ARCHIVAL THEORY AND MACHINE READABLE RECORDS:

SOME PROBLEMS AND ISSUES

By

Catherine Aileen Bailey

B.A, The University of British Columbia, 1986

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARCHIVAL STUDIES

in

THE SCHOOL OF LIBRARY, ARCHIVAL AND INFORMATION STUDIES

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

AUGUST 1988

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

School of Library, Archival and Information Studies
~~Department of~~

The University of British Columbia
Vancouver, Canada

Date 11 August 1988

ABSTRACT

It is a common feeling among archivists that the basic principles of archival theory may have to be examined and redefined in light of the development of computer technology. This need exists not so much because archivists are currently faced with a new and unfamiliar medium, but because the new technology changes not only the uses made of the information but also the way in which we perceive it.

The many attempts to approach the problem in the last twenty years have tended to be focused on single archival functions rather than on fundamental archival principles, and the solutions proposed were essentially practical. This thesis takes a global approach to archival theory and tries to answer the general question: are traditional archival principles valid as a guide in the treatment of machine readable records? In order to answer this question, the thesis puts into relationship the terminologies of computer and archival science, analyses and reconciles them, and proceeds to examine the basic concepts of the nature of archives and records, their life cycle, their appraisal, arrangement and methods of communication, and studies their application to machine readable records. The conclusion of this study is that the theoretical foundation of archival science is valid for the management of all archival documents regardless of their physical medium, and that any differences in treatment are a result of practical application of the theory. However, the development and diffused use of computer technology have opened new areas of concern to all archivists which must be explored from a theoretical perspective. Archival science does not, therefore, need to be redefined, but merely expanded.

Table of Contents

List of Figures

Acknowledgements

Several people have made the production of this thesis a much easier task. Very special thanks go to my supervisor Professor Luciana Duranti for all of her input, editing skills, and solid support; her first year in the program made my last year memorable. Thanks also to my classmates, particularly Valerie Billesberger, who offered a wide range of helpful comments when I presented some of my ideas in class. Finally, to my husband Doug, who patiently helped me reconcile computer science and archival theory on more than one occasion.

Introduction

> Many archivists ignore the new technology, hoping to retire before it invades their archives. They may be right on an individual basis; collectively, we cannot ignore the monster in our midst.
>
> Trudy Huskamp Peterson
> American Archivist 1984

Modern society has undergone many changes since the end of World War II, none more momentous than the invention of the computer. From the first vast, room-sized machines dedicated to the 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 private and government record keeping in particular is so all-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 papers 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 in the intellectual control of its holdings, and through the appraisal, acquisition, preservation and communication of the results of computer activities, more commonly referred to as computer or machine readable records.

Twenty five years ago, the thought of actually acquiring machine readable 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

tools designed for rapid processing of information and machine readable records transitory records created in the processing stage between paper input and paper output. Because of the transitory nature of machine readable records only the paper input and output were accepted in 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. Machine readable 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 machine readable records vanished or were destroyed before the archival profession began to grapple with the issue and during the first phase of orientation and enquiry. Confusion over the methods of handling computer records for archival purposes predominated, fueled by the strong divergent opinions expressed by authoritative representatives of the archival profession. Some archivists argued that the magnetic medium was 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. Others expressed the belief that computer records were of the same nature as paper records and could therefore be treated according to the same theoretical principles. 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 machine readable 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 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 machine readable records on archival functions from appraisal to reference with the result that many key issues still remain unexplored and neglected.

The root of this confusion is in attempting to find solutions to questions or problems which are unclear themselves. Only when a specific question has been determined and outlined 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, we need to provide answers to a number of more specific questions such as: how does the physical medium of the record affect appraisal, arrangement, description and access? What role does records management play in the treatment of electronic information? In order to decide if archival theory as it exists now 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 machine readable records.

The object of this study is to identify problems arising from the application of archival theory to the treatment of machine readable records and to establish whether those problems need to be solved through a redefinition of archival principles or through an adaptation of archival practices. Chapter One examines the

definitions of key archival science terms and compares them with homonyms and similar terms found in computer science in order to define a standardized, common terminology for further discussion. Chapter Two focuses on the development of appraisal criteria for computer records and compares them with those used for traditional paper records. Chapter Three deals with the relationship between records management and archival theory as to the scheduling of electronic information. Chapter Four looks at the archival practices of arrangement and description of machine readable records, and compares the archival principles governing them with those applied by librarians for the control of machine readable data files (MRDF). Finally, Chapter Five outlines three separate issues related to the dissemination of the information contained in computer records: reference services and access, transborder data flow, and the use of computer records and printouts as evidence in court. Emphasis is placed on the Canadian archival scene and particularly on the work of the National Archives of Canada[1] wherever possible. Terms in italics are defined in the glossary located in Appendix A following the conclusions.

This is by no means *the* definitive work on the application of archival theory to machine readable records. It is designed to focus archivists' attention on those problem areas needing work, and to stimulate further discussion and revisions within the profession as a whole. Such discussion is vital if we are to deal successfully with machine readable records before they are lost or destroyed. The time has come that we can no longer afford even to pretend to ignore the "monster in our midst."

---

[1] The Public Archives of Canada became the National Archives of Canada as a result of the National Archives of Canada Act, proclaimed 11 June 1987. Throughout the course of this thesis, I will refer to the Public Archives of Canada (PAC), unless the work cited or event was published or occurred after June 1987.

Chapter One: A Discussion of Terms and Concepts

Archival science is not characterized by clarity of terminology or easy communication of ideas. Many archival terms which are in current use have roots which stretch as far back as the record keeping practices of the Ancient World, and over the intervening centuries, they have reflected the constant changes in archival theory. Archival science has also borrowed terminology from other disciplines whenever the need arose.

Like any other specialized vocabulary, archival terms rarely retain the exact meanings given to them when they were created. Due to the long history and constant evolution of archival science, it is also not uncommon to find a new term being used in North America which describes a concept already well-known throughout Europe. Clearly, if the terminology is fluid and ever-changing, discussions of archival theory can easily lead to confusion and misunderstandings. As archivists can become confused in dealing with the terminology of their own discipline, it is not surprising to find that they have problems in dealing with computerized information, where they must also contend with the language of computer science. The causes of these problems are found both in the history of archival terminology and in the interaction of archival terms with computer science terms.

Archival terminology has always been a flexible entity largely dependent upon social context. The term "archives", for example, could mean very different things depending upon the use of records in each particular culture. The ancient Greeks, for example, considered that archival repositories had political and judicial purposes. Archives consisted in large part of stone or metal tablets upon which the

laws of the state, the treaties, and any information which was public by nature had been inscribed and displayed to the citizens. The ancient Sumerians, on the other hand, accumulated archives for administrative or housekeeping reasons in addition to the political ones, and gave them a patrimonial character. They preserved a wide variety of records generated by the king's bureaucracy, including tax records, land records, and evidence of administrative actions, so that the government could function more smoothly.

While these examples illustrate different perceptions of archives developed in two totally separate cultures, it is equally valid to suggest that the evolution of archives and the terminology of archival science are both closely linked to the social, political and judical changes within one country. Consider the development of archives as an institution in France. During the feudal period, when there was no central government, documents served as proof of rights, privileges, and power, and were therefore carefully preserved either by the rights holder or someone designated by him, usually a priest. Archives were treasured and passed down from father to son like any other possession, such as a parcel of land. Later, as central authority coalesced in the form of an absolute monarch, the need to preserve documents as proof of rights and privileges lessened. Archives became administrative in nature, and were designed to provide information to the king's bureaucracy in order to help him rule; they did, however, remain the property of the king. In the French Revolution, archives changed once again from being the private patrimony of the king, archives became the nation's heritage. The revolutionaries went even farther than that and drew a distinction between historical and current administrative archives which still exists in France today. At each of these stages, the terms which the record keepers used to refer to archival functions and material changed, reflecting the broader social and conceptual changes wrought upon archives.

Another factor which proves to be as powerful as the social context in which an archival term is coined is language barriers. The direct translation of terms from one language to another is often a major problem, because there are words that sound the same in different languages, but have totally dissimilar meanings. This problem, which has plagued archivists for centuries, is best illustrated using examples drawn from modern archival science. The English term "series", for example, refers to the internal subdivisions of a fonds. However, the French term "série", while it is the direct or literal equivalent of the English word, refers instead to each class in which the total holdings of an archival repository are subdivided according to a general alphabetic classification scheme, each representing a group of fonds. An example would be Série U in French departmental archives, where one would find all the fonds received from the law courts. A French "sous série", on the other hand, would generally correspond to a fonds, while the English "sub series" would mean a further breakdown of the fonds into its constituent groups of files. Similarly, in English, "finding aid" is a general term for the tools used to facilitate access to the materials, while the French, who have no such general term, would call them "guide", "répertoire", "inventaire sommaire", "inventaire analytique", or "état sommaire", depending upon their level of detail.[1] As a determinant of the terminology of archival science, language cannot be ignored or underrated.

In North America, archival science and its terms have been affected by the historical, cultural and linguistic factors which shaped European experience, but the effects of these factors has been concentrated in a much shorter period of time. The few centuries of North American history, compared to the thousands of years of European and Middle Eastern societies, has meant that North American archivists

1 Michel Duchein, "Theoretical Principles and Practical Problems of *Respect des fonds* in Archival Science," Archivaria 16 (Summer 1983): 80-81.

have had to rely greatly upon theories and terms already developed and brought by settlers and immigrants to Canada and the United States. North American archival science is therefore very much a prefabricated and transplanted discipline which has developed in different directions to suit new conditions.

North America has for most of its history had a different kind of society than Europe has had. Europe evolved through various stages of government, from feudalism to absolute monarchies, democracy, and in some cases, totalitarianism. North America, particularly the United States, has always been a democratic society. The political and juridical doctrine dominating American society has had an effect on archival science, as exemplified by the development of the term "public document". In Ancient Greece and in Rome, that term meant that the citizens had the right to know the information contained within the document. From the Middle Ages to the French Revolution, it meant that the document had been issued by a public authority (one having jurisdiction over the land and the people). Today in Europe, the term still preserves this meaning. The North American meaning of "public document" is that citizens have the right to consult it by virtue of their ownership of that document.

The greatest problem facing North American archival science is not, however, constituted by historical or social differences between the Old World and the New, for the effects of those differences can be mitigated through the awareness of their existence and the knowledge of their nature. The difficulty lies in the terms of archival science which archivists in the same country use to communicate. Archivists today feel the need of incorporating terms from other disciplines formerly completely foreign to archives into archival science terminology in order to handle some new forms of modern records. Bringing computer science terms into the terminology of archival science can be as difficult as translating a specialized

archival term from Dutch to English.

Nevertheless the exercise must be attempted, for any discussion of the application of archival theory to computerized information must involve the merging of computer science and archival science terminology to a certain degree. By comparing the key archival terms of "records", "archives", "file", "item" and "document" with their homonyms from computer science, and by examining the concepts contained within each term, we can establish some clear definitions on the basis of which we can discuss the application of archival theory to computerized information.

The definition of the term "records" constitutes the fundamental concept of archival science, and as such, it is the starting point for any archival theorization. By defining what a "record" is archivists essentially describe what is "society's documentary heritage". There are several definitions of records, but the most accepted in North America is that given by American archivist Theodore R. Schellenberg in his 1956 work Modern Archives: Principles and Techniques.

According to Schellenberg, there are three main elements to take into consideration in defining a record, the first being physical format. Schellenberg considered that as far as an archivist was concerned, the information contained within a record was much more important than its external aspects. He maintained that books, papers, maps, photographs, and any other "documentary materials" are equally eligible to be records. Thus, even though he did not specifically mention information held in electronic or computerized form, such forms should not be excluded from the definition.

The second element concerns the manner in which the information came to be recorded, that is, the circumstances of the record's creation. Schellenberg declared that records are documents which have been either accumulated or created by an agency or person in order to accomplish a practical activity. Records are therefore recorded evidence of completed transactions.

The third element concerns preservation. If materials are to be considered records, Schellenberg argued, they have to be preserved by the creator for his own reference, that is, to assist him in his functions by providing him with the evidence of his past actions, not to provide information of a historical or cultural nature to other users.[2]

Implicit in the three elements of Schellenberg's definition of records are two other concepts which are of importance. One stems from the assertion that records are created or accumulated during the course of any action which accomplishes a purpose. Records are not only evidence of a transaction, but also an integral part of it because they are the means to carry it out. Without the information they contain, the action could not take place. A record is therefore inextricably tied to the action which determines its creation.

A direct consequence of this concept is the notion of the stasis of the record. Once the information has been recorded and through such recording the action is completed, a record becomes a static element; information, form, and medium are held together as a single unit, and they are as invariable as they are inseparable. As evidence of the transaction, that unit cannot be changed in any of its parts, because if it was later altered, the record would become a forgery. We can, however, use the

---

[2] Theodore R. Schellenberg, Modern Archives: Principles and Techniques (Chicago: University of Chicago Press, 1956; reprint ed., Chicago: Midway Reprint, 1975), p. 13,16.

three elements separately; the information can be used in another activity, just as the form (in the case of guide letters) or the medium (through erasure) could be. What results is a *new* record, completely and totally separate from the first; we shall see later on in the chapter that such is not the case with electronic information.

Now let us compare the archival term records with the same term as it is used in computer science. In computer science, "record" is short for logical record, a term which describes a storage unit for computerized information. It is

> a unit of *data* ... forming the basic unit of a file... and consisting in turn of a number of inter-related data fields.[3]

Since this term includes only the means of storage of information, any computer scientist using it would probably not be at all interested in the reasons for which the information was gathered, whether it done during the completion of an official function of the record's creator or not.

Many archivists may argue at this point that the two terms cannot be reconciled, and that the use of one or the other must be discontinued if we are to deal with computer materials without confusion. The two terms are not, however, as irreconcilable as they first seem. In fact, they can be used together. If we ignore for the moment the linguistic convention that states that a term may not appear in its own definition, we may refer to computer materials in the most general sense of "recorded information" as a whole of machine readable records, which are in turn composed of (logical) records. The terms are complementary because the former refers to the function of the computerized information and the way in which it was generated, while the latter gives an indication of the form of its physical storage in

---

[3] Peter Walne, ed., International Council on Archives: Dictionary of Archival Terminology (New York: K.G. Saur, 1984), p. 136.

the computer. An example of this could be found in a social services ministry. All of the information contained within the ministry's computers is created during the course of the practical activities of the social workers and other employees, and therefore can be called "machine readable records". The information which makes up a specific letter, memorandum, case history or report, is stored in the computer in the form of a number of logical records consisting of *data elements* and *fields*. Often a single logical record will contain the contents of an entire record; this is how a machine readable record in the archival sense of the word can also be a computer science record.

The main problem in attempting to reconcile these two homonymic terms arises when the people discussing the subject do not make clear the sense in which they use them, either archival or computer science. This kind of communication difficulty occurs when an archivist speaks to a computer scientist, or even a "new" archivist fully conversant with the language of computer science talks with a "computer illiterate" traditional archivist. In both cases, the parties involved will fall back on their own discipline or personal knowledge in order to provide meanings for terms unfamiliar to them, unless they are directed to do otherwise.

If we accept that computerized information can legitimately be called machine readable records when it is generated during the completion of a transaction, we are brought into a second area of possible conflict between the disciplines. I already pointed out that, according to the implications of Schellenberg's definition of the term, records are static entities; once the transaction is completed and the information has been recorded, the information, form and physical format become a single unit which cannot be changed without destroying it as evidence of that transaction. Such is not the case with computerized information. It is very easy to alter electronic information at the touch of a few buttons; in fact, it is this ease of

manipulability which accounts for the computer's increased importance in modern record keeping. Computer records cannot be considered static entities because of the way they are structured, a point which will be discussed in greater detail later in this chapter in connection with the term "file".

Having determined that it is quite possible for information stored in a computer to be considered records in the archival sense of the term, we can now consider what records constitute "archives".

In his Manual of Archive Administration (1922), the English archivist Sir Hilary Jenkinson defines archives very carefully. First, he defines the constituent parts of archives: documents. According to Jenkinson, a document is any manuscript in any form regardless of physical characteristics of either the medium or the script. Therefore, he says, archives are any such document drawn up in the course of an action or transaction of which it formed an integral part and subsequently preserved for the information of the document's creators and their legitimate successors.[4]

Jenkinson's definition of archives is essentially Schellenberg's definition of records. In fact, the two definitions share three main elements: disregard for physical format, creation during the course of a transaction, and preservation for future use. However Jenkinson adds two concepts to his definition of archives that are not found in Schellenberg's definition of records. The English archivist states categorically that "Archives were not drawn up in the interest or for the information of Posterity."[5] This means that the information was recorded in order to bring an action to completion, not to give account of the action to history

---

4 Sir Hilary Jenkinson, A Manual of Archive Administration, 2nd Revised Ed. (London: Percy Lund, Humphries & Co. Ltd, 1965), p. 6-7.

5 Ibid., p. 11.

and scholars of the future.

Jenkinson further states that all archives share two qualities. One of these derives directly from the fact that archives are not written for posterity. Because they were not created with the purpose of influencing future perceptions of the action, archives are imbued with the quality of impartiality; that is, they can convey only the truth to people who later use the documents for purposes unenvisioned by the creators. The other quality extends from the preservation of documents. Jenkinson says that since archival quality is dependent upon providing possible proof of an unblemished line of responsible custodians, archival documents have inherent authenticity as to their creators. Because they are by definition "properly" preserved, they are free from any suspicion of tampering.[6] If a document is later found to be a forgery, we can be sure that it was the creator who did it.

Jenkinson's definition may be difficult to apply fully outside the juridical context of the English society. Nevertheless, it has had an effect on what North American archivists have preserved, since many of them are familiar with Jenkinson's ideas. Schellenberg, however, felt that in light of the tremendous increase in modern public records since World War II a new definition of archives was needed, so he proceeded to offer a definition based on his definition of records. The two definitions are in fact complementary; the definition of archives would be misleading in the absence of the one for records. Schellenberg states that there are two components to his definition of archives: the identity of the record creators, and the reasons for which the records are preserved. It is in the elaboration of the values records must have to deserve preservation that he diverges from the Jenkinson definition of archives and his own definition of records. In order to become archives, he maintains, records must be preserved for reasons other than those for

[6] Jenkinson, Manual, p. 12-13.

which they were first created or accumulated, and those reasons may have a legal-administrative and/or a cultural character.[7]

Here then is the fundamental difference between Jenkinson's and Schellenberg's definitions of archives. Jenkinson, by defining archives, not records, implies that every document created during the course of a transaction has an archival nature since the moment of its creation. Schellenberg, on the other hand, argues that records *acquire* archival quality when they have been consciously selected by an archivist for preservation in an archival institution on the basis of their research values. He *implies* that records can only become archives when they are no longer current.

