how any system for computer records description may be incorporated into a larger, integrated descriptive system for all forms of archival material. Finally, they must develop and implement standard policies and procedures for description, including authority lists and standard vocabularies. These conclusions may be applied equally well to paper records as to electronic records; in fact, through better planning of descriptive tools, the implementation of standards, and the development of an integrated system of description, archivists are once more urged to treat computer records and paper records in the same fashion.

Improved descriptive tools for electronic records is only one aspect of a subject which archivists need to address in more detail: public service for computer records. In the 1970s and 1980s, archives have tended to act more like data libraries or records centres than like archives when providing public service for electronic information. If a researcher manages to find out that an institution holds datafiles of interest and requests to use them, the response has been to provide a copy of the tape and the documentation manual. There is usually no personal contact between the archivist and the researcher; the assumption seems to be that if the researcher gets a copy of the tape and has access to a computer, he or she can figure out how to extract the information in the fashion desired. At the same time, the lack of personal contact exacerbates the isolation of the electronic record from complementary material in other media. Without proper inventory series descriptions and finding

aids which provide the necessary archival context for computer records, the entire electronic holdings of an archives is reduced to a group of discrete items, a situation which greatly reduces the datafiles' evidential and informational values.

Archivists cannot continue to provide public service for computer records in the 1990s in this fashion without doing a serious disservice to their users. While a major shift in archival reference practices cannot be implemented overnight, there are several areas in which they might begin to be improved; description is one of them. Since the manipulability of the electromagnetic medium makes it much easier to provide anonymized copies of records under the provisions of Access to Information and Privacy Acts, archivists should undertake more discussion of the circumstances under which such copies are provided and what their archival value is. As well, there are three other areas affecting archivists' ability to provide public service for computer records which can be examined more closely; they are copyright, the use of computerized information as evidence in court, and transborder data flow.

The question of the application of copyright affects both the software used to process the records and the records themselves. Software, as a single work containing the ideas expressed by one or more authors, is considered to be protected under the provisions of the Copyright Act of 1924; therefore, how can an archives process and make available to researchers a series of electronic records if the records are dependent for their interpretation upon a software package which is subject to copyright? The records are indecipherable without the software, yet the archives does not have the right to make copies of the program for the users without becoming guilty of software piracy, something which legislators have sought to curtail. Even if the archives seeks to avoid this issue by accepting for accession only those records which are software independent, they are faced with another growing problem resulting from the continuous development of computer technology. Because copyright provisions protect only the *form* of the work, not the ideas contained within it, how can archivists apply copyright legislation to the ever-changing information found in a large, shared database? Does copyright reside with one person or group owning the database because it is a single "form"? Does each person who entered information into the database maintain individual copyright privileges? Can archivists even begin to apply the legislation to an entity which is so fluid that it can be rearranged into countless new forms at any time, each with its own copyright? The resolution of these issues is of vital importance to all archivists, especially those faced with acquiring electronic records from large systems in the near future. If the dissemination of computer records is hindered by copyright legislation, this will have a profound effect on the decision of which records to keep, and how to make those selected records available.[16]

Another issue which affects both the appraisal and dissemination of an electronic record is its legal status, or its acceptance as evidence in court cases. Anyone trying to introduce evidence of any kind into court and have the judge declare it admissible must be concerned with two rules: the Hearsay Rule and the Best Evidence Rule. The Hearsay Rule is a portion of common law designed to prevent the introduction of "statements" made outside of the confines of the court where the persons making the statements could not be cross-examined and the statements corroborated. Strict application of this rule would exclude almost all of the evidence put in front of the court from being heard by the judge, even that evidence which cannot be presented

in any other manner. The Best Evidence Rule states that the evidence being offered to the court must be either the original "writing" or a duplicate; the latter is acceptable only in cases where the party introducing the record can prove to the court's satisfaction that the original is unavailable.

To have computer evidence accepted in court, the party presenting it must satisfy the court that, in spite of the fact that they are technically hearsay, electronic records are acceptable as evidence. In order to do this, the party must lay out foundation evidence which will attest to the fact that the electronic records were created in the course of regular business proceedings. In the United States, foundation evidence, consisting largely of documentation explaining how the system works, what data entry procedures are used, what security is in place, and how the printouts are produced, is considered absolutely essential in establishing the regularity of record preparation. William A. Fenwick and Gordon K. Davidson, in their 1978 article on computer evidence,[17] give several points which they considered have to be answered to the judge's satisfaction in order to create proper foundation evidence for computer printouts. Those points have since been taken as the standard preparation for introducing electronic records as evidence in the United States.