Whichever definition of archives we choose to accept brings us into conflict with computer science, where the term "archives" has a very different meaning. To a computer analyst, archives are periodic copies of the information present in a computer system, copies taken to serve as backups in case of accidental damage or erasure. There is no selection on the basis of research value as Schellenberg prescribed, nor is there any concern as to whether the information was created in the course of a transaction as Jenkinson defined. When a computer facility experiences a power failure, or a malfunction in the computer's *hardware* or *software*, the information stored in the computer at the time is sometimes damaged or even erased. The systems analysts therefore take the most recent "archives" of the system and re-mount the information onto the computer so that the users may continue to carry out their functions. Clearly, if the *data* in the backups can become active again in the event of system failure, it cannot even be considered "noncurrent records". In addition, if we persisted in calling

---

[7] Schellenberg, Modern Archives, p. 16.

computerized information archives in spite of this point, we would be faced with the old problem of confusing the holdings with the term for the building or the agency responsible for their preservation.

It is interesting to note that in the Society of American Archivists' "A Basic Glossary for Archivists, Manuscript Curators, and Records Managers" (1974), the authors appear to have avoided the whole issue by designating computer data "machine readable records/archives."[8] Their apparent unwillingness to decide for one of the two, records or archives, makes this designation both cumbersome and confusing. It combines all of the terminological problems already outlined about records and archives, without making any attempt either to solve the problems, or even to indicate that they exist.

The terminological difficulties between archival and computer science are not limited to the terms "records" and "archives". The term "file", too, has a dual meaning, depending upon whether it is used in an archival or a computer science sense.

Once again Schellenberg provides a useful definition. According to him, file is a term of classification used to describe a group of records on the basis of the function which they carry out. Usually a separate file "unit" or folder is established for each transaction, whether it relates to a person or institution, place or event.

The material within a file has a common root; all of the individual records relate somehow to the person, institution, place, or topic which is the subject of the transaction. Files may also, in certain cases, have more generally stated relationships such as when an action involves not one but a whole class of people,

---

8 Society of American Archivists, "A Basic Glossary for Archivists, Manuscript Curators, and Records Managers," American Archivist 37 (3) (July 1974): 425.

or when an important event affects many people. In those cases, files are usually designated according to the transaction itself, rather than the people involved. Files are considered the smallest grouping of interrelated records. They can be grouped into larger (unnamed) units relating to the activity carried out, and still later, these larger units can be grouped into series according to function. A graphic example of this is seen in Figure 1.

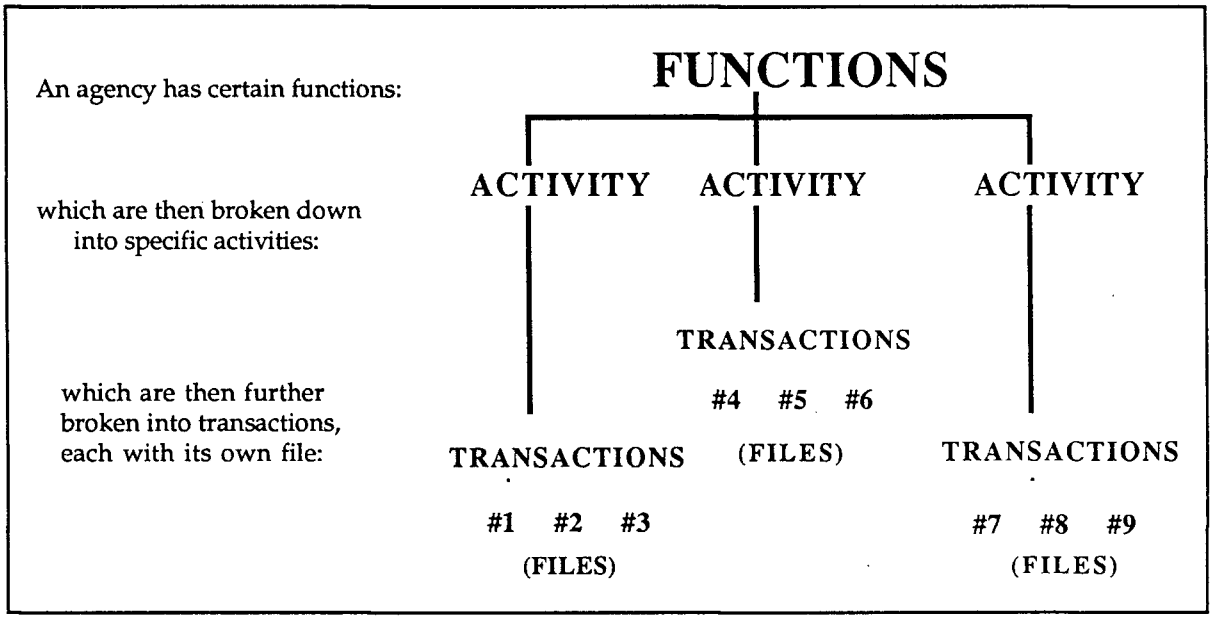

Figure 1: A Graphic Representation of Schellenberg's Definition of "File"

The SAA Glossary takes Schellenberg's "common root" concept a little farther to state explicitly that a file is maintained in a "predetermined physical arrangement". This point proves to be very important when it comes to the computer science definition of file, and for that reason, it deserves a closer examination.

Predetermined physical arrangement is really both physical and conceptual. Each paper record is put into a particular file folder according to an order which will

facilitate retrieval of the records. The arrangement could be as simple as placing the most recent document at the front of the folder, or it could be more complex, such as linking related records in a specific pattern as it is done with some medical records.

Conceptually, the placement of each record in the file links that record with the others preceding or succeeding it within the folder. It creates an immutable relationship which is further enhanced by examining the function each record carries out with respect to the others. This linkage is a necessary and vital part of the file because it is the fixed interrelationships between the records that give the file its archival qualities. We can learn something about the agency or person who created these records simply by looking at the physical placement of the records which mirrors the links between their contents. The informational content of the file is therefore more than a sum of the information contained in its constituent parts; it is provided by both the records of the file and the type of relationship which exists between them. If the relationships between the records were ever altered, we would lose a great deal of information.

Let me suggest an example to illustrate this point. Supposing there were a file containing the case history of four people seeking welfare benefits. The records are arranged first into four sections, one for each person, then within these sections they are arranged chronologically to show how each person came to need welfare. If we were to rearrange these records, grouping them perhaps according to common work experiences, or living places, or the number of dependents, we would no longer have that same progression of records which composed the original file, and the information inherent in the original arrangement would be lost. A file in the archival sense of the term is, therefore, a static element just as the archival record. Its most important quality is the arrangement of the constituent records because the

relationships between the parts gives the file its meaning.

Again, the archival term is in direct contrast to the computer term. Like a logical record, a computer file (sometimes short for datafile) is a term used to describe the means of physical storage of information within a computer. Specifically, it is constituted by "two or more data [logical] records treated as a unit."[9] Imbedded within this computer science definition is the concept that the constituent parts of the file, the logical records, are not at all bound by any static physical relationship to each other.

As was said earlier in the chapter, computer records cannot be considered static elements because their contents can be changed so quickly and easily. That is, after all, one of the reasons computers have become so important in this age of information. Equally, in a datafile, information can be and is changed and rearranged to fill new needs.

This means that because the records can be changed around so easily, there can be no information in the relationship between records due to their physical position as there would be in a paper file folder. Physical arrangement, the first element of the archival definition, has no relevance in computer science. Instead of set physical arrangement, indexes are used to locate and arrange records from any part of the file according to the instructions stored within the index, which can be created by any user of the information. Therefore, a new arrangement, or the addition or deletion of an individual record has no effect on a machine readable datafile, while it would seriously damage a traditional paper file.

Clearly, what follows from this point is the observation that computer files

---

[9] SAA, "Glossary," p. 2.

cannot be considered archival unless we can find some way to fix the relationships between individual records. This can in fact be done. If we were to take an historical snapshot of the information in a system at a specific time (what computer analysts refer to as "archiving"), we freeze the position of the records and the information deriving from it. The information in the records can conceivably become re-activated, as was noted before in discussing the computer science definition of "archives". The records themselves, however, always remain inactive because the relationships between them are frozen during the "archiving" process. By ensuring that the index used to find and order the records is accurate, we can create something in machine readable form which resembles a traditional file.

These indexes play a very large role in the use of computer files as archives. When we look at a traditional file, all the information we need is contained within that physical entity. The physical arrangement we observe tells us about the conceptual relationships between the records, while an examination of the records gives us information about their subject. To get the same information in a computer setting, we must look at two or more separate physical entities together: the datafile and its index(es).

It is actually the index, not the file, which defines the relationships between the records in the file. Because it takes each record from storage and puts it in its proper place according to its instructions, an index could be considered like the arrangement of a traditional file. Assuming that there is more than one index for a datafile, it is almost as if we are faced with several different files containing the same information within the records, but providing different information as collective entities because of the different arrangement of the records. As an example, think of the datafile which the personnel department of a large corporation would use to

keep track of the employees. The logical records in the file contain data elements for name, address, age, sex, physical characteristics, and length of employment. One index could be used to order the records by surname; another could order them by address, while yet another could order the information according to the length of employment. Thus, we have a datafile with at least three arrangements and potentially thousands of interrelationships. In addition, that file could easily be considered archival if it is removed from active use as an historical snapshot. Each of these different arrangements naturally gives a different slant to the information in the file; in order to achieve the same thing with paper, we would need to have copies of the file, each physically arranged a different way. One computer file may therefore correspond to an indefinite number of paper files.

The last terms which need to be examined with regards to computerized information, document and item, do not actually appear at all in computer science. Documents and items are generally spoken of in conjunction with each other, and even in the same context as a record. The SAA Glossary's definition of document shows how close the terms are:

> recorded information regardless of medium or characteristics. Frequently used interchangeably with record. (2) A single record or manuscript item.[10]

The definition of an item, the "smallest unit of record material,"[11] combines with the definition of document to suggest two points of discussion regarding electronic information.

First, a document is such independently of its medium. This harkens back to Schellenberg's definition of records and Jenkinson's definition of archives, and

[10] SAA, "Glossary," p. 421.

[11] Ibid., p. 424.

reinforces the point that the physical form of a record is irrelevant in archival theory. Secondly, if we accept that a document and an item are essentially the same thing, then we must consider that documents are the smallest *indivisible* unit of record material. As I have said before in discussing the static nature of the record implied by Schellenberg's definition, the completion of a transaction results in the creation of a single entity. The information, the form of its recording, and its physical format or medium are locked into a unit. This unit cannot be broken down any further because to do so would destroy the document's value as evidence of action by removing part of the information necessary to its understanding. A single paragraph from a memorandum may convey the *sense* of the whole, but it cannot be used in place of the entire memo with any sort of accuracy or authenticity.

On the contrary, a computer record can be designed to be broken down farther than that and still be useful. Following the idea that an item or document is the smallest indivisible unit of record material that documents a transaction, its equivalent in computer science is extremely variable. If we were talking about a single memorandum or letter, the computer term used could be a logical record. If our item was a bound volume containing the correspondence register of an agency with one entry for each piece of mail received, the equivalent could be a datafile.

The reason that this confusion exists is that the "smallest indivisible unit" is not constant in computer terms. While it would destroy the value of a paper record to have it subdivided into smaller parts, a computer record can usually work just as well when it is broken into fields, data elements, *characters, bytes* and even *bits* as when it is whole. Computers have an ability to deal easily with small portions of information through the use of indexes that paper records lack.

What should be apparent from all this discussion of terms in archival and computer science is that archivists cannot hope to bring computer materials under control until they are certain of what they are talking about. Since there is obviously a fair degree of confusion in terminology, archivists must make an effort to define their technical vocabulary. There have been several attempts to standardize the terms used in archival science so that archivists can communicate amongst themselves, perhaps even without the problems brought on by cultural and linguistic differences. The ICA's Dictionary of Archival Terminology and the SAA Glossary are two examples of this trend.

When it comes to the terms used in archival science with regards to machine readable information, however, the situation is not so encouraging. Instead of attempting to reconcile archival science terminology with that of computer science by translating the latter into the former on the basis of an examination of equivalent concepts, archivists have simply merged the two vocabularies and confused them. Terms are created for computer material which retain enough elements of archival terminology to reassure the traditional archivist, yet they also introduce elements of computer science in order to deal with the qualities of the new medium. The result is a great deal of the confusion as the different concepts of two disciplines are brought into a single term. The SAA's definition of "machine readable records/archives" is a very good example of the attempt to create a useful term which instead simply increases communication problems and general confusion.

This is a complex, involved and often confusing area of study. I have examined several key archival terms and attempted to reconcile them with the same terms and with correspondent concepts found in computer science. As a conclusion, I offer some clarification of the meanings of the terms which will be used throughout

the course of this study.

In archival science, machine readable (or computer) records are to be considered any information recorded in a machine readable form during the course of an official activity or transaction. These records must form an essential part of the transaction and subsequently be stored for future use. Records may be physically stored in a computer at various levels, from a data element or a field, to a logical record or even a datafile. What is important is the purpose of information creation; the means of its storage is of secondary concern.

Machine readable archives, on the other hand, will be taken to mean those computer records which have ceased to be in current use and are therefore removed from the system along with their respective indexes, thereby fixing both the content of records and the relationships between them.

An archival computerized file is a complex or whole of archival records stored in machine readable form which documents one single transaction, and those indexes which determine the arrangement of the file's constituent parts. A computerized file could be as small as a single logical record, or as large as a datafile containing as many logical records as a series of paper records has folders. Throughout this study, I propose to avoid the use of the terms document and item except in a strictly traditional archival sense. The term machine readable record that I have outlined better expresses the concepts important for the examination of computerized information.

Archivists are not necessarily faced with the need to redefine their working vocabulary completely before they can go farther in handling computerized information. They simply have to become more aware that these terminological

problems exist, and attempt to express themselves more clearly. Computer science terminology will have to be better integrated with archival terms; indeed, in some areas of archival science the attempts have been relatively successful. One is the appraisal of electronic information.

Chapter Two: Appraisal

Appraisal is the process through which recorded information is assessed for preservation according to its current and lasting values, its arrangement and its relationships to other records. Many consider it to be one of the most important of all the archivist's functions; as Hans Booms points out, it is the archivist who is entrusted with assembling and preserving the recorded evidence of society's actions, its historical sources.[1] The appraisal function, he suggests, "carries the greatest social significance, and unmistakably characterizes and defines the professional image of the archivist of today."[2]

Appraisal, however, is also the most difficult of the archivist's many tasks.. Deciding what to preserve and what to discard is naturally largely dependent on the social context of both the archivist and the records; it also depends on the archivist's personal values as a result of that social context. This situation leaves the selection process wide open to charges of bias. Archivists have therefore devoted a great deal of time and energy to making the process as impartial as possible, relying on specific values the test of which can be applied by any archivist to any records.

With regards to the appraisal process itself, I believe that there are actually two different but complementary approaches, both resulting from the nature of the society which creates the records. One is the selection of documents for destruction that Jenkinson outlines in the Manual; the other is selection of records for permanent preservation as archives that Schellenberg discusses in Modern Archives.

---

1 Hans Booms, "Society and the Formation of a Documentary Heritage: Issues in the Appraisal of Archival Sources," trans. Hermina Joldersma and Richard Klumpenhouwer Archivaria 24 (Summer 1987): 76.

2 Ibid., p. 71.

In the judicial English society of which Jenkinson wrote, a great deal of emphasis was put on the role of creating agencies in the preservation of documents. As I mentioned in Chapter One, Jenkinson considered that a key element of his definition of archives was that the documents were to be preserved for the information of their creators or their legitimate successors. As a result, he argued, no one can criticize those "ancient" archival collections which have been placed in the safekeeping of the archivists of the Public Records Office for what was not preserved along with them, unless that lack of preservation violated contemporaneous record-keeping standards. Furthermore, since archives are preserved for the creators' information, an archivist can only legitimately destroy material on two grounds, both of which are difficult to determine and to apply. One is that the documents duplicate others already in existence, the other is that they are of no historical value. Jenkinson concluded that due to the biases inherent in the latter, only word-for-word duplicate documents could be safely selected by the archivist for destruction.[3]

While Jenkinson says that it is acceptable for archivists to assess "ancient" collections for duplication and hence destruction, he argues that as far as "modern" archives go, the situation is entirely different. Archivists should not and indeed cannot be responsible for the selection of modern public documents for destruction. The only legitimate place for the destruction of documents is the creator's office long before they ever reach the archival stage because only the administrator truly knows what is essential to his functions. It is therefore the role of the archivist to ensure that the creators of documents are neither destroying too many records nor preserving too much.[4]

---

[3] Jenkinson, Manual of Archive Administration, p. 139-47.

[4] Ibid., p. 147-51.

Schellenberg's position that records must be selected for preservation, not destruction, also springs from his own definition of archives, just as Jenkinson's position does from his definition. Since Schellenberg's ideas on appraisal have had such a strong influence on North American archival practice, we should look more closely at them.

Schellenberg stated that archives were those records "adjudged worthy of permanent preservation for reference and research purposes" and deposited in an archival repository. Since the records had originally been produced or accumulated for a specific purpose and then retained for purposes other than those for which they were created, he concluded that public archives have two kinds of values: primary and secondary.[5]

Primary values are the values that the records have to their creators, and usually relate to the reasons for which the records were created. Administrators generally preserve records which are of value to them in one or more of three main areas: administration, finance and law. Records with administrative value document the basic policies, decisions and actions of an agency, providing continuity to an agency's functions. Those with financial value document the agency's financial authority, its various transactions and obligations, and must be kept in order to act as verification in case of audit procedures. Finally, records with legal values are those which provide proof of an organization's rights and responsibilities with respect to the law. They may be kept in case of future legal action, or because of statutory requirements.

As long as a record is needed by its creator, it is considered to have primary

---

[5] Although the emphasis here is on records generated by public administration, it should be noted that these values also may be applied to "private" records in certain circumstances. The ideas summarized in the following pages are taken from: T.R. Schellenberg, "The Appraisal of Modern Public Records," Bulletins of the National Archives 8 (October 1956): 254-59, and Schellenberg, Modern Archives: Principles and Techniques, p. 139-60.

values, and Schellenberg, like Jenkinson, says that only the agency and its record officer can determine these values. Once the primary values expire, however, the archivist must be brought in to appraise the secondary values to determine if the records are of enduring archival value. Naturally, the moment at which this occurs is variable. In some cases, records must be retained permanently by the creating agency for legal reasons and are therefore "current" or "semi-current" despite their age. An example of this is some land title records. Conversely, some records are of short term value to the agency; for instance, some administrative policy documents become valueless for current purposes when they are superseded by new policy.

Secondary values are the values that the records have to other agencies and private users once the primary values have ceased. It is on the basis of these values, which are further subdivided into evidential and informational values, that the archivist selects records for permanent preservation as archives.

Records with evidential value contain evidence on the organization and functions of the creating agency, which can then be used by an outside researcher in order to study the agency's administrative history. Any archivist faced with the need to preserve only the barest minimum of an agency's records should first preserve those records with evidential value; without them, there would be no accurate means of reconstructing the basic organizations and functions of that agency.

Informational values, on the other hand, do not at all concern agency but focus instead on the persons, places and subjects with which the agency dealt as it carried out its functions. Schellenberg therefore suggests that informational values should be judged independently of their creator. In order to do so, the archivist must have a wide knowledge both of research methods and interests and an awareness of their

possible future trends. There are three tests which may be used to determine the record's informational value: uniqueness, form and importance.

The "uniqueness" of the record has two meanings. First, it means that the information in the record does not exist in a usable or complete form in any other sources, such as published works or manuscripts. Secondly, it means that there should be no physical duplication of the record itself; in other words, only one copy of the record should be retained by the archives.

The "form" test also applies to both the information and the medium of the record. The archivist must assess the form or nature of the information to determine if it is extensive (contains many similar facts about many subjects), intensive (many facts on a few), or diversified (many different facts on many scattered subjects). In appraising the medium of the records, the archivist must choose to preserve the physical form which will provide the easiest access to users other than the creators.

The test of importance is considerably less specific than those for uniqueness and form. The archivist, drawing upon all his personal and professional experience, must decide if a record has a particular significance which makes it more valuable than other similar records. For example, a file on a prominent politician could be considered more important than that of an ordinary citizen, or a court case in which a legal precedent was set could be more valuable than a routine case of the same nature.

Evidential and informational values are by no means mutually exclusive. The evidence of an agency's actions which leads to the archival preservation of a record could at the same time provide valuable information to a researcher on a particular

subject.[6] Here what is important is that records become archives *only* when they are selected for preservation on the basis of these values so that they may be used for purposes other than those envisioned by the creators.

Schellenberg's thoughts on the appraisal process have had a much greater influence on North American archival theory and practice than Jenkinson's emphasis on the creator's role in selection of documents for destruction. This influence is a result of the nature of North American society and its archives in much the same way that archival science terminology is. Here in North America there is a much stronger cultural component to archival preservation than there is in Europe. North American archives are perceived by most members of society solely as a means of preserving the nation's history or its "cultural heritage". Only secondarily, it seems, are archives considered the arsenals of law and administration as they are in Europe. For this reason, North American archival appraisal puts a great deal of emphasis on the secondary values of archives.

One important fact which we must not ignore when trying to apply Schellenberg's appraisal techniques to machine readable records is that they were originally designed primarily for paper records. True, according to Schellenberg, records and therefore archives could be any medium; Jenkinson, too, argued that the physical form of a document was irrelevant. Nevertheless, machine readable records have long suffered under the misapprehension that they are not really materials suitable for permanent preservation. Only within the last twenty years has the situation changed.

In his article "The Evolution of an Appraisal Theory for Automated Records" (1987) Thomas Elton Brown explains that the question of the long-term value of machine readable records was raised in the United States National Archives and

---

6 Schellenberg, "Appraisal of Modern Public Records," p. 59.

Records Service[7] (hereinafter referred to as NARS) long before the computer was even invented. In 1936, NARS concluded that the punched cards generated by the Bureau of Census were "not record material" and could therefore be destroyed without hesitation.[8] Therefore, until 1956, the NARS policy on machine readable records was to destroy them all without question.

In 1956, Schellenberg wrote an internal report called "The Appraisal of Modern Public Records." In it, he argued on the basis of his definitions of records and archives that machine readable punchcards were in fact records since they contained valuable information despite their medium. He immediately negated this position, however, by arriving at the conclusion that notwithstanding the value of their contents, machine readable records could not pass the form test. Saying that records preserved in archives should be in a form which would allow users to consult them easily and without any need for special equipment, Schellenberg concluded that "punchcards and paper recordings are commonly unusable without resort to expensive equipment"[9] and should therefore not be preserved.

With this statement, Schellenberg essentially gave archivists carte blanche to destroy all machine readable records. At the same time, however, he felt that the possible re-aggregation or re-analysis of old data on persons, subjects and other phenomena allowed by them meant that machine readable records could have very high informational values. Yet as before, he came to doubt their usefulness and endorsed the policy of wholesale destruction on the grounds that

---

7 Like the National Archives of Canada, the National Archives and Records Service of the United States has recently undergone a name change; since 1984, NARS has been known as the National Archives and Records Administration (NARA). Throughout the thesis, the acronym NARS will be used unless the work cited or event was published or occurred after 1984.

8 Thomas Elton Brown, "The Evolution of an Appraisal Theory for Automated Records," Archival Informatics Newsletter 1 (3) (Fall 1987): 49.

9 Schellenberg , "Appraisal of Modern Public Records," p. 257.

> if a Government agency that created the records for statistical purposes did not fully exploit them, it is hardly likely that anyone else will; for scholars outside the government do not ordinarily have the resources for the costly exploitation of such records.[10]

Schellenberg's attitude and his conclusions may seem terribly wrong and short-sighted to many present-day archivists dealing with computer records, but to dismiss them summarily would be equally wrong. When he expressed those thoughts, Schellenberg could have had no idea how versatile, inexpensive or widespread computers and their records would become. The fact that he defined records (and hence archives) as any documentary material "regardless of physical form or characteristics" meant that he was far-sighted *for his time.* In applying his appraisal criteria to machine readable records, we should always attempt to keep them in the proper context.

The fact remains that until the early 1960's, NARS policy on machine readable records was one of complete destruction. At that time, the appraisal treatment of machine readable records began to change. A report to NARS in 1960 suggested that computer records were not "non record material" but "interim media", a step between conventional paper *input* and paper *output.* The report's author, Richard Jacobs, believed that although machine readable records were in fact records, computer tapes, like punched cards, would "lose their usefulness in time,"[11] hence their designation as "interim". How these cards would lose their usefulness is not made clear; nevertheless, Jacobs urged the preservation of essential information in paper rather than electronic form.

This position was the last major stand against the preservation of machine

---

10 Schellenberg , "Appraisal of Modern Public Records," p. 274.

11 Meyer Fishbein, "Appraising Information in Machine-Language Form," American Archivist 38 (January 1975): 37.

readable records in their original electronic form. Many researchers at that time actually wanted to convert conventional records to electronic form for easier analysis in statistical research. Archivists began to realize that not only were there more and more records being produced in machine readable form, but that machine readable records actually had advantages over paper in terms of storage space and ease of manipulability.

Nevertheless, it was not until 1978 that an archivist came up with solid guidelines for the appraisal of machine readable records when Charles M. Dollar's pioneering work "Appraising Machine-Readable Records" appeared in American Archivist. While the ideas in the article were not exactly new in themselves, Dollar was the first to articulate a clear and concise method for appraising the new medium. He treated it as a logical, step-by-step process, beginning with some technical considerations such as the readability of the data, before addressing more traditional archival considerations. He refined this appraisal process into a flowchart, which is reproduced in Figures 2 and 3.

Dollar's work was timely and of great practical value to archivists struggling with the beginning of the flood of machine readable records. Unfortunately, in his discussion of the values present in computer records, he seriously undermined their position as materials worthy of archival preservation. Dollar argues that since very few machine readable records "impinge on legal rights or document significant agency decisions and programs accomplishments,"[12] they should be appraised solely for their informational values to secondary users. This implies that machine readable records can only be used for research purposes and can therefore be

12 Charles M. Dollar, "Appraising Machine-Readable Records," American Archivist 41 (October 1978): 424.

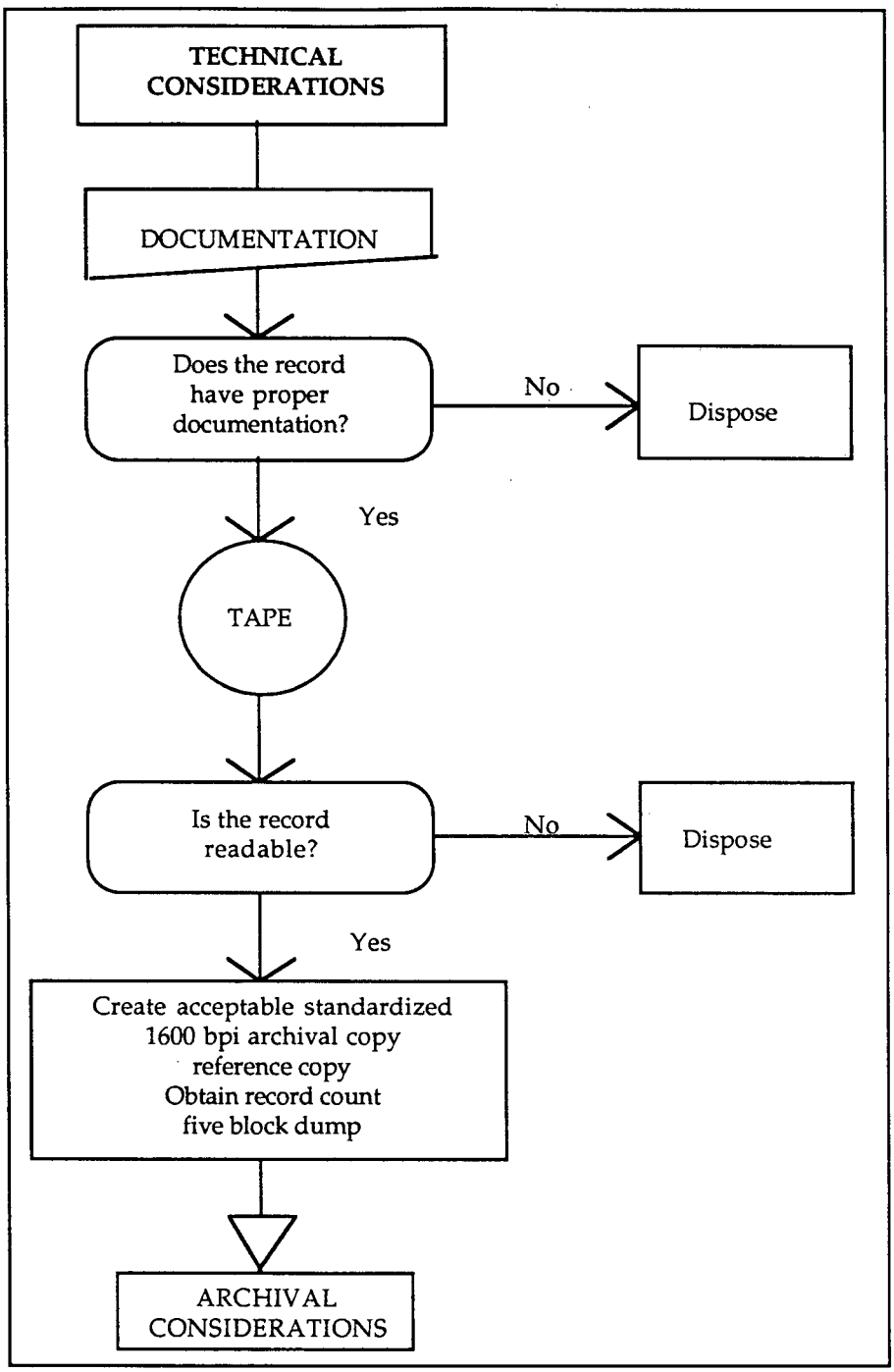

Figure 2: Technical Considerations Flowchart from Charles M. Dollar's "Appraising Machine-Readable Records" (1978)

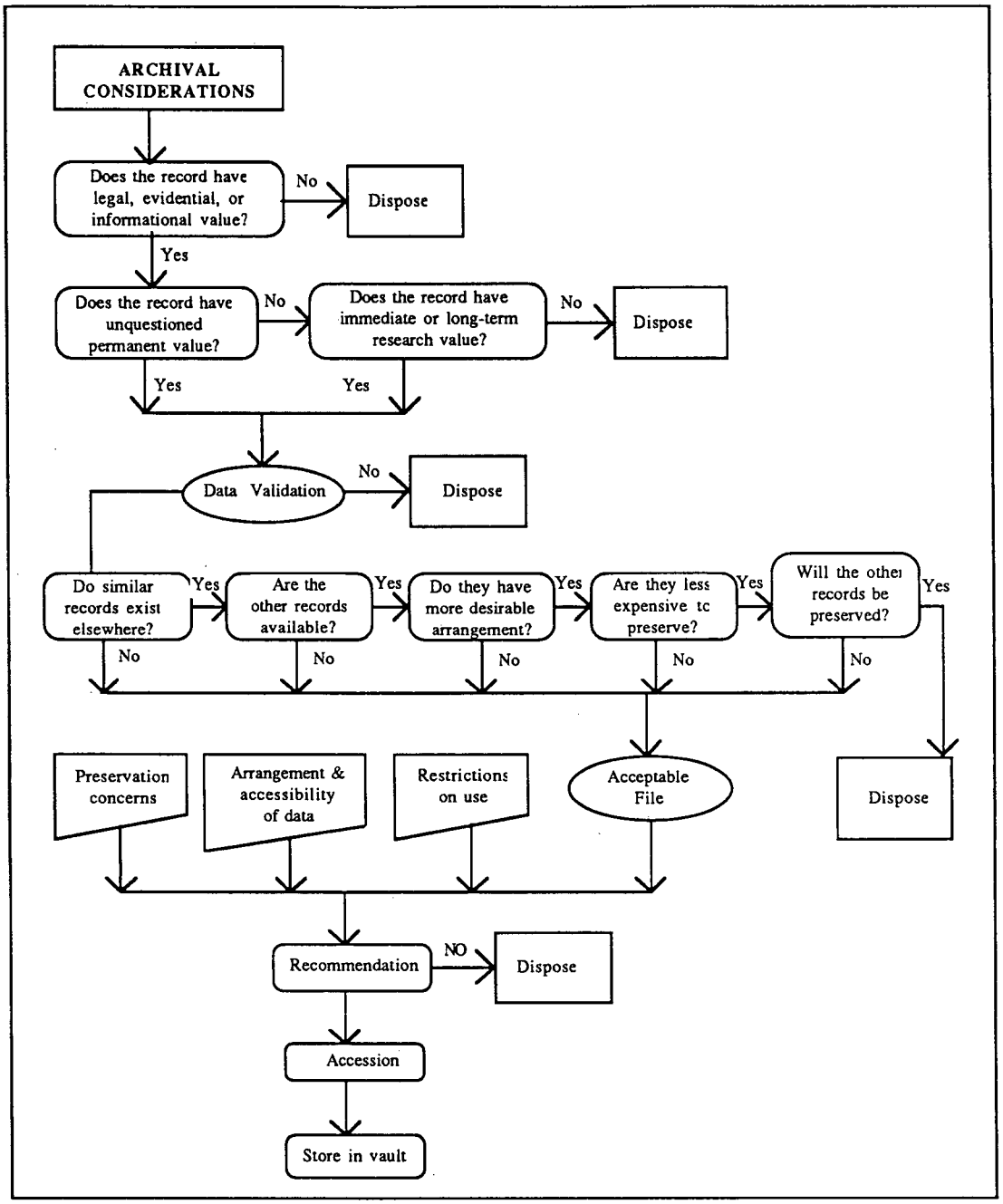

Figure 3: Archival Considerations Flowchart From Charles M. Dollar's "Appraising Machine Readable Records" (1978)

appraised without regard for their provenance, something which I think seriously compromises the appraisal process. It also subtly removes machine readable records from the collective whole of an agency's record keeping system and creates the possibility of preserving an incomplete record of that agency's functions and activities.

In 1981, however, Dollar's position was strongly challenged by John McDonald and Katherine (Sue) Gavrel of the Public Archives of Canada's Machine Readable Archives Division who issued guidelines for appraisal in their division and elaborated upon Dollar's technical analysis. They also made two important points which went against the accepted views on machine readable records. First, they disagreed with Dollar's view that computer records should be appraised only on the basis of their informational value and said that like any other record, computer records should be evaluated on the basis of their evidential and legal values as well. Secondly, they proposed that in appraising administrative housekeeping data, it was not always appropriate to destroy machine readable form because the same records in conventional, "human-readable" form would be destroyed.

Later in 1983, Harold Naugler took McDonald's and Gavrel's conclusions a step farther in The Archival Appraisal of Machine Readable Records: A RAMP Study with Guidelines. Naugler's study, which is essentially the PAC guidelines modified by information on the situation in the United States and Europe with some discussion of issues relating to appraisal of machine readable records, such as privacy and transborder data flow, appears to have been accepted as a standard by the profession. In view of this fact I shall examine more closely the points that he makes.

As I said before, appraisal of records requires that the archivist examine them for their evidential and informational values, regardless of the form of those records. As McDonald, Gavrel and Naugler point out, the appraisal process for machine readable records must, because of the medium, be broken into two component parts: technical analysis and content analysis. Content analysis equals the "traditional" process of appraisal, with some added features, while technical analysis is a detailed examination of the record's medium carried out both before and concurrently with content analysis.

There are two main components to the first stage of technical analysis: determining the readability of the data, and ensuring the proper documentation is available. Determining readability is quite simple in theory and absolutely essential in its application. The tape, *disc* or *drum* is mounted on the proper hardware with the proper software to convert the electronic information into a "human readable" form. Obviously, if there are any errors that prevent the information from being read, the records cannot possibly be appraised for their internal archival qualities.