The Canada Evidence Act, however, is very vague on the subject of electronic records. Creation "in the usual and ordinary course of business" is the only condition which determines their admissibility. There is no set standard of foundation evidence necessary to establish the reliability of the records as there is in the United States. Since many judges are unfamiliar with the elements of computer record keeping and tend to view computers as eminently reliable, the onus of the burden of proof is put, not on those seeking admission of evidence, but on those people trying to prevent admission,[18] notwithstanding the fact that the information necessary to prove unreliability or reliability is much more easily obtained by the people wishing to use the system's records in court than by the opponents.

The vagueness of the phrase "usual and ordinary course of business" creates another problem for computer evidence in Canadian courts. Because it can be interpreted in a myriad of different ways, judges depend upon case law, previous decisions and precedents to determine what constitutes suitable computer evidence and its foundation. Consequently, persons presenting computerized evidence are never completely sure about the quality and quantity of foundation evidence necessary. One judge may be satisfied with the testimony of a single expert witness, whereas another might demand full documentation along the lines of that suggested by Fenwick and Davidson.

The archivist who is more often involved in these problems is the corporate archivist. As more and more large corporations come to depend upon computer record keeping, they will find an increasing need to produce computer printouts as evidence in litigation. Good foundation evidence suitable for court proceedings requires documentation of all the stages of a system, from its design to its inactive stage. This implies the existence of proper user manuals, design specifications, schedules, security procedures outlines, and classification links to the paper records, to name only a few elements. The archivist may well be called upon to back up the foundation evidence with testimony as an expert witness; he or she may vouch for the normal operation of the system, or its security procedures, or authenticate the printouts created by the system. Similarly, if a government archivist were called upon to

provide inactive government records in machine readable form to a judge, he or she would have to prepare foundation evidence documenting the procedures carried out during the processing and public service stages. The archivist could also be called as an expert witness in any case where the records of a defunct agency held within that archives are called into court. Actually, this function of the modern archivist is not different in principle from the function of the traditional (particularly European) archivist, who was called to authenticate paper records on the basis of a diplomatic criticism of the medium, the text, and the documentary procedure attested by the chancery notes. Consequently, if it is true that the modern archivist deals with the technicalities of the medium, the different structure of the text, and the specific procedures governing the creation, maintenance, and use of computer records, it is also true that the theory governing his authentication function is unchanged.

Given that more and more countries are accepting computer records as evidence in court, archivists must remain aware of new developments in this area of the law. If an archivist can be placed on the stand, either as an expert witness or to provide basic foundation evidence, the appraisal criteria must have been sufficiently detailed and rigorous to ensure that the archivist has the necessary information at hand in the years following the records' accessioning to allow him or her to carry out those required duties.

Archivists responsible for the acquisition and dissemination of electronic records should also give more thought to the effects of transborder data flow, defined as "the movement across national boundaries of data and information for processing and storage in computer systems."[19] The transfer can be text or electromagnetic tapes or discs sent physically through the post, or electronic impulses transmitted through telecommunications equipment. Whether the transfer across a national boundary involves the data of a multinational corporation, an international union, a service or credit bureau, or the electronic transfer of funds by a bank, transborder data flow has a distinct effect on four aspects of a nation's existence: national economy, national sovereignty, individual privacy, and cultural identity. Of these four, only the latter three have a serious impact on archives.

Archivists' concerns over transborder data flow as it relates to national sovereignty are tied directly to the issues of freedom of information and protection of individual privacy. Information is essential for a nation to determine the direction of its social, political, economic, and cultural policies; if information is not readily available because it has been transmitted to another country for processing, the nation's independence is undermined. The information could even become completely inaccessible to its creators, thereby frustrating government policy, the courts, and scholarly research. The information-generating nation becomes dependent upon the goodwill of another state and upon the latter's degree of internal control over such things as industrial sabotage, power failures, civil unrest, strikes, and unemployment. More damaging, however, is the fact that a nation exporting data to another country for processing cannot apply its own laws regarding freedom of information, copyright, or privacy to its own information within the jurisdiction of the storage country. When information is sent out of the country for processing, archivists may find themselves unable either to guarantee or to safeguard the security and integrity of personal information in electronic form within their holdings, because copies of it are subject to the different legislation of the foreign country. It may happen that a

researcher who has been denied access to such material restricted by Canadian law obtains such access in another country holding copies of the same material; in the absence of an international agreement, the archivist and his or her government, or the subject of the information, are powerless to stop the researcher. Therefore, transborder data flow can only be controlled through the development of specific national and international legislation ensuring the security of information.