If the data can be successfully converted into a readable form, the next step in the preliminary technical analysis is to ensure that the proper *documentation* is available. Documentation explains the various factors necessary to make the computer file(s) function properly, including the hardware *operating system* and software, the contents of the records, the arrangement of them, and the coding of the information found within them. Without documentation, a machine readable datafile can be totally incomprehensible to all but the computer programmer who created the basic structure. In order to begin the appraisal of a machine readable data file, the minimum documentation is considered to be a *codebook* and a record layout. The codebook is relatively straightforward; it gives an explanation of all the

particular codes which stand for more extensive information. For example, an "a" in a particular field could mean that the respondent in this survey lives in the province of British Columbia, while a "b" could indicate Alberta. A record layout, on the other hand, is a list or diagram of the basic structure of the record, showing how long each field is, what it contains and where it is located in the file.[13]

Once the two conditions of readability and proper documentation have been satisfied, the archivist can proceed with content analysis. Content analysis is composed of an examination of the record's evidential and informational values mentioned by Schellenberg, as well as a new one: legal value.

The evidential value of machine readable records is easy to identify. A computer record, like any other record, has evidential value if it provides testimony of an agency's existence, or evidence of its organization, functions or activities. In this respect, appraisal is identical for paper and electronic records.

The informational value of machine readable records, as with paper, refers to the usefulness to secondary users of the information contained in the records about the people, places and subjects with which the agency dealt. Again, the three tests of uniqueness, form and importance should be examined, but before an archivist can reach a final decision on the informational value of machine readable records, he should consider three more elements not often taken into account in dealing with paper records. They are the manipulability of the records, the level of data aggregation, and the linkage to other records.

Manipulability of the records refers to the ease with which the information in the records can be re-arranged in order to re-analyse it. As McDonald and Gavrel

---

13 Harold Naugler, The Archival Appraisal of Machine Readable Records: A RAMP Study with Guidelines (Paris: UNESCO, 1983), p. 146, 155.

point out, it is quite common to find machine readable records classified as having value for statistical purposes simply because of this ease of manipulability, while the same kind of information in paper form would be considered valueless because of the difficulty to manipulate it.[14]

The level of aggregation of the information has a different significance in machine readable records and in paper records. In paper records, summarized information is considered to be more useful than basic, unsummarized information, largely due to the difficulty of manipulability. With machine readable records, instead, because of the power and versatility of modern computers, non-aggregate data is considered valuable since it can be re-organized countless times in order to carry out new studies.

Finally, the linkage of records to other records is another important element in the appraisal of machine readable records. In the past, archivists have tried to consider possibilities of linking various paper records to related series while doing their appraisal, but the results were less than successful. Record linkage required a great deal of time and energy to examine the records for common characteristics, usually for the purpose of statistical research. As a result, it was rarely carried out in a formal manner. As more and more standardized data elements began to appear in computer datafiles, however, and with the speed of computers, records can be linked much more easily than ever before. For example, machine readable data files dealing with unemployment insurance could be linked with similar files kept on income taxes by comparing the fields in which a social insurance number is found. In this way, the informational value of both files is increased.[15]

---

14 John McDonald and Katherine Gavrel, Appraisal Guidelines in the Machine Readable Archives Division (Ottawa: Public Archives of Canada, 1981), p. 6.

15 Ibid., p. 6-7.

Along with their evidential and informational values, machine readable records must also be appraised for their legal value. Schellenberg spoke of legal value as an aspect of a record's primary value to its creator, but in the context of computerized information, legal value does not have the same meaning. Naugler refers to it as the implications that machine readable records have with respect to the law after they have ceased to be of importance to their creators. There are three areas to be examined here: the admissibility of machine readable records as evidence in court, their connection with copyright laws, and the restrictions imposed on certain types of records by statutes.

Few countries presently admit machine readable records as acceptable evidence in a court of law, but from all indications this is changing. In many cases, it will become unavoidable, since important records may soon be kept only in electronic form. Thus, in appraising computerized records, the archivist must remain aware of recent developments in this area of the law. Changes in the laws governing the admissibility of evidence may require such things as more in-depth appraisal, not only of the records themselves but also of their documentation in order to be able to report to the court how the data were entered, how the records were processed, and who had access to the system.[16] This puts much more emphasis on good appraisal procedures for machine readable records.

Computer records are also relevant to copyright laws because of the software programs which are used to run them. Older copyright laws were written to deal with traditional records, such as papers, manuscripts, and books, but recently, these laws have been revised to reflect the changes in modern records media. These laws include provisions for the use and re-use of computer programs, and since most if not all of the datafiles accepted by archival institutions for preservation are run by

[16] Naugler, Appraisal of Machine Readable Records, p. 38. This subject will be discussed in greater detail in Chapter Five.

various commercial software packages, archivists could find themselves in difficulties concerning the dissemination of that information. An archives could be prevented from making datafiles available to users without first obtaining the proper distribution rights to its software. If the information cannot be disseminated, then this could have a serious effect on its value.

Archivists appraising machine readable records must also consider whether those records are affected by the various statutes which prescribe that certain types of records must be preserved for a specified time period to fulfill legal requirements for the creators. These statutes, like copyright laws, are slowly being revised and often include a specific reference to EDP records. If new legislation includes computer records, stating that they must be retained for perhaps seven years for financial reasons, then this would clearly affect the appraisal of their values. The point I wish to make here is that in light of the fact that attitudes towards machine readable records are changing rapidly, not only among creators and users but also among legislators, archivists must work to keep abreast of any changes in the legal status of these records. These three facets of the legal value of machine readable records may also affect their traditional paper counterparts to a certain degree and should therefore be considered. In this respect, there is very little difference between paper records and machine readable records. It is in technical analysis where the differences become more apparent.

As I said earlier, technical analysis is really divided into two stages: one which must precede any content analysis of the evidential, informational and legal values of the records, and the other which is carried out concurrently with that content analysis. In the second stage of technical analysis, the archivist relies much more heavily on personal value judgements to arrive at an appraisal decision for machine readable records. There are three broad areas of consideration: acquisition

and processing implications, preservation concerns and restrictions. Although they are separate areas, each of the three is linked directly to the other two; an adverse decision in any one of them could conceivably prevent the datafile's acceptance for preservation.

When looking at the acquisition and processing implications of a machine readable datafile, the archivist is really appraising it in light of the institution's resources, both financial and professional. Two major considerations are the size and complexity of the file, and its degree of dependence on hardware or software. If the file is particularly large, the archivist might want to consider taking only a sample portion of it. This brings up the entire question of the effect of sampling on the informational value of the file, as well as the utility of sampling machine readable records in general, all of which I shall examine in greater depth later in the chapter. While the datafile's internal arrangement has no effect on its archival value, it does have implications for its processing. Processing consists of comparing and verifying the information contained in printouts from the datafile with its codebook and record layout from the first stage of technical analysis. In most cases, files containing standardized codes and an easily understood layout are considered more acceptable than complex, highly individualized coded files which are difficult to decipher.

Hardware or software dependency is another vital concern when assessing a datafile. When a file can only be run on a certain type of hardware system, or with a specific software package (which itself may be hardware dependent), the archivist is faced with a dilemma. If he accepts a dependent datafile because of its great archival value, he must either maintain the necessary hardware and software to run it as well, or convert the data to an independent format. Both of these options are expensive and time-consuming, and usually leave the archives little choice but to

reject dependent files. If the archives does choose to reject these files, valuable information could be irretrievably lost. Conversely, if they accept them as they are, researchers will have to deal with the technical problems on their own, and the files may never be used. Clearly there is no single acceptable solution; each archivist appraising machine readable records must consider the capabilities and limitations of his own institution.[17]

Preservation concerns also are a vital area in the appraisal of computerized records. Here, the total costs of permanently preserving the datafile are considered. Unlike paper records, the preservation of magnetic media is an active process. Since the expected lifespan of magnetic tape (the accepted means of archival storage) is estimated to be only 10-20 years,[18] constant efforts must be made to extend that lifespan through cleaning, recopying, and reformatting. A paper file, on the other hand, may be accessioned and then left to sit for years in the proper environmental conditions without any loss of information.

Magnetic tapes are not only short-lived but also extremely sensitive to dust, dirt and smoke particles. These particles may accumulate on the tape and prevent its being read by the hardware. To prevent this, the tapes have to be stored in scrupulously clean conditions and be run through a tape cleaning machine on a regular basis. Should any errors appear as a result of tape degradation, the tape is usually reformatted as soon as possible.

Magnetic media are also very sensitive to any kind of magnetic field such as that produced by an electric motor or a magnet. Since the information is magnetically

---

17 McDonald and Gavrel, Appraisal Guidelines, p. 18; and Naugler, Appraisal of Machine Readable Records, p. 59-60.

18 Lisa L. Fox, "Archival Preservation in the Age of Technology," Provenance III (Spring 1985): 25.; and Sidney B. Geller, Care and Handling of Computer Magnetic Storage Media (Washington, D.C: U.S. Government Printing Office, 1983), p. 65.

encoded onto the tape or disc, exposure to any external magnetic field may seriously damage the tape beyond repair. For this reason, the tape storage area has to be in a separate part of the archives well away from any such sources. I should also point out that there are dangers from magnetism within the tapes. Since several hundred layers of a magnetic tape are placed directly on top of each other on a reel, the magnetic signal on one layer could "bleed through" to its neighbour. If left untouched for extended periods of time, computer tapes could conceivably damage themselves as much as if they were exposed to an external field.[19] This is yet another reason for the active nature of machine readable record preservation.

Finally, apart from these specific technical considerations in the care of machine readable records, the storage conditions for paper records, which form the bulk of most archival holdings, are not suitable for computerized media. Paper's recommended storage conditions are 13 to 18°C at 55 to 65% relative humidity,[20] while those for the long term storage of magnetic tapes are 17 to 20°C at 35 to 45% R.H.[21] While it is generally recognized that paper can withstand a small range of temperatures and relative humidity, it is also observed that the most important factor in the environmental conditions for machine readable records is the avoidance of fluctuations. Sudden changes in temperature or moisture can cause stress damage to the tape and hence lead to a loss of information. For that reason, some suggest that computer tapes be acclimatized or "relaxed" in the temperature of the computer room (which is usually warmer than the storage area by about 3°C for human comfort) for at least 24 hours before they are mounted and used.[22]

The point behind all of these technical conservation details, which only begin to

---

19 Michael Roper, "Advanced Technical Media: The conservation and storage of audio-visual and machine-readable records," Journal of the Society of Archivists 7 (October 1982): 111.

20 Ibid., p. 107.

21 Geller, Care and Handling of Computer Magnetic Storage Media, p. 79.

22 Ibid., p. 79

show how much is involved in preserving machine readable records is really quite simple. Preservation concerns are perhaps the most important part of the technical analysis of every computer file. The question which the archivist must ask himself is: do the evidential, informational, and legal values of the file justify the tremendous outlay of time, expertise and funds to preserve it? If the answer is yes, then the file should be acquired; if not, it should be rejected. The same sort of decision is made with regards to paper records in need of conservation treatment; it is usually only the degree of the commitment which varies.

One final point of consideration in appraising machine readable records is the question of possible restrictions on the information. The kinds of restrictions which apply to paper records may also apply to computer records. However, restrictions put on a datafile, either by its creator or donor, or by outside agencies such as legislative bodies, are not treated in exactly the same fashion. It is possible, in cases where only certain information is restricted, to remove it and then manipulate the remainder of the datafile so that it can still be made available to users. Since many computer files contain information of a personal nature, the removal of personal identifiers, or "anonymization" is considered a viable alternative to complete restriction of the whole file. This process will be discussed in more depth later in Chapter 5 in the context of privacy and confidentiality issues as they relate to computer records. What is important for the archivist appraising machine readable records with possible restrictions on their use is to assess the impact which the removal of part of the file will have on the value of that which remains. Does the removal of personal identifiers diminish relatively the informational value of the file? Does it destroy any value as evidence in court that the file may have? If it does, the evaluation of the file will change accordingly.

These then are the components of appraisal of machine readable records: the two

part technical analysis, and the content analysis of the information. Let us now look more closely at how much the appraisal of machine readable records differs from that of paper records.

The most important point to make is that the criteria for appraisal of computerized information are very clearly rooted in traditional archival theory and practices. Machine readable datafiles are examined for their value as evidence of an agency's existence and its actions, for the value of their information to research and scholarship, and for the value of their information with regards to the law. The content analysis of computer records, removed from its companion, technical analysis, could easily be applied to paper with no ill effects to the latter.

Even if we consider technical analysis, which many "traditional" archivists may argue does not exist as such in dealing with paper, we find striking similarities between traditional archival theory and the principles governing the treatment of machine readable records. Granted, the appraisal of paper does not initially require a determination of its readability, nor does it need documentation. However, if we look at the secondary stage of technical analysis, the acquisition, processing, preservation and restriction implications, the similarities between paper and machine readable records are much more obvious. Archivists appraising computer records ask the same sorts of questions as those dealing with paper. How big is the file? How is it arranged? How difficult will it be to process it? Has the institution got the resources to store, preserve and disseminate it? Does it have restrictions that might impair its value?

The difference between appraisal of paper and machine readable records is not in the theory but in its practical application. The apparently heavy technical nature of the appraisal is due to the simple fact that a computer and its files are more

technologically complex than pieces of paper. Machine readable records require a different expertise and approach. I believe that the same *kind* of shift in thinking had to have taken place when papyrus and parchment replaced clay tablets and paper replaced parchment as the main medium found in archives, since those new media had to be treated in a completely different fashion from clay. It happens that in this case, the archivist has to go outside his field to computer scientists for assistance in the more technical aspects not found within archival expertise.

But what about these seemingly "new" concepts of record linkage, manipulability, and aggregation, some traditional archivists may ask. They played a very minor role in the appraisal of paper not because they were theoretically unimportant, but because their influence on the overall value of the records was slight. The three concepts are most important in terms of statistical research, where large quantities of data are re-manipulated by users other than those who created the records for completely different purposes. Statistical research did not become important or even feasible until the advent of the computer, whose speed and accuracy made it possible to manipulate in seconds or minutes amounts of information that would have taken humans years to complete. Before this time, there was much less emphasis on statistical research and hence less importance given to linkage, manipulability and aggregation.

There is no fundamental difference between the appraisal of paper records and of machine readable records, but there are some problems which do arise.

One of the biggest problems in the appraisal of machine readable records is determining when to appraise them. In Chapter One, I said that Schellenberg's definition of archives implied that they were constituted by inactive records which no longer had any value to the creators. I also pointed out that to computer

scientists, archives was merely a term used to describe system backups taken in case of accidental damage or erasure; the data in those "archives" could conceivably become active again in the event of system failure. If the data can be re-activated at any time, when, if ever, do the records reach the inactive stage where they are to be transferred to archives for permanent preservation?

Some may argue that given this situation, machine readable records never truly become inactive. This is not the case. What we have to remember is that there are really two separate components of machine readable records: the record and the information contained within it. With paper, the record's form, medium and information are all bound together in a single package which cannot be changed without destroying it. The information may be used to *create* a new record, but the original record remains intact and inactive, once it has lost whatever value it had to its creator. With computer records, the information and the record are usually considered as separate entities. The record becomes inactive when it is removed from the computer system along with the indexes which fix the relationships among the parts of the datafile. Even when this is done, however, the information is considered to be active and available for use at any time simply by putting it back on the system. Indeed, the computer encourages this re-use and re-manipulation.

What the archivist must appraise, then, are those records which are removed from the system in concert with their indexes and have no further use to their creators *in that form.* These are the only truly inactive machine readable records. However, to ensure that what reaches this inactive stage are records worthy of archival preservation, it is often necessary to attempt to appraise the *information* contained in the active records before they are removed from the system. I shall go into this point in more depth later in this chapter and in Chapter Three, which deals with scheduling electronic data.

We are faced with another conflict if we follow Schellenberg's views that the need for appraisal has increased dramatically due to the expansion in public records since World War II, and that archives are only those few records carefully selected to make the best use of limited space. Lionel Bell, in discussing the question of the appraisal of large databases in his article "The Archival Implications of Machine-Readable Records," says that computerization has increased the amount of interesting and valuable information available to archivists. Computers have also made possible the storage of vast quantities of information in smaller and smaller spaces.[23]

The problem is this. Because computerized information takes up very little space, an archives may find it easy to accept an entire datafile when only a very small percentage of the contents has true archival value. This situation leads to a relaxation of appraisal standards. In many large computer systems, however, the constant replacing or updating of new information for old means losing information unless regular backups are made. But complete systems backups are often difficult or even impossible in larger systems, so, to ensure the information is retained, archivists have to decide early in the records' life cycle what elements or information will be important and therefore need to be selected for preservation. This necessarily creates a need not for a relaxation of appraisal standards, but instead for the refinement and enhancement of those criteria.

The problem can be resolved relatively easily by reinforcing stringent appraisal criteria to ensure that anything preserved is worthy of preservation. If the information is not archivally valuable, it should not be kept no matter how easy it is to do so because of its medium. For every tape of questionable value that is

---

23 Lionel Bell, "The Archival Implications of Machine-Readable Records," Archivum 26 (1979): 87.

preserved, valuable information may be lost due to reduced institutional resources. In addition, there should be a greater emphasis on the appraisal of machine readable datafiles much earlier in their life cycle to ensure that information is not accidentally lost.

If we cannot save everything that is on a computer simply because it is easy to do so, can we instead use sampling techniques on machine readable records? Sampling is somewhat of a controversial subject in archives, partly because each archivist seems to have his own conception of what sampling really is, and partly because many techniques are wide open to the charges of bias. In his study The use of sampling techniques in the retention of records: A RAMP study with guidelines[24], 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.[25] Of these four methods, both specimen and purposive sampling are considered debatable techniques in terms of their obvious bias. They attempt to preserve the most significant or important parts of a record series, but as a result they create a skewed and subjective view of the whole. While it would be possible to do the same kind of sampling on machine readable 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. Systematic sampling is somewhat better. It depends upon the establishment of a particular pattern of selection in that every nth file is preserved, or 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.[26]

---

[24] Felix Hull, The use of sampling techniques in the retention of records: A RAMP study with guidelines. (Paris: UNESCO, 1981).

[25] Ibid., p. 10.

Hull's main argument against the use of sampling techniques on what he calls "non-conventional records", such as machine readable records, is based upon their appraisal process, which I outlined earlier. Because datafiles are accepted into an archival institution after a very careful selection process based on their physical properties, to ensure that only accessible information in terms of supporting documentation and of readability is retained, Hull says that "a later, secondary selection process involving sampling is purposeless."[27] He and many others appear to hold to the idea that sampling machine readable records is a waste of time and effort precisely because it is easy to store and retrieve computerized information, so there is no need to reduce bulk. Even if it would be a simple matter in theory to carry out a true random sample of a particularly large file of uniform records, we must not ignore the practical technical difficulties which could arise. In the case of an ideal datafile, all we would need to do is to generate a set of random numbers and pick the records identified by those numbers until the desired size of the sample is reached. There is no need to spend a great deal of time and energy in preparation as there would be in numbering conventional files before selection; each computer record already has an internal, implicit, unique identifier used to reference the record.