Archivists ought to be equally concerned with an aspect of transborder data flow which has not been discussed in much depth in the literature: the issue of cultural identity. Culture itself is not easy to define, and the effects of transborder data flow upon it are difficult to pinpoint. Marguerite A. Vogel points out that culture, or a people's self-image, is greatly influenced by what people are told and are shown.[20] It follows, therefore, that mass market advertising techniques, which rely heavily upon computerized information designed to categorize people according to demographics, living standards, and buying patterns, often shape and mould a nation's culture with or without its citizens' permission. If that influence comes from another country, using information gathered and sent to it for processing, it is the processing nation's view that will predominate. American influence in Canada is a good example.

Another aspect of cultural identity affecting archivists is that of scholarship. Transborder data flow is common in various sections of society which produce records containing a great deal of information valuable to scholars. International unions, for example, regularly transmit their electronic records out of the country. A researcher studying the impact of union development in Canadian society, therefore, could be faced with gaps in the archival record which Canadian archivists are unable to prevent. Such a researcher, in order to get complete information, would have to go to another country to find material vital to the writing of Canadian history. Furthermore, the possibility exists that the researcher may be denied access to that information under the other nation's legislation. This situation gives a great degree of power or control over the information of a nation to a foreign state; by manipulation or denial of access, the foreign country can shape Canadian history to suit its own needs and perceptions.

Transborder data flow is an issue which archivists need to examine more closely, as it affects their ability both to acquire a proper archival record and to make it available to researchers promptly, securely, and fully. Harold Naugler touches on the subject of transborder data flow briefly in his book *The Archival Appraisal of Machine Readable Records*, yet he treats it not as a separate issue but only as an appraisal criterion, urging the creation of more national and international legislation.[21] Archival strategy and practice need to be developed in this area, because transborder data flow presents problems related not only to the dissemination of electronic records but also to their appraisal, arrangement, and description.

The question underlying this study is whether modern archival theory really requires extensive revision before it can be applied fully to computer records. If archivists take a close look at what the theoretical principles behind archival practices actually entail, the answer must be "no". The appraisal process as it applies to electronic records shows that, despite the different media, there is no fundamental difference in evaluating paper or computerized information; both are assessed according to their administrative, legal, evidential, and informational values within

the context of the record-keeping system in which they were created. In terms of the arrangement of electronic records, the theoretical principles merely need to be reformulated, not redefined, to reflect a more conceptual, less physical orientation. The principles governing description and public service actually encourage the equal treatment of paper and computer records, since a good system of archival description should focus on the context of the information itself, not its physical form. In fact, traditional archival theory, if properly applied, will make a significant improvement in the present practices for electronic records adopted by some archives in the 1970s and 1980s. Archivists do still need to discuss various aspects of archival theory and electronic records in more depth, but what should be apparent is that the major changes brought about by the new medium are in archival practice and strategy and planning, not in archival theory.

Notes

* The substance of this article is taken from the author's Master of Archival Studies thesis entitled "Archival Theory and Machine Readable Records: Some Problems and Issues" (M.A.S. thesis, University of British Columbia, August 1988). The author wishes to thank Luciana Duranti for her help on the original, and Terry Cook for his comments and suggestions on this article. The views presented herein are not necessarily those of the National Archives of Canada.

1 For an example of the former position, see Richard Kesner, "Automated Information Management: Is There a Role for the Archivist in the Office of the Future," *Archivaria* 19 (Winter 1984-5), pp. 162-172; for examples of the latter, see Trudy Huskamp Peterson, "Archival Principles and the Records of the New Technology," *The American Archivist* 47 (4) (Fall 1984), pp. 383-393, and Terry Cook, "From Information to Knowledge: An Intellectual Paradigm for Archives," *Archivaria* 19 (Winter 1984-5), pp. 28-49.

2 In the interest of brevity, the author has chosen not to discuss, except in passing, the issue of conservation as it relates to electronic records, since there is already some very good literature available on the subject. Conservation affects all aspects of electronic records, from appraisal through processing to public service; readers are advised to look at the following articles as a sample of the information available on the subject: John Mallinson, "Preserving Machine Readable Archival Records for the Millenia," and Katherine Gavrel, "Preserving Machine-Readable Archival Records: A Reply to John Mallinson," *Archivaria* 22 (Summer 1986), pp. 147-52 and 153-55; Sidney B. Geller, Care and Handling of Computer Magnetic Storage Media (Washington, D.C.: U.S. Government Printing Office, 1983); Lisa Fox, "Archival Preservation in the Age of Technology," *Provenance* III (Spring 1985), pp. 23-29; and Margaret Hedstrom, *Archives and Manuscripts: Machine Readable Records* (Chicago: Society of American Archivists, 1983).