This would be the ideal situation. In reality, each file must first go through a technical analysis to determine which records are readable and therefore accessible. Unless this is done, sampling would be a useless endeavor, since a large percentage of unreadable information could be selected; the unavoidable selection process which precedes the taking of a random sample would falsify that sample, since it would not be representative of the original file.

---

[26] Hull, Use of sampling techniques, p. 13-15.

[27] Ibid., p. 35.

There is a more serious drawback to sampling electronic records that relates directly to the structure of a computer file. Unlike a conventional paper file, which is treated as a single unified entity, a computer file is composed of two separate but related parts: the datafile containing the records, and the index(es) which defines the relationships between the records in the file. Sampling paper records requires two steps: the random number generation determines which files are to be selected from a series, then the file folders are removed from their original boxes and recombined in the same order in new boxes, creating a smaller version of the series. Achieving similar results with computer files requires an additional step. Since the index actually determines the proper arrangement of the file, sampling begins with a random sample of the index, not the datafile. Once the index sample reveals which specific records in the datafile have been selected, those records are copied or removed from the datafile. The result of this process should be a smaller version of the original index and that portion of the datafile which the new index arranges. At this point it seems that the only difference between sampling paper and electronic records is that in practice, computer records require an additional step, that of selecting records indicated during the index sampling. The sampling of computer records can, however, become more complex in cases where a datafile has more than one index.

Indexes can be created by any user to impose an order on the records of a datafile according to his or her specific needs. The indexes may only deal with a portion or a subset of the whole file. When sampling a datafile with multiple indexes, the procedure described above for a single index file would simply be repeated for each of the file's indexes. The more indexes a file has, the more likely it is that records will be chosen more than once when the index samples are complete. Once the selected records are re-combined to form the new datafile, this duplication can be

eliminated easily by ensuring that only one copy of any record is kept in the new datafile. A datafile with a number of indexes should differ from a single index file only in its size and not so much in its structure.

It should be noted that the sampling of machine readable datafiles may not result in appreciable storage space reductions. It is possible that in the case of multi-index files the number of records indicated by the index samples could require the preservation of the whole datafile or a large portion thereof. In addition, the space saving on single index files could be only a matter of magnetic storage space. If the datafile and its index are stored on a single magnetic tape, the sampled file and index would still take up the same amount of physical space as before (i.e. the space of a single magnetic tape).

It appears that there is no fundamental difference between sampling paper and computer records. Electronic records only require practical modifications of sampling procedures in order to deal with the new medium. Hull's conclusion that the sampling of machine readable records is not a useful process is valid, yet his arguments undermine his position.

Hull states that for purely practical reasons, machine readable records must be dealt with on a tape-by-tape basis. Since each tape occupies a set amount of physical space, sampling would not result in a space saving and is therefore an unnecessary operation. In addition to this point, computer records undergo a strict appraisal process before they are ever accepted into an archival repository; during that process, the archivist looks not only at the readability of the material but also at its legal, administrative, and informational values. Sampling accessioned files is therefore not justifiable for two reasons. One has already been mentioned: if only readable records are accepted, a later random sample would be unrepresentative of the

original file. The second reason is complementary; the appraisal process ensures that the archives only accepts those records deemed worthy of permanent preservation. The applicability of sampling techniques to such material is, in Hull's words, "not proven."[28]

The implications of Hull's arguments against sampling machine readable records are disturbing. He says that the rigorous appraisal of computer records ensures that there is no need for further sampling, and that such a process may be worthless anyway because it does not usually result in a physical space saving. At the same time, he fails to point out that paper records, theoretically, also undergo a stringent appraisal process before being accessioned. In spite of this, paper records are routinely sampled to reduce their bulk, even though the methods used are considered questionable by many archivists. Since the appraisal process applied to machine readable records (excluding the specific practical technical concerns of technical analysis) is essentially that which is applied to paper, it is not valid to argue that machine readable records cannot really be sampled while paper can because the former undergoes "strict appraisal". Hull's remarks reinforce the popular misconception that machine readable records and paper records, because they reside on different media, must therefore be treated differently. This is not the case.

The main argument against sampling computer records does not refer to the specific medium, but to the efficacy of the sampling process in general. The main argument against sampling paper files is that a sample may not reflect the true nature of the original file or series; this argument may be applied equally to computer files. The medium of records under consideration for sampling should not affect the decision to sample; it should only affect the practical application of sampling procedures once the decision to sample has been made.

---

28 Hull, Use of sampling techniques, p. 37.

I think that what is necessary for the proper preservation of machine readable records is not better sampling techniques but 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 I propose 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 machine readable records, only EDP (*Electronic Data Processing*) 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 provisions for early appraisal of machine readable records, however, less of this loss might occur.

The second stage of the appraisal should encompass the process outlined in this chapter, 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 stage one, 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 it to store 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 machine readable, and in view of the fact that records can conceivably lose their value, datafiles should be re-appraised 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 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 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 de-accessioning, and that reckless destruction is not practiced, but it still should be done.

Of the three stages, I believe that the first is the most vital because without an identification of the importance of information soon after its creation, machine readable records may never even reach the second stage of a full archival appraisal. Already countless valuable machine readable 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.

The first stage is also the most difficult to achieve because it, like the development of suitable terminology for machine readable records in archival science, requires archivists to cooperate with specialists outside their discipline. Archivists understand appraisal principles, but they must be able to convince the computer scientists and administrators responsible for machine readable records of

the need for and value of early appraisal. And the way to do this is to enlist the assistance of a discipline related to archival science which is already making great progress in the controlling of information: records management.

Chapter Three: Records Management, Archives and Machine Readable Records

While the history of archival science stretches back over hundreds of years, the history of records management as a discipline begins in the middle and later twentieth century. It is generally accepted that records management was initiated in the United States in the early 1940s as a response to the enormous records production brought on by World War II. Since that time, both records managers and archivists have been aware that their areas of expertise are connected. While records managers are mainly concerned with the efficient retrieval and use of an organization's current records, they also realize that the inactive records they wish to dispose of often have lasting values to others outside the creating organization. Archivists, in their turn, see that the effective management of current records affects the quality of the records which will become archives.

Yet, despite their mutual respect and an awareness that records management and archival science are not only conceptually linked but complementary, archivists and records managers have long believed that they actually practice separate disciplines. Archivists, notes Jay Atherton in Archivaria in 1985[1], have traditionally viewed themselves as scholars, the partners of historians and researchers; for that reason, they generally avoided any involvement with business administration or management which might detract from this status. As a direct result, many archivists felt (and still feel) that records management was simply one of the means to ensure the preservation of materials with enduring values, an instrument of better appraisal.

---

[1] Jay Atherton, "From Life Cycle to Continuum: Some thoughts on the Records Management - Archives Relationship," Archivaria 21 (Winter 1985-86): 43-51.

Records managers took the opposite view. Seeing themselves not as scholars but as managers and administrators whose duty was to serve the needs of business, records managers considered the archival function to be the final stage of the records management process. They believed that by concentrating on the efficient management of current records in an organization, and by developing systematic and reasonable disposal procedures, records management naturally results in archival materials: to them this was not the ultimate goal of the process, but merely a pleasing side benefit.[2]

Given that there may be elements of truth in both of these positions, the idea that records management and archival science are separate disciplines is gradually being replaced by a new perspective. Changes in record keeping practices since World War II have lead many to suggest that both disciplines are actually elements of a broader, more general discipline known as "information management." As a result, emphasis in records keeping is slowly shifting from a concentration on the management of static records to the control and manipulation of a more fluid entity, information. Both archivists and records managers are beginning to realize that in actual fact, their functions are not just complementary but very closely entwined. Cooperation, which was once touted by both sides as a professional courtesy, is now seen as essential. The shift towards a greater integration of archives and records management is clearly visible in the shift from the "life cycle" model of record keeping to the concept of the "continuum".

The life cycle model of record keeping assumes that every record moves through a set sequence of regulated steps from the moment of its creation to that of its final disposition. Until very recently the life cycle was the accepted model of both the National Archives of the United States and the National Archives of Canada. It consists of eight distinct stages of which the first four are grouped into a records

2 Atherton, "Life Cycle to Continuum," p. 43-4.

management phase. They are: creation or receipt, classification, maintenance and use, and disposition (destruction or transfer). The second set of four stages, selection, acquisition, description, and preservation and reference, constitute the archives phase.[3]

Looking at the life cycle model, it is easy to see how records managers and archivists could consider themselves as practicing separate even if related professions. The elements of the records management phase focus almost exclusively on current records. Records managers are more concerned with the efficient retrieval and use of information in the records than with their eventual archival values; once the records have been disposed of, the records management phase is considered to be over and the function of the records manager completed. Archivists, accepting this sharp break between the two phases, then take control and concentrate on caring for and making available the end residue of permanently valuable records. There may be some consultation or cooperation between the two groups of professionals, but in most situations, the distinctness of their functions prevails.

Recently, however, more and more archivists and records managers are switching from the life cycle to the continuum model of record keeping. A continuum is a unified approach to records keeping which fuses records management and archival functions. Atherton gives a good overview of a continuum in his article. Instead of the life cycle's rigid eight stage, two phase structure, the continuum consists of four broad stages or functions in record keeping: creation or receipt of records, classification, scheduling of information, and maintenance and use of records. It is irrelevant whether the records are maintained in a creator's office, in an inactive records centre, or in an archives. Instead of

---

[3] Atherton, "Life Cycle to Continuum," p. 44.

focusing on the record as an object to be cared for, the continuum model encourages archivists and records managers to concentrate on providing service to anyone wishing to use the records for whatever reason.[4] Records therefore do not require special care at various stages of their life cycle as if they were living entities. The records never change; only their users and their purposes do.

The importance of the continuum model to both archives and records management is revealed in two observations. First and foremost, there is no rigid separation or break between records management and archival functions either stated or implied within the continuum model as there is within the life cycle concept. Because there is no clearly marked separation of the two professions, archivists are no longer limited by theory and tradition to the role of passive recipient of a records management activity's residue. They can become more involved in various aspects of the management of current records such as classification and scheduling on the grounds that those records will later be entrusted to their care. Conversely, records managers may interest themselves in what has been regarded as archivists' tasks, if they understand that their service is directed to the general public, that they are responsible to all citizens, not just to their organization.

Archivists and records managers can combine their talents, experiences, and training to good effect. A records manager has considerable experience in the everyday use of records within his organization. He knows what records are created in carrying out what functions, how long records have a value to the agency and when they can be disposed of. He is therefore competent to develop policies, procedures, classification systems and other tools for current records keeping so that when the time for scheduling approaches, he has firm intellectual control over the

4 Atherton, "Life Cycle to Continuum," p. 48-9.

material. This allows him to apply retention periods to the records according to the organization's needs. Archivists, on the other hand, have a much greater knowledge of the long term values of records because of their prolonged contact with scholars and other users. Their appraisal techniques take into consideration legal and administrative concerns as well as research needs. The exercise of archival appraisal during records scheduling, therefore, not only ensures that records are not destroyed simply because they have ceased to be of value to the organization; it may also give records managers an insight into how they could modify current record keeping practices, taking into consideration the general purposes of records. It seems, therefore, that only through the combination of their functions can archivists and records managers be sure of preserving permanently valuable records.

The shift from the life cycle to the continuum model, and the growing cooperation between records managers and archivists is not limited to Canada. In 1981, English archivist Derek Charman wondered if archival and records management functions were actually separate. After discussing the various problems faced by both in the United Kingdom, he concluded that that they were really part of the same discipline.[5]

While the conclusion is not really surprising, the arguments are quite interesting. Throughout the article, Charman discusses the increasing complexity of modern records keeping which necessitates greater cooperation between records management and archives; he argues that archivists need to become more actively involved at an earlier stage to ensure the preservation of valuable documents. What is significant is that even though he speaks only about traditional paper records, he could easily be speaking about computer records instead.

---

5 Derek Charman, "Archives and Records Management - an interface?" Journal of the Society of Archivists 6,7 (April 1981): 423-27.

Even if the life cycle model of records keeping may still work for traditional paper records, archivists and records managers alike are beginning to look more favourably upon the continuum concept as a better means of controlling computer records. The two professions are being literally forced into greater cooperation by the fact that computer records cannot reasonably be handled using the life cycle model.

There are several reasons why machine readable records do not fit easily into the life cycle model, the first having to do with the difficulty of determining whether a record is active, semi -active or inactive. Paper records in the life cycle model are considered active from creation through 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 "re-activated", but this is a rare occurrence.

Computer records, on the contrary, cannot be classified quite so neatly. They can be considered active if they reside on a computer system which is used daily. They can be called inactive if they and their indexes have been removed from the system and frozen in a certain form. But when can we classify computer records as semi-active? Let us consider as an example the case of a large database, consisting of several dozen files and thousands of logical records. Of those many files and records, there are a couple of files and several dozen specific logical records which have not been consulted for months or even years. If similar records in paper have their status altered from active to semi-active and inactive based on a reduction in the numbers of times they are consulted, how do we classify these electronic

records? Are they active simply because they still reside on an active system? Are they semi-active or inactive because they are rarely or no longer used in conducting everyday business? Obviously, this is a difficult distinction to make. Yet the life cycle requires that this distinction be made because the treatment of records is clearly divided into separate phases: active and semi-active phases for records managers and an inactive one for archivists. It seems that with machine readable records, the designation semi-active as it is currently used is not valid.

The question of the status of backup and update copies also causes problems. If proper backups of the files are taken and removed from the system along with their indexes, the records become inactive as long as they are never put back onto the system and re-activated. But what happens if records are copied to backup tapes for safety purposes, but also remain in their original form on the system, unconsulted and unaltered, along with clearly active data? Would we then have active and inactive copies of the same records at the same time? Conversely, we must also look at the opposite situation, where a system's logical records are updated constantly so that only the most recent form of the information is preserved on the system. If no effort is made to capture previous versions of the records, can there ever be anything but active records?

These observations are not designed to show that it is completely impossible to fit computer records into the life cycle model, but merely to indicate that there are many more considerations to make with machine readable records than there are with paper records. One that we should always bear in mind is that with machine readable records, it is quite possible to deal with two separate elements, the information and the record, something which is not possible with paper. The information contained in backup copies can be reactivated if it is put back on the computer in the event of a system failure. Yet the records themselves can remain

inactive at the same time because the relationships between them, as indicated by the indexes, were frozen during the "archiving" process.

Another purely practical reason why machine readable records do not fit easily into the life cycle model is that they reside on a volatile medium which requires early attention to ensure permanent preservation. While records managers and archivists could, without serious loss of information, exercise control over paper records sequentially during their life cycle, the same cannot be said for computer records. In addition to the normal threats to paper records from fire and flood, computer systems are subject to catastrophic losses of data through electrical malfunctions, magnetic interference, or hardware or software failure. They are also extremely vulnerable to human error through incompetence or lack of awareness. For example, a system which maintains only the most up-to-date version of the records and its information could be losing vast quantities of valuable data unless someone created policies and procedures to ensure the preservation of older versions of the records. Similarly, a computer technician unaware of the long-term value of the backup files could erase them when re-cycling the tapes on which the files were located in order to save the organization the cost of buying new tapes.

Because it is so easy to lose valuable information in the early stages of computer records' existence, which is traditionally the domain of the records manager, archivists can no longer accept the role of passive recipient. They cannot wait until inactive machine readable records are offered to them for appraisal as they might have for paper records; too many computer records have vanished by then. The sheer volatility of machine readable records should be a powerful inducement for archivists to accept the increased involvement in the scheduling process which is offered to them by the continuum model.

A third reason why machine readable records do not fit as well into the life cycle concept as they do into the continuum has to do with those responsible for their creation and maintenance. Unlike paper, computer records require a reasonable amount of specialized technical expertise to maintain, expertise which is found outside the area of traditional records management. Within most organizations, "EDP systems people" are not usually within the jurisdiction of the records manager, despite the fact that the latter theoretically is responsible for all of the records of an organization. Since the life cycle concept refers only to the two functions of records management and archives, there appears to be no place within it for the specific needs and concerns of the systems people. Usually, these people have little or no idea of the principles of records management and archival science; in addition, their own practices and terminology sometimes clash with those of the other two disciplines. If the continuum concept instead of the life cycle one is used for computer records, EDP systems people will be encouraged to take part in the management of information and records as members of an integrated discipline encompassing both records management and archives.

A final problem which is related to the isolation of EDP systems people is that these people often have their own conception of the term "life cycle" which, like many other computer science terms, is at odds with the records management and archival terms. There are actually two interpretations of the meaning of "life cycle" with regard to computer records: that of the systems people and that of the users. To most systems people, the term refers to the life cycle of software, a series of five distinct stages which mark the development of a software system. The stages are: specification, design, implementation, testing and operation/maintenance.[6] At any time in the development process the software may be modified to remove flaws or add improvements; the result is a new version of the original program. The

---

6 I. Sommerville, Software Engineering (Reading, Massachusetts: Addison-Wesley Publishing Company, 1982), p. 3.

term version should be familiar to purchasers of microcomputer software packages, such as WordPerfect Version 5.0 or DOS 3.2.

A user's interpretation of the life cycle is slightly different. Generally speaking, users refer to the backups or updates made from a computer system when they speak about life cycle. They are not concerned with the version of the software which produced the information, nor do they care if the records are designated as active, semi-active or inactive by the records manager. Their only concern is the *age* of the information, which is defined as its "generation". If, for example, a system is updated every two weeks, the first and therefore oldest update is called the "first generation", and is followed by the "second generation", and so on. Once a certain number of generations is reached, the earliest ones are usually destroyed as being too out of date.

Thus, in dealing with machine readable records we are faced with not one but two different interpretations of the term life cycle which conflict with the records manager's viewpoint. Depending upon the influence of the individuals involved, or the organizational structure of the agency, any one of the three views may be held as the "correct" one. If the agency's system people are predominant and concerned only with emphasizing their interests, the term might refer only to the software life cycle. In this case, there would be a direct conflict with the records management term, which is concerned with the records' frequency of use, not the program which records and manipulates it. If the users' viewpoint prevails over all others, it could determine destruction of "out dated" backup copies. In any event, if we continue to use the records management life cycle concept for machine readable records without taking into consideration the other interpretations available, we would be faced with more terminological problems such as those outlined in Chapter One.

The response to the problem of the treatment of machine readable records within records management should be evident by now. Archivists and records managers alike must accept the underlying principles of the continuum model, dictating that there should be no clearly marked division between their two disciplines, and that both groups of professionals should combine their skills, experience and efforts throughout the entire process of records keeping. Scheduling of electronic information is the most suitable area with which to begin, for it is here that materials of permanent value are identified and controlled to ensure their preservation. Once this area is examined and developed, archivists and records managers, as well as other professionals such as EDP systems people, can branch out into other areas according to the continuum model.

A great deal of work on the scheduling of machine readable records has been carried out at the National Archives of Canada. Under the guidance of John McDonald, PAC[7] developed a theoretical approach and various procedures for scheduling computer records that have been accepted and adapted by several records management programs across the country, particularly in the provinces of Alberta and British Columbia. As was the case with the PAC's appraisal guidelines, the PAC's procedures for scheduling EDP records have had such an impact that it is opportune to look at them in detail.

The procedure developed at PAC is commonly referred to as the "system overview" approach and embodies several important concepts for the scheduling of computer records. These concepts are meant to guide the actual scheduling process. The most important concept behind the system overview approach is that the scheduling of an organization's traditional paper records and of its machine readable records must be treated in a unified and integrated fashion. The physical medium

[7] The work on scheduling electronic records was carried out at the Public Archives before it became the National Archives in 1987, hence the reference to PAC.

of the records is irrelevant to that specific purpose, and neither paper nor computer records can be examined and scheduled in isolation without adversely affecting the treatment given to the rest of the organization's records.

In practical terms, this concept of unity and integration of paper and electronic record translates into the "team approach" to scheduling electronic records. The National Archives insists that not only records managers but archivists, EDP systems people, users and specialists on relevant legislation such as the Access to Information and Privacy Acts (ATIP) must be involved in the scheduling process from the very beginning. The pooling of knowledge and experience allows for specialized concentration on key areas of the scheduling process without harming the overall results. No matter what their background, the ultimate goal of all those involved is the preservation of the information in the computer system.

A second important concept is that the system overview approach is designed to give a generic picture of the information contained within a computer system, as well as the major stages through which that information passes as it is created and processed. The appraisal is based upon the premise that the best way to handle machine readable records is not to concentrate on individual records, files, and tapes, because this tactic would leave the records manager, the archivist, and their colleagues overwhelmed with specific details. Those details would prevent the team from seeing the overall pattern and function of the records that they are attempting to schedule.

A system overview approach, therefore, requires the use of techniques designed to work from the top of a system downwards. This means that the overview begins by describing very basic information, such as the name of the main computer system and its overall purpose and functions. It then becomes more detailed as it describes

the system's component parts such as the number, names and purposes of the subsystems, or the contents of various levels of processing files. Inherent in this top down approach is the idea that for maximum effectiveness, a computer system only needs to be described down to the functional level which will permit full intellectual control over the information contained within it.

Another facet of a system overview approach is the emphasis placed on appraisal. Appraisal is regarded as the cornerstone of scheduling electronic records because it ensures that only the most valuable and important information is preserved, as is the case with paper records. Given the volatility and fragility of magnetic media, it is not surprising to discover that the National Archives encourages the appraisal of electronic information as early as possible, perhaps even at the system design stage, before the specific information is even created. Treated as the first step in an appraisal process, such as that proposed in Chapter Two, appraisal at the design stage would ensure that electronic information is not inadvertently lost through mismanagement.

Finally, the system overview approach requires a strong policy framework supported by senior management personnel. Because the scheduling process requires the cooperation of many individuals outside the traditional jurisdiction of records managers (such as EDP systems people and users), there must be some means of ensuring that the cooperation takes places even if those involved are personally unwilling. In the federal government, for example, the Treasury Board developed specific sections of its Administrative Policy Manual to deal with computer systems and their records. In Chapter 440 (7), departments are invited to develop suitable records schedules, to submit requests for records disposition to the Dominion (now National) Archivist, and to store and handle EDP material in a proper fashion.[8] Since Treasury Board has the power to approve or reject a

department's funding, compliance with the policy should be widespread.

Unfortunately, Treasury Board's Chapter 440 (7) has been found to be inadequate to fill the needs for which it was designed because of its hazy definitions and lack of serious attention to the part that systems people play in scheduling. It was to be replaced by Chapter 461, a chapter dealing with computer records and complementing the existing Chapter 460, which deals with traditional paper records, providing policy guidance in terms of their organization, retrieval, retention, protection, disposal and archival management. As of August 1985, however, John McDonald reports that, since Chapter 461 continues to exist only in draft form, the Public Archives lacks authority to extend the influence it has over the management of paper records to the management of computer records.[9] The specific steps of the scheduling process devised by PAC are seen in Figure 4. Figure 5 lists the various elements present in most system overviews, while a modified example of a system overview is given in Appendix B following Chapter 5.

The National Archives' procedures have also proven to be quite adaptable to other situations, particularly that of a provincial government. Over the past two years, the British Columbia Government and its Records Management Branch have been working to adapt these procedures and implement the strong policy framework that the federal government appears to lack. To date, their efforts have proven quite successful. As in the federal government, British Columbia's policy framework resides in specific directives from the Treasury Board. Unlike Chapter

---

8 Treasury Board of Canada, Administrative Policy Manual: Chapter 440: Electronic Data Processing. Section 7 (Ottawa: December, 1978), p. 4.

9 John McDonald, Scheduling Data in Automated Systems: A Final Report (Ottawa: Public Archives of Canada, 1985), p. 19.

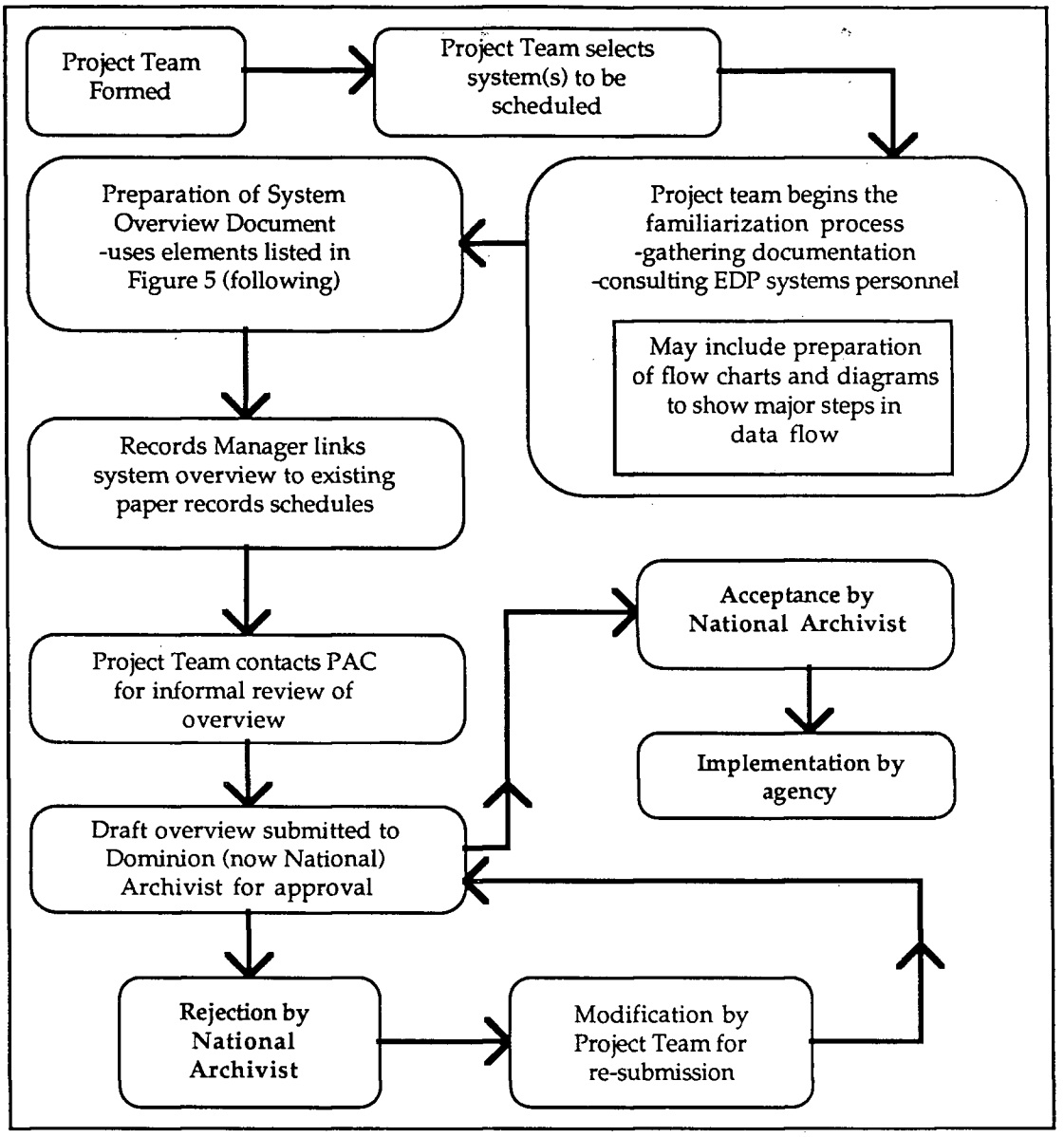

Figure 4: Summary of Steps involved in Public Archives of Canada's "System Overview"

461, however, the Information Systems Management Framework (ISMF)[10] has been approved. The ISMF states that every computer system in the B.C government, the funding of which is authorized by Treasury Board, must be

10 Province of British Columbia Treasury Board, Information Systems Management Framework (Victoria: Queen's Printer, 1986).

1. the name of the computer application
2. the purpose of the application
3. the agency responsible
4. the informational content of the records
5. a summary of the inputs, outputs and processes
6. a summary of retention and disposal periods for each file
7. the hardware and software environment
8. a list of major technical milestones in the application's development
9. a list of technical personnel to contact
10. a summary of available documentation and other manuals
11. cross references to related paper record series
12. cross references to Classes of Records and Personal Information Banks

Figure 5: Elements of the Public Archives of Canada's System Overview

described in two separate inventories: the general Information Systems Master Inventory (ISMI) and a specific Information Systems Descriptive Inventory (ISDI).[11] The ISMI provides the application name, responsible branch, application description, hardware environment and operating costs of every system in the government that is funded with a sum exceeding twenty five thousand dollars. It also identifies the type of system (accounting, inventory, word processing etc.), and specifies whether the system contains administrative or operational records. This information is gleaned from Information Systems Plan (ISP) documents filed yearly with Treasury Board by each ministry requesting initial or continuing funding.[12] The ISMI is currently maintained jointly by the Records Management Branch and

11 Information Systems Management Framework - Standards Manual, p. 3-4.

12 Ibid., p. 2.

the Treasury Board. In theory, the directives offered in the ISMF provide for some semblance of intellectual control over the information within a computer system before it is even fully designed; the functional statements within the inventory will alert the Records Management Branch to the existence of potentially valuable records so that they can be appraised early. The Information System Descriptive Inventory (ISDI) is in fact the Public Archives' system overview modified slightly to fit the context of a provincial government and the specifics of the British Columbia situation. The first ISDI completed was that of the Land Title Office's Automated Land Title Office System (ALTOS) in the summer of 1987. It is this "overview" which appears in Appendix B. The ALTOS application proved to be an excellent model for subsequent ISDIs. Because the application had no processing files, the records analyst preparing it could concentrate her efforts on refining other elements, such as the retention and disposal summary, or the classification links between electronic and paper records. Another advantage for the development of the project, which underlined the integrated approach to records management taken by both the National Archives and the Records Management Branch, was that the Land Title Office's paper records had already been controlled in a draft schedule, Land Title Office: Operational Records Classification System; thus, the ALTOS ISDI (overview) became a part of that schedule. The Records Management Branch has clearly shown that the Public Archives' procedures are sound in practical application as well as in theory. The real strength in the British Columbia program, however, is the effectiveness of its policy framework, which is lacking at the national level. Acceptance of the ISMF, with its two offspring, the ISMI and ISDI have shown that it is indeed possible to gain early control over electronic information. It is to be hoped that in the future this model will be emulated by other records management programs.

Although the successful application of PAC's system overview procedures in

British Columbia has shown the theory to be sound and applicable, there are still some problems in the scheduling of electronic data which threaten our ability as archivists and records managers to preserve a complete record of our society. First and foremost are the practical difficulties of bringing the theory behind scheduling to those unfamiliar with the basic concepts of records management and archival science. It is always hard to alter ingrained attitudes, but doubly so if the subjects under discussion come from another discipline, and have difficulties determined by terminology.

However, the obstacles are not only constituted by the people not conversant with the principles governing records management and archives; the scheduling process itself can provide practical problems. In cases where a system is large and complex, with dozens or hundreds of master files, processing files, and output files, scheduling could go on for weeks or months as the various project team members struggle to familiarize themselves with the system and its contents. It is possible in situations such as this (and they appear to be the majority of cases) that valuable information is lost before the schedule is fully implemented, unless steps are taken to prevent it.

The most serious problem we face, however, is not related to the present application of the system overview theory of EDP scheduling but to its application in the near future. At the present time it is 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, a system overview is relatively straightforward to carry out.

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 combine their resources to create and maintain a large, anonymous system or database which can serve all of their diverse but related needs at once. Electronic information becomes so fluid that not only is it difficult to determine the active, semi-active and inactive stages of records, 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 the system overview approach becomes complex and difficult. Where do we begin to schedule the contents of these shared databases? Can we legitimately break them down into smaller units fit for individual schedules or overviews, or will this destroy the true nature of the system? Or will such a system require a completely different scheduling technique than that of the system overview?

Until we master the most basic procedures for scheduling electronic data which is linked to a single, dedicated application, we will not be able to develop the theory and techniques necessary to deal with the more complex data management computer systems of the future. Cooperation between records managers, archivists and other related professionals is, therefore, not only encouraged but demanded, for without it preservation of electronic records will suffer greatly. With computer records and paper alike, archives and records management must be treated as part of the same discipline, that of information management in general. When archivists, records managers and their colleagues succeed in gaining early control over electronic information, archivists will have to devote themselves to the analysis of the problems involved in arrangement and description of machine readable records.

Chapter Four: Arrangement and Description

While archivists have written a great deal on the arrangement and description of paper records, there is a distinct paucity of literature on the arrangement and description of their modern counterparts, machine readable records. It is true that the use of the computer as a major tool in record keeping is fairly recent. Nevertheless, it was interesting for me to discover, in a limited survey of ten English language archival science journals from Europe, New Zealand, Australia, Canada and the United States from 1948 to 1986,[1] that only one article dealt with the arrangement of machine readable records. Furthermore, that one article did not even address the issue of the application of the principles of the archival theory of arrangement to computer records; rather, it discussed exclusively the United States National Archives attempts to create inventories of the physical medium, the magnetic tapes on which those records were stored, preferring a practical approach to a theoretical one.[2]

Why has there been such a gap in the archival literature on machine readable 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 machine readable 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 machine readable records. The real reason is much more profound.

---

1 Catherine A. Bailey, "Computers and Archives: A Survey of the Literature," paper prepared for Archival Studies 500 course, University of British Columbia, 3 April 1987.

2 Everett O. Alldredge, "Inventorying Magnetic-Media Records," American Archivist 35 (July 1972): 337-46.

The principles of archival arrangement are not as relevant to machine readable 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 machine readable records, suffers from arrangement's 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.

Before we can discuss the application of the theory of archival arrangement to computer records, we must examine some of the key principles involved. For this, we have to refer to one of the classic works of archival science, the Dutch Manual for the Arrangement and Description of Archives by S. Muller, J.A. Feith, and R. Fruin.[3] Originally published in 1898, and designed to deal with paper records in the context of the Dutch archival system, the Manual discusses principles and offers practical suggestions which remain sound and may be applied equally well to many of the situations encountered by modern archivists.

The cardinal principle of archival arrangement espoused in the Manual is the principle of provenance, or *respect des fonds*. In the 18th century, archival collections were often placed into rigid subject classification systems, with no regard for the actual origin of the records. Subject classification, however, proved to be destructive of the organic nature of many archival collections;[4] consequently, archivists in the middle of the 19th century began to reinforce the idea that all records created by a single administrative entity had to be kept together and treated as a single unit or fonds. The Manual clearly states that under no circumstances

---

[3] S. Muller, J.A. Feith and R. Fruin, Manual for the Arrangement and Description of Archives, trans. Arthur H. Leavitt (New York: H. W. Wilson, 1968).

[4] Muller, Feith and Fruin take careful note that "collection" does not mean a group of things arbitrarily brought together by a collector, but "an organic whole of records received or produced by an administrative body or one of its officials" (p. 13).

must there be a mixing of collections[5], or a splitting of a single collection between archival institutions; it also states that, wherever possible, it is necessary to re-construct collections which previously have been split up.[6]

Muller, Feith and Fruin devote the first part of the Manual to the origin, composition and qualities of archival collections. Every archival collection is described as a living organism which is created through the activities of an administrative body, and reflects the function(s) of its creator and any change in them.[7] Archival collections are, therefore, highly individualistic and unique; as a result, they cannot be arranged according to pre-determined classification schemes, but must be treated by the archivist according to the original rules which governed their organization. The guiding principle of archival arrangement therefore, according to Muller, Feith and Fruin, is that the archivist must examine the collection very carefully and understand the functions and organization of its creator and arrange the records accordingly. This principle, known as the respect for original order, is based on the assumption that the order given to the documentary body by its creator was not arbitrary or casual, but the natural result and reflection of the organizational structure and functions of the producing agency.[8] If the archivist were to alter this original order, the result would be the destruction of the collection as a single, complete entity, or at the very least, the loss of valuable information about the records inherent in their arrangement.

There are two corollaries to this principle. The Manual states that only when the original order has been restored as far as possible through examination of the records and research into the history of the creator can the archivist deal with particular situations, such as the placement of undated documents.[9] Secondly, if

---

5 Muller, Feith and Fruin, Manual for Arrangement and Description, p. 33.

6 Ibid., p. 38.

7 Ibid., p. 19.

8 Ibid., p. 57.

the archivist does decide to do any re-arrangement because in the original order files or documents were misplaced, he must take special care to ensure that each deviation from the original order is noted in the description of the collection,[10] and that no new mistakes arise from his rearrangement, for there may be a perfectly good reason for and possibly consequences of the original "error" which he corrected.[11]

Muller, Feith and Fruin emphasize the need to understand fully not only the records themselves, but also the functions and motives of their creators, factors both linked to the records' creation.