3 There are many homonymic terms used in computer science and archival theory which cause communication difficulties for archivists and a variety of computer scientists, from technical support personnel to data managers and designers and users. The term "records" can mean entirely different things to an archivist (who sees them as static recorded evidence, in any physical format, of a completed transaction), a technical computer person (for whom record is short for logical record, a term describing aspects of physical storage and layout), and a data designer (who often thinks of records in terms of records as building units in the development of a model outlining the flow and use of information in a system or database). Likewise, the term "archives" can cause confusion. While archivists consider archives to be information in any form which has a lasting evidential and/or informational value, computer users think "archives" are periodic copies of the information present in a system taken to serve as backup copies in case of accidental damage to or erasure on the system. There is no selection, nor do the users have any concern as to whether or not the information was created in the course of fulfilling a function. These are only two of several examples of potential terminology problems; for a fuller discussion of this subject, see Chapter One of Bailey, "Archival Theory and Machine Readable Records: Some Problems and Issues".

4 Charles M. Dollar, "The Appraisal of Machine Readable Records," *The American Archivist* 41 (October 1978), pp. 423-30.

5 Katherine Gavrel and John McDonald, "Appraisal Guidelines in the Machine Readable Archives Division." (Ottawa: Public Archives of Canada, 1981).

6 Harold Naugler, *The Archival Appraisal of Machine Readable Records: A RAMP Study with Guidelines*. (Paris: UNESCO, 1983).

7 Public Archives of Canada, "Guidelines for the Scheduling of Data in Automated Systems." (Ottawa: Public Archives of Canada, 1986), p. 6.

8 Theodore R. Schellenberg, *Modern Archives: Principles and Techniques*. (Chicago: University of Chicago Press, 1956; reprint ed., Chicago: Midway Reprint, 1975), p. 16.

9 Gavrel and McDonald, "Appraisal Guidelines," p. 6.

10 Since modern archivists are faced with an enormous information universe from which they must select the most valuable records to preserve, the next question that should be asked is this: if archivists cannot save everything that is on a computer simply because it is easy to do so, when appraising computer records can they, instead, use sampling techniques on them? In his study *The use of sampling techniques in the retention of records: A RAMP study with guidelines* (Paris: UNESCO, 1981), Felix Hull outlined four recognized methods of sampling used in archives to reduce bulk: the taking of samples or specimens; purposive (qualitative) sampling on a pre-determined pattern or bias; and two forms of quantitative sampling, systematic on a pre-determined basis, and true random sampling using a scientific objective process (page 10). While it would be possible to do the same kind of systematic sampling on electronic records that is done on paper files, it is questionable if this method of reducing bulk is effective, or whether it simply destroys the representativeness of the series. It depends upon the establishment of a particular pattern of selection in that every nth file is preserved, or that all the records for a chronological period or a certain geographical area are selected. It is not to be confused with true random sampling where files are selected on the basis of a random number table and every item has an equal chance of being chosen (pages 13-15).

Theoretically, a true random sample of a particularly large body of homogeneous records in an electronic form should be simple to do, yet archivists must not ignore the fact that there would be technical difficulties because of the form of the records. For a more detailed look at this issue, see Bailey, "Archival Theory and Machine Readable Records," Chapter Two, pp. 51-55.

11 Bureau of Canadian Archivists, *Toward Descriptive Standards: Report and Recommendations of the Canadian Working Group on Archival Descriptive Standards* (Ottawa: Bureau of Canadian Archivists, 1985), p. 7.

12 *Ibid.*, p. 7.

13 *Ibid.*, p. 8.

14 *Ibid.*, p. 8.

15 Sue Dodd, *Cataloguing Machine-Readable Data Files — An Interpretive Manual* (Chicago: American Library Association, 1982).

16 The points raised in this section are based upon Gina LaForce, "Archives and Copyright in Canada: An Outsider's View," *Archivaria* 11 (Winter 1980-81), pp. 37-51.

17 William A. Fenwick and Gordon K. Davidson, "Use of Computerized Business Records as Evidence," *Jurimetrics Journal* 19 (1) (Fall 1978), pp. 9-27.

18 Kenneth L. Chasse, "The Legal Issues Concerning the Admissibility in Court of Computer Printouts and Microfilm," *Archivaria* 18 (Summer 1984), p. 168.

19 "Transborder Data Flow: Its Environment and Consequences," in *Transborder Data Flow Policies: papers presented at the IBI Conference on Transborder Data Flow Policies*, Rome [Italy], 23-27 June 1980 (New York: UNIPUB, 1980), p. 580.

20 Marguerite A. Vogel, "Transborder Data Flow: A Canadian Focus" (M.A. thesis, Simon Fraser University, 1984), p. 24.

21 Naugler, *The Archival Appraisal of Machine Readable Records*, pp. 88-90.