While these principles function extremely well for paper records, they are somewhat more difficult to apply to the archival arrangement of machine readable records. The principle of provenance has become so ingrained in archival science that its application to computer records is 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 will be treated as a single fonds. It is much more difficult, however, to apply the principle of respect of original order and all of its various corollaries to machine readable records because of the nature of the medium.

Arrangement is of little concern for machine readable 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 we knew which index had been in place when the

---

9 Muller, Feith and Fruin, Manual for Arrangement and Description, p. 59-60.

10 Ibid., p. 64.

11 Ibid., p. 71.

datafile was first made operational, we 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 itself gives fundamental information about the records and the functions and structure of the administrative body which created them. With machine readable records, information about the records derived from their arrangement comes from examining the indexes. As long as those indexes are removed from the computer system, verified as accurate, and remain frozen and unaltered, the information we seek from the file's arrangement will remain intact.

The relative unimportance of the arrangement of the records in the computer file therefore leads archivists to *process* rather than arrange datafiles once they are accessioned by an archival institution. Processing is a much more technical operation than arrangement, focusing mainly upon aspects of the physical medium rather than on the records and the information they contain. During processing, datafiles are first copied to a blank tape to provide a master datafile and a security backup copy. The archivist then takes a sample of the records from the file and begins verification, not of the information itself but of the integrity of the records in terms of *codes, frequencies, tracks, density,* and *parity.* The file and the record sample are then checked against the file documentation, which must be provided with the file when it is acquired. If any discrepancies occur between the file and its documentation, they are noted, but not changed or "cleaned up". Finally, once the documentation package or manual is prepared for public use through careful editing

and the addition of notes containing information gleaned by the archivist during processing, the datafile is ready for users.[12] Throughout the processing phase, the archivist assumes a passive role in the sense that he seeks only to verify the accuracy and integrity of the records in front of him, not to do extensive research into what *might* have been their original arrangement.

Arrangement of computer records, then, has come to mean something slightly different than arrangement of paper records. Arrangement means two things to an archivist dealing with paper records: it is the intellectual order of records in series and series in fonds, and the process of imposing an order on a body of records, be it the original order determined through research and records examination by the archivist, a variation of that original order, or a new order devised by the archivist in the clear absence of any evidence of an original order. This is not the case with machine readable records. Since the order of the records in a datafile can be changed quickly and easily at any time through the creation of a new index, an archivist dealing with computer records would think of arrangement neither as the intellectual order of the records, nor as the process of reconstructing or devising such an order. He would think of arrangement as the actual physical positioning of the logical records on the magnetic tape before they are placed into any intellectual order by an index. Arrangement of machine readable records therefore indicates if the logical records are *blocked, coded,* or *hierarchical.* Any of these states will naturally affect the usefulness of the datafile for secondary use by determining how the logical records can be searched and retrieved; it will also affect the processing of the file, since different techniques must be used for each kind of recording. Hence, arrangement is not considered a process succeeding accessioning but one of many appraisal criteria considered within the detailed appraisal process applied to machine readable records.

---

[12] The Machine Readable Archives: An Overview of its Operations and Procedures, (Ottawa: Ministry of Supply and Services Canada, 1980), p.11.

If arrangement as a status and as a process has little significance for the archivist dealing with machine readable 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.[13] 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.[14] Finally, archivists assume that specific finding aids are usually more appropriate to certain levels of arrangement than to others.[15] 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.[16]

Bearing these assumptions in mind, we can then imagine the ideal archival descriptive system. Each archival fonds, regardless of the physical form of the

---

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

14 Ibid., p. 7.

15 Ibid., p. 8.

16 Ibid., p. 8.

records, is carefully arranged according to the established principles of archival arrangement such as are found in the Manual, or processed in the case of machine readable records. Then the fonds is described in a series of different finding aids, proceeding from the general to the specific level. At the most general level, the fonds is described in a finding aid which lists the entire holdings of the institution and gives a title and an overview of its contents. Next, description focuses on the fonds itself, providing the full scope of the contents, a history of the body which produced the records, and a brief summary of the series. The inventory could even be so detailed as to include both series and file lists, thereby allowing users access to the most basic element of the fonds, the individual item. In addition to these more basic finding aids, the archivist often produces indexes to one or several related fonds to provide easier access to individual items according to their subject, name, date or relevant geographical region. In the ideal situation, the system of finding aids is carefully planned out in advance, with proper cross references to ensure that a user can enter the system at any point and still find what he wants quickly and easily. There is consistency, both in the structure of the descriptive tools themselves and in the terms used to describe the records. Above all, no archival fonds is ever put into the institution's stacks or made available to the public until it has been described at a very basic level; a main entry card showing the fonds' title and physical extent is usually sufficient.

Unfortunately, the ideal situation is often just that: an ideal. Most systems of archival description are not planned before their initiation but simply grow and change to meet the immediate and current needs of the institution. Often there is little connection between finding aids; when there are connections, they are not consistent. Authority files, standard terms for headings, and policies and procedures governing the preparation of finding aids are also rare; what standard subject headings exist are usually adopted from the field of library science. The result is that

archival descriptive and indexing practices are highly idiosyncratic, varying widely between institutions and even individual archivists.

Despite this rather gloomy picture, however, it must be said that archival description suffers mainly from a lack of concentrated organization and of standardization. The Bureau of Canadian Archivists' study, conducted in 1985, found that most Canadian archivists capture and record the same kinds of information about all of their fonds and collections regardless of the type of record, the physical medium, or even the type of finding aid being produced.[17] It is therefore more a question of how best to arrange the information than one of changing the kind of information collected.

Let us look more closely at archival description as it relates to machine readable records. 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 machine readable records. Improper or sloppy description of paper records does not necessarily preclude a user from finding the information he 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 manpower and funding forces many institutions to forgo extensive description. That kind of exercise is much more

17 Bureau of Canadian Archivists, Towards Descriptive Standards, p. 53.

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 he sees 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 can reach the same stage. This leads us to believe that there is a minimum level of description necessary for machine readable 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 he is quite familiar with the computer system which runs the file, he may need only the name of the computer to proceed. If, however, he is not familiar either with computers in general or the system in particular, he may require a set of documentation, ranging from a simple codebook to a series of user manuals outlining all the computer's commands, before he can begin extracting any information from the datafile.

With machine readable 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.

In this endeavor, archivists dealing with machine readable records are assisted greatly by the research already done and the results reached in the field of library science. In the past ten to fifteen years, librarians, faced with growing numbers of machine readable datafiles (MRDF) within library holdings, have been forced to develop cataloguing rules for material whose bibliographic information is not necessarily neatly printed on a title page as it is with paper monographs and serials. Work by a subcommittee of the American Library Association's Cataloging Code Revision Committee resulted in the inclusion of Chapter 9 –Machine-Readable Data Files– in the second edition of Anglo-American Cataloguing Rules (AACR2). That Chapter now constitutes the standard for bibliographic description of MRDFs in data libraries and is also used by several archives, including the National Archives of Canada, to catalogue their datafiles. These standard library practices for the description of machine readable datafiles are by no means limited to brief bibliographic descriptions intended for card catalogues. Sue A. Dodd's seminal work Cataloging Machine-Readable Data Files - An Interpretive Manual[18] not only assists readers in the interpretation of the AACR2 guidelines, but also outlines a multi-level system of descriptive record keeping which looks remarkably similar to a system of archival description. While the proposed system is not in complete harmony with the principles of archival science, it should not be rejected outright on the grounds that it was developed by librarians instead of by archivists; there is no need to "reinvent the wheel."

According to Dodd, there are two aspects to this descriptive record keeping system: the organization of the data elements used in bibliographic control, and the various finding aids in which those elements are used. The data elements for

18 Sue A. Dodd, Cataloging Machine-Readable Data Files - An Interpretive Manual (Chicago: American Library Association, 1982).

description can be broken down into six groups depending on whether they need to identify, to describe, to classify, to access, to analyse, or to archive or maintain a machine readable datafile.[19] The elements themselves can be seen in Figure 6. Dodd then goes on to explain where to find this bibliographic information in the absence of a title page on the datafile or its documentation, and how to prepare the results according to standard bibliographic conventions.[20] Her specific instructions throughout stress brevity, clarity and completeness.

Having outlined the elements necessary for full bibliographic control of machine readable datafiles, Dodd turns to an examination of the system of finding aids in which those elements are included. Again, as with archival description, the system is divided into organizational levels proceeding from the general to the specific; in this case, there are four. First is the catalogue record, which may involve not only the catalogue entry itself, but also worksheets for in-house control generated during the processing phase.[21] Next there is the data abstract, a brief form of documentation designed and used to lead a user from the even briefer catalogue entry to the third level of description, that of the datafile specific documentation. Data abstracts in printed or electronic form are quite often used to disseminate information about a data library's holdings to other institutions and users. Documentation, as opposed to a data abstract, allows a user to examine the actual contents of a datafile once he has determined that the file meets his needs. It is not unusual to find these documentation packages standing alone in place of their datafiles in other data libraries. Finally, the fourth level of record keeping consists of processing characteristics, an in-house record of the technical information needed to run the datafile.[22] As Dodd points out, there may be some overlapping of data

---

19 Dodd, Cataloging Machine-Readable Data Files, p. 157.

20 Ibid., p. 157-9.

21 Ibid., p. 180-9.

22 Ibid., p. 197.

Level 1: Bibliographic Identity

- Corporate or personal author
- Title, subtitle, and other title information
- General material designation
- Edition, plus appropriate statements of responsibility relating to edition
- Production statement, including place, organization, and date of production
- Distributor statement (if appropriate), including place, organization, and date of distribution
- Size of file
- Series statement (title and numbering within series, if appropriate)
- Notes

Level 2: Data Abstract

- Unique identification number
- Type of file (numeric, text, computer program, etc.)
- Bibliographic citation of MRDF
- Methodology (universe sampling, unit of analysis, etc.)
- Geographic coverage
- Time period
- Date(s) of data collection (if unique from other dates)
- Summary (subject matter description)
- Derived source of data (if from printed source or other MRDF)
- File size
- Complete citation or reference to accompanying documentation
- Related publications based on study or use of MRDF
- Terms of availability
- Contact Person

Level 3: Classification

- Subject classification (applied locally)
- Index terms, descriptors, or key words (applied locally)
- Geographic headings (applied locally)
- Library of Congress Classification Number
- Dewey Decimal Classification Code
- Library of Congress Geographic Classification

Level 4: Access of MRDF (technical information)

- Mode of access
- Type of data carrier or storage medium
- Memory or storage requirements
- Recording density, blocking factors, etc.
- Computer compatibility
- Software compatibility
- Peripheral requirements
- Special formats or system files

Level 5: Analysis or Use of MRDF

- File structure/ sort sequence
- Condition of Data
- Restrictions on use
- Intended audience or level of expertise
- Applications of the file or program
- Linkage with other files or programs
- Unit of analysis
- Sampling procedures
- Citation and location of documentation

Level 6: Archiving or Maintaining MRDF

- Archival study number
- Personal or organizational donor of MRDF
- Date received
- Date processed and entered into collection
- Retention status (if temporary)
- Access code (publicly available, restricted etc.)
- Cost for file duplication/ dissemination
- Frequency of updates or additions
- Holdings note (for serials or serial-like MRDF)
- Processing history (changes, revisions, modifications, etc.)
- Documentation number or shelf location

Figure 6: Elements of Library Science Description of Machine Readable Data Files adapted from Sue A. Dodd, *Cataloging Machine Readable Data Files- An Interpretive Manual* (1982), p. 158-9.

elements from one level of record to another depending upon the institution; however, in most cases the bibliographic data elements previously mentioned in Figure 6 remain as they are grouped.

What should be immediately apparent to all archivists studying this system of descriptive record keeping is that not only do librarians capture the same kind of information about machine readable datafiles that archivists do, but also that the method of organizing the resultant documents in levels from general to specific to guide the user to his goal is very similar to the ideal situation of archival description. This is not to suggest that the system can be adopted wholesale by the archival community. For one thing, librarians generally concentrate on the item (tape) in hand and do not place the datafile within the context of the rest of the records produced by the same administrative body. This is because libraries tend to receive single datafiles containing survey and other research data rather than series of operational or administrative records created by an agency. This concentration on the technical details of the individual file would detract from more archival concerns such as the functions of the records and their place within the context of the organization of the administrative body. What this observation does suggest is that if archivists can copy the library science example, add the necessary archival elements relating to the creators of machine readable records, and adapt any other necessary data elements, the only real work left in establishing descriptive standards for computer records will be developing policies and procedures, and dissemination.

It must be kept in mind, however, that the library science model of multi-level record keeping is not pure description but a system of reference tools designed primarily to allow users to look at the datafile from a series of different perspectives. It happens that those reference tools carry quite a bit of information about the file,

but they are all directed solely to getting the user to the datafile, whereupon he may decide finally if the information it contains suits his needs. Even the package of documentation is designed to help the user *use* the datafile, not necessarily to understand the composition of the information or the circumstances surrounding its creation.

What a system of archival description for machine readable 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 machine readable records and their related paper records, which often include the input and output documents of the computer system. An archival system of description needs to treat the documentation manual, the equivalent of an inventory for a paper records fonds, as the focus of attention instead of concentrating on the card catalogue entry as librarians do.

The system of description for machine readable records at the National Archives of Canada is a good example of the necessary blending of library science and archival science methods. According to an overview of its functions published in 1980, the Machine Readable Archives initially relied upon the same multi-level documentation system used in the library field: card catalogue entries, data abstracts, documentation packages, and records of the file's logical/physical characteristics.[23] By 1983, however, the situation had changed. While archivists still adhered to the AACR2 method of preparing catalogue entries, they also developed a system of reference tools more in line with traditional archival practice. Four main elements - an Automated Inventory of Canadian Machine Readable Datafiles, an accession register, catalogue entries, and published inventories - now contain enough

[23] Machine Readable Archives: An Overview, p. 12.

description of the datafile to allow a user to learn that a file exists, where it is located, and to decide whether the information it contains suits his needs. For more detailed description, the user is directed by the system to consult the basic finding aid, the documentation manual. The manual encompasses many of the data elements formerly held by librarians in separate records such as the data abstract and the list of physical characteristics. It also provides administrative information, printouts of the records and file structures[24], and may even give a brief administrative history of the creating agency.

Herein lies the difference between a library and an archival system of description for machine readable records. While a library system concentrates on the file itself, the archival system steps back and attempts to describe the datafile as a part of the entire group of records created by an agency. For this reason, the archival system employs specific tools such as the National Archives of Canada's Record Group Index to provide links to other related archival materials such as papers, maps, and photographs.[25] Within the documentation manual, there are cross-references to other datafiles and related publications.[26] Without these elements, there would be a serious threat to the archival status of machine readable records; treated according to easily applicable library techniques, as valuable as they may be, computer records would retain the stigma of being so physically different from paper records that they could not be treated in the same theoretical manner as other archives.

There are several things that archivists can do to improve current descriptive practices for machine readable records. First, they can accept the basic elements of the library system, focusing their energy on adapting various elements to reflect a

---

24 Machine Readable Archives Bulletin 1 (3): 2.

25 Ibid., p. 1.

26 "Description Form: Abstract for Machine Readable Data File." Photocopied handout distributed at the Workshop on Machine Readable Records, School of Library, Archival and Information Studies, Vancouver, B.C., 2-3 March 1987, p.4.

proper archival viewpoint. Second, 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 machine readable 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.

There is one matter specifically related to machine readable records that must be stressed. It was stated previously that there is a minimum level of description for machine readable datafiles which varies according to the user's level of expertise. Since it would be nearly impossible to anticipate that level of expertise with every user, I believe that archivists should make every effort to ensure that, before any datafile is entered into a catalogue or inventory of holdings, the documentation manual be as complete as it can be made. The preparation of a preliminary catalogue record before datafile processing is complete, as is done at the National Archives of Canada, could cause problems for users. Knowing of the datafile's existence and confident of his ability to manipulate it, a user could conceivably make use of the file but miss valuable related information normally noted in the documentation manual. As well, experience has shown that in times of reduced manpower and resources, many archivists cannot afford the time to go back and complete the "preliminary description". Archivists should avoid these dangers and take great care to ensure that their techniques of description will aid, not hinder, them in providing reference services for machine readable records.

Chapter Five: Reference and Other Issues

One of the prime functions of the archivist is to make his institution's holdings available to a wide variety of people who wish to use them: employees of the archives' sponsoring agency, academics, genealogists, government officials and members of the general public. It is a task which requires a clear understanding of the principles and practices governing reference service, and an awareness of those broad issues in modern society that have an impact on archives, namely the questions of copyright, freedom of information and individual privacy. The development of computer technology has also created two new areas of concern for archivists: the flow of computerized information across national boundaries, commonly referred to as transborder data flow; and the admissibility of computer generated information and printouts in a court of law. We shall examine each of these issues in turn in order to determine how they will affect the archivist's ability to make machine readable records available to users.

A: Reference Services and Access

The principles guiding the activity of reference service are quite straightforward within the context of a modern democratic society. First and foremost, archivists have a responsibility to make materials preserved in their institutions available to all researchers on equal terms of access. Archives may have a primary clientele on whose needs they focus their programs and publications; a university archives, for example, may devote a great deal of time to assisting academic staff and students, while a corporate archivist's first duty is to the employees of his company. The principle of equal access to archives, however, means that archivists must not give selective preferential treatment to some researchers on the basis of their affiliation,

interests, or abilities while excluding or ignoring others who do not meet arbitrarily established requirements.[1]

A corollary to the principle of equal access is that researchers should be allowed full access to all of the records preserved in an institution which bear on their area of study, with the exception of those proscribed by the law. Unfortunately, this principle is sometimes difficult to respect to its fullest extent. It is unethical for an archivist to withhold deliberately information he knows is useful to a researcher's project if he is aware that the records containing it are in no way restricted; however, the archivist may fail to bring all the material which is relevant to the researcher's topic to his attention because of an inadequate finding aid system, or the inability of the researcher to express his topic clearly, or his desire to be secretive. In this last case the archivist is unconsciously withholding sources, but he may just as well be accused of unethical behaviour. Such a thing did occur in 1968, when a researcher accused the Franklin D. Roosevelt Presidential Library of unfair and unjust treatment through the "'deliberate and systematic' withholding of certain items that they knew he needed for his work."[2] Although the Library was eventually partially exonerated by a joint committee of the American Historical Association and the Organization of American Historians, it was concluded that the archivists "should have made sure" that the researcher saw what he needed, even if it was located not in the Presidential Library but the Library of Congress.[3] This ruling puts archivists in a difficult position. They devote a great deal of time and energy to making all their materials available to users to the best of their abilities, yet, because of the possibility of human error, they may be charged with obstruction. While this

---

1 Sue E. Holbert, Archives and Manuscripts: Reference (Chicago: Society of American Archivists, 1977), p. 2-3.

2 Ibid., p. 4.

3 Ibid., p. 4. For a full report of the event, see *Final report of the Joint AHA-OHA Ad Hoc Committee to Investigate the Charges Against the Franklin D. Roosevelt Library and Related Matters* (Washington: AHA and OAH, 1970).

is not the place for a full discussion of the implications of incidents like that mentioned above on the work of archivists, it may be appropriate to point out that researchers should be partially if not wholly responsible for finding all the materials they need.

Reference can be provided *about* an institution's holdings as well as *from* them. Dissemination of information about its holdings is an archives' primary function and encompasses virtually all levels of archival description. Archives often include brief descriptions of their fonds in national guides such as the Union List of Manuscripts in Canada and the National Union Catalog of Manuscript Collections in the United States so that researchers who live far away from the repository will know about its materials and be able to decide whether a research visit is worthwhile. Some institutions also duplicate or even publish finding aids for their more heavily used fonds and distribute them to researchers and other archives. Public awareness or education programs are also frequently a part of this primary aspect of reference.

The core of this aspect of reference is constituted by the descriptive finding aids made available to users in the archives' reading room and by the assistance in the use of those finding aids provided by the reference archivist. As was suggested in Chapter Four, with a well-planned, carefully integrated system of finding aids, a researcher can learn a great deal about the archives' holdings without having to look at the records themselves. If the researchers have difficulty interpreting the descriptions within the finding aids, they can consult the reference archivist, who will have learned something of the nature of the researcher's topic through the course of an entrance interview and should be able to help with suggestions about records relevant to his topic, as well as assist him in the proper use of the finding

aids. The same entrance interview is also an excellent opportunity to inform the researcher of the institution's policies on access restrictions, security, and reproductions. In their turn, researchers can point out to the archivist errors they find in the finding aid system, the correction of which may make the system much more efficient and easier to use, or make suggestions for improvement. Feedback from users is an integral part of the reference process.

The second aspect of reference, the provision of information from an institution's holdings, is a contentious issue. Archivists are not supposed to engage in extensive research for users, in effect doing their work for them; neither can they refuse in principle to answer enquiries for information within the records. Some telephone or mail enquiries of this kind may contain only simple and straightforward questions from people who cannot come to the archives to do the research themselves. Some persons may require only a certified copy of a birth, marriage or death certificate, or a copy of a homestead record, or even just the briefest information from either record. Others, in contrast, may request information which, by its nature or the amount, involves hours of research on the part of the archivist. In order to ensure that reference requests are not handled in a discriminatory or arbitrary fashion on the basis of criteria such as the archivist's personal interests (which would contravene the principle of equal access), each archival institution should implement written policies regarding the amount of attention in terms of archivist's time that each reference request shall receive. Simple requests, such as those from church or homestead records, can be carried out within reasonable time. If, however, it becomes apparent that a user is abusing the system by sending an extensive series of simple questions one at a time, thereby forcing more research from the archivist than is considered acceptable, some action must be taken. In such cases, as with situations involving extensive or complex

reference requests, a polite form letter could be sent to the user, stating the archives' policy of not undertaking to do extensive research for patrons and enclosing a list of local research assistants available for hiring, should the patron be unable to come and do his own work. Insofar as it is possible, a user should always be encouraged to visit the archives rather than to request the archivist to do research for him.

Reference service provided about and from machine readable records is similar in theory to that provided about and from paper records, but it differs in practice. The principles of equal and full access apply as well, and so does the use of an integrated system of finding aids. However, the need of a personal contact between the researcher and the reference archivist is too often not felt in dealing with machine readable records; this makes the entrance interview rare and affects the use by the researcher of other types of archival sources. The computer technology which supports the magnetic medium is no longer confined to large institutions and has spread out to private individuals; this fact, coupled with the facility of the magnetic medium, allows large amounts of information to be copied and disseminated fast with little effort and with great accuracy. Although researchers can still choose to travel to the archives and look at the datafiles there, they are no longer forced to do so. After looking at the finding aid sent out by the institution, such as a national machine readable datafile database, a user can chose the file that he feels fills his requirements and, by simply writing or phoning the institution, can have a copy of that file and its documentation sent to him. He does not need to see the reference archivist once during the transaction.

This kind of reference procedure requires the archives to provide a special set of services. Michael E. Carroll, referring in particular to the Public Archives of Canada and to a few other institutions preserving machine readable records, says that there are three services which can be provided to a researcher either in the archives or at a

distance, depending upon the computer facilities, software and staff or expertise available to each party: tape copying, data extraction, and data analysis.[4] Depending upon other factors, such as records restrictions or the need to protect personal information contained in the records, researchers may either carry out at their own place or have carried out at the archives any of the three services. Carroll urges that in the absence of restrictions, whenever a user can provide the required facilities, he should be encouraged to accept a copy of the file and extract himself the information he needs because archivists should not do a user's work for him.[5] Tape copying and limited data extractions are therefore the two services most often performed in machine readable archives.

One practical difference between the provision of reference service about and from paper records and the provision of reference service about and from machine readable records is that the first is free while the latter imposes a charge on the researcher in order to offset the cost of materials and computer time. The Public Archives of Canada, for example, uses a series of formulas to determine the total cost for tape copying and data extraction based on the number of files and reels of tape, and on the amount of time required by those operations (Figure 7). By charging for its services, the PAC hoped to make files as accessible as possible to users, while at the same time reducing "indiscriminate and superfluous requests".[6]

---

4 M. E. Carroll, "A Perspective of Machine-Readable Archives and Public Service," Automatic Data Processing In Archives 2,1 (1976): 7.

5 Ibid., p. 9.

6 Ibid., p. 10.

Tape Copying Formula: 20x + 20 y = total cost

where x= the number of magnetic tapes involved
y= the number of files involved

Example: three files on one reel of tape would be
20(1) + 20 x 3 = 20$ + 60$ = 80$

Data Extraction Formula: 20x + 20y + (z-7) = total cost

where x= the number of magnetic tapes
y= the number of files
z = number of person hours greater than 7

Figure 7: Formulas Used to Calculate Cost for Machine Readable Records Services at the Public Archives of Canada[7]

If there are differences between reference service for paper records and for machine readable records due to the medium, there are also significant similarities mainly related to the archivist's ability to make records available to researchers, that is, the issues of restrictions, copyright, freedom of information and privacy.

Restrictions on the consultation and/or use of records may be imposed in a variety of ways and for different reasons. Records may be sealed by statute for certain periods of time in order to protect the security of the government, ongoing activities of administration, or the privacy of citizens, as is the case with census records. Donors may want to include in the contract of gift a clause limiting the access to their records in order to protect their personal reputation, career or business secrets. Likewise, private people who deposit their papers in a public institution may make it a condition of deposit that researchers ask them for

7 Machine Readable Archives Bulletin 1 (4) (Winter 1984): 2.

permission to use the materials and that only accredited users be allowed access. A milder form of restriction is the limiting of quotation from or publication of the records as a condition of access. Restrictions may be applied equally to paper and machine readable records. Whatever the medium, archivists must consider carefully the full impact of the restrictions, especially those imposed by donors, to ensure that they do not greatly affect the usefulness of the records in the long term and hamper the archivist's function of communicating his institution's holdings to researchers. Unreasonable and excessive restrictions are to be avoided whenever possible, no matter what the medium.

However, the computer has made it possible to administer restricted machine readable materials with much more ease than paper. If a fonds of paper records contains some restricted files, it is quite likely that the entire fonds, or at least entire series, would be closed to prevent the possibility of accidental disclosure. The manipulability of the magnetic medium, instead, allows archivists to make all the unrestricted material in the fonds available to researchers. The computer needs only to copy the whole file without including specially tagged records, creating as a consequence a researcher-usable copy.

Another important and vexing issue facing archivists, whether they handle paper or machine readable records, is the question of copyright. Copyright, the exclusive legal right to reproduce, publish and sell the matter and form of a literary, musical or artistic work, is guaranteed in Canada by the Copyright Act of 1924.[8] Whether a work is published or unpublished, the Copyright Act protects the ideas' unique expression, but not the ideas themselves. In an article in Archivaria,[9]

---

8 See Canada. Laws, statutes, etc., *Revised Statutes of Canada, 1952*, Chap. 55, s. 4, ss.1. Bill C-60, an act to revise the Copyright Act of 1924, is still before the House of Commons.

9 Gina La Force, "Archives and Copyright in Canada: An Outsider's View," Archivaria 11 (Winter 1980-81): 37-51.

Gina La Force gives a good summary of the provisions of the Copyright Act as they affect archivists. Canadian copyright legislation, because of its date, contains no provisions for works created on media developed in the last sixty four years, such as sound recordings, photographs other than wet chemistry, and computer datafiles. Photocopying and microfilming, techniques widely used by archival institutions for communicating and protecting their holdings, are regulated through the extension to them, by analogy, of those sections of the Copyright Act which refer to hand copying.[10] Apart from these obvious limitations, the legislation of 1924 contains many points which are still relevant to the treatment of archival records regardless of their medium, and are of serious concern to all archivists.

According to the Copyright Act, the duration of the copyright for published material is fifty years after the date of publication; therefore, unpublished material, which constitutes the major part of archival holdings, is protected by copyright in perpetuity, or until it is published. This means that, although researchers may have access to all non-restricted material in an archives, they cannot legally quote or copy any part of it without permission of the copyright holder or his heirs. Clearly, this is a serious limitation on the use of archives. Thus, the Keyes-Brunet Report of 1977[11] proposed that a special provision for unpublished materials deposited in an archives become part of any new copyright legislation. They argued that copyright of unpublished materials should exist

> until publication ... and for 50 years thereafter, but ... the total term of protection [shall] not exceed 75 years after the death of the author, or 100 years after his death where the work has been deposited in an archives.[12]

---

10 La Force, "Archives and Copyright," p. 38.

11 A.A. Keyes and C. Brunet, Copyright in Canada: Proposals for a Revision of the Law (n.p: Consumer and Corporate Affairs Canada, 1977).

12 Keyes and Brunet, Copyright in Canada, quoted in La Force, "Archives and Copyright,"p. 40.

This recommendation is somewhat flawed, insofar as it places an extra twenty-five year restriction on unpublished works simply because they have been deposited in an archives. What is important, however, is not the actual figure given for the duration of copyright for unpublished materials, which may be amended, but the underlying principle that there must be a finite term of protection on such material when it is preserved in an archival institution.

The question of the ownership of copyright further complicates an already complex issue. Archival fonds contain different kinds of documents: letters, diaries, reports, photographs, memoranda, etc. Each document is protected by a copyright invested in the creator of that work and his heirs. Thus, the copyright of a fonds containing hundreds of letters would not be entirely held by the creator of the fonds (private or public), but would be split among all of the authors of the individual documents. Moreover, a donor of archival material might not even be the person responsible for the creation of the fonds, and would therefore hold no copyright privileges himself which he could transfer to the archives. In such a case, the search for copyright holders, which is necessary before researchers can use the materials properly, is often impossible.[13]

Another area of copyright legislation which affects archives and archivists is commonly referred to as "fair dealing". Section 17(2)(a) of the Act states that fair dealing "with any work for the purpose of private study, research, criticism, review, or newspaper summary" is not an infringement of copyright. Archives, which have traditionally relied upon this section to photocopy and microfilm their holdings, are actually in a tenuous legal position, being responsible not only for their use of the reproductions they make, but also for the use that researchers make of materials copied by archives on their behalf.

---

13 La Force, "Archives and Copyright," p. 44-5.

All of the problems outlined above affect the treatment of all archival materials regardless of their medium, but they acquire a broader dimension for the archivist handling machine readable records. Software, for example, as a single work containing the ideas expressed by one or more authors, is considered to be protected under the provisions of the Act; how can an archives process and make available to researchers a series of machine readable records if they are dependent upon a software package which is subject to copyright for their interpretation? 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 a growing problem resulting from the continuous development of computer technology. Because copyright provisions only protect the *form* of the work, not the ideas contained within it, how can we 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 we even begin to apply the legislation to an entity which is so fluid that it can be re-arranged into countless new forms at any time, each with its own copyright?

The resolution of these issues is of vital importance to all archivists. The problems posed by the characteristics of machine readable records, coupled with the difficulties of interpretation of the 1924 Act as to unpublished materials make it imperative that new legislation be implemented to clarify the position of archives and assist them in the performance of their duties, free from most threats of legal

sanction. Archivists are presently making great efforts through their national associations to ensure that their concerns are met in any new statute.

The Copyright Act is not the only legislative text which currently affects the archivist's ability to make records available to researchers. In 1983, the Canadian government passed two new laws which had a significant impact on archives: the Access to Information Act and the Privacy Act. The first law places the archivist in the position of arbitrator between records creators and researchers, the latter between citizens and citizens.

In democratic societies, government records must be properly preserved so that in time the people may have access to them.[14] A democratic government is accountable to the citizens, who are the ultimate owners of its records; therefore, if a certain degree of confidentiality or secrecy is necessary in such areas as foreign affairs and defense, in general, the government is not to conceal its actions by prohibiting access to their records for an inordinate length of time.

The Access to Information Act states that all citizens are granted the right of access to all government records, according to the principles outlined above, with the exception of those classes of records listed in Sections 13 thorough 24. These specific exemptions, shown in Figure 8, are designed to protect information, the disclosure of which could affect the public good or private interests,[15] and are clear, few in number, and "consistent with and confined to the genuine need for confidentiality in the governing process."[16] In order for records to be "protected", they must be submitted to an injury test, to see if the release of the information they

14 Robert J. Hayward, "Federal Access and Privacy Legislation and the Public Archives of Canada," Archivaria 18 (Summer 1984): 48.

15 Ibid., p. 50.

16 Honourable John Roberts, Legislation on Public Access to Government Documents (Ottawa: Minister of Supply and Services, 1977), p. 9.

contain would prove to be harmful either to the public or to individual interests. The records may also be submitted to a class test, where the exemption is invoked on the grounds that they belong to a class outlined in Sections 13 to 24. It should be noted that material already preserved in the Public Archives, the National Library, or the National Museum are not considered within the bounds of the Act.[17]

	Exemption Class	Section	Nature of Exemption
(1)	Information obtained in confidence from another government	13	Mandatory
(2)	Information injurious to federal-provincial relations	14	Discretionary
(3)	Information related to international affairs and defence	15	Discretionary
(4)	Information related to law enforcement and investigation	16	Discretionary
(5)	Information that is threatening to the safety of individuals	17	Discretionary
(6)	Information that is prejudicial to economic interests of Canada	18	Discretionary
(7)	Personal Information as defined in the Privacy Act	19	Mandatory
(8)	Third-party information (trade secrets or commercially sensitive information)	20	Discretionary
(9)	Advice or recommendations	21	Discretionary
(10)	Information relating to testing or auditing procedures	22	Discretionary
(11)	Information subject to solicitor-client privilege	23	Discretionary
(12)	Information subject to statutory prohibition against disclosure	24	Mandatory

Figure 8: Exemptions to Canada's Access to Information Act (1983)[18]

Any person wishing to obtain copies of records under the Access to Information

17 Hayward, "Federal Access and Privacy Legislation," p. 51.

18 C. Ian Kyer, "The Federal Access to Information Act: a statute to be aware of," Canadian Computer Law 2,9 (July 1985): 170.

Act must apply in writing to the access coordinator of the responsible agency, providing as much information as possible to aid in the search for the records. If his request for access is denied, the petitioner may appeal to the office of the Information Commissioner, who will carry out an investigation.[19] Throughout the whole process, the onus is placed upon the government agency wishing to deny access, rather than on the person seeking the information.

But how does the Access to Information Act affect archivists if the information already in the Public Archives is not considered to be within the scope of the Act?

The Access to Information Act has an impact on archives by making records management a vital part of government record-keeping. The Act requires all government agencies to list the classes of records they hold in the yearly Access Register; this is an excellent finding aid not only for researchers but also for archivists faced with requests for government records not contained in their holdings. In addition, the possibility that they may be called upon to produce any document at any time and within a time limit set by the law, forces government departments to improve their records classification systems and scheduling procedures. Within this context, the growing acceptance of the continuum model of record keeping means that archivists can get much more involved in the early stages of the scheduling process; freedom of information concerns naturally become part of the appraisal criteria for both paper and machine readable records. All of these factors ensure a smooth flow of records from the government to the archives and eventually make the task of arranging (processing) and describing materials much easier.

Under this legislation, it also becomes much clearer who is responsible for access

19 Kyer, "The Federal Access to Information Act," p. 169-70.

to records. In the past, the access to government records that were held in the Public Archives and less than three decades old, was still controlled by the transferring agency, not by the Archives. Access is now governed by the holder of the records; physical transfer to the Archives is accompanied by transfer of responsibility for access.[20] Such a responsibility acquires a broad social dimension when it refers to the communication of private information contained in government records. This issue directly concerns machine readable records. According to the joint Department of Communications/ Department of Justice Task Force set up in 1972 to study the effects of computers on privacy, there are three main categories of privacy: territorial, personal, and informational.[21] Archivists are most concerned with the third category, informational privacy, hereinafter referred to simply as privacy. On the basis of the principle that information about a person is fundamentally owned by the person, privacy means that each individual has the right not to have information taken from him forcibly or without his knowledge and an interest in knowing who has access to the information, as well as how and where it is used.[22] Proper protection of privacy, therefore, requires that the person in question gives his consent to the gathering of personal information, that the information be held in a secure state with no indiscriminate access, and that the person be notified if the information is to be divulged to a third party.

Governments are naturally the largest potential misusers of private information because of the vast quantities they collect and manage in order to provide various services. In order to curb violations of privacy, both in the government and in the private sector, various pieces of legislation have been created. In 1978, Canadians were, under the provisions of Part IV of the Canadian Human Rights Act, granted

---

20 Hayward, "Federal Access Legislation," p. 53.

21 Privacy and Computers (Ottawa: Information Canada, 1972), p. 13-4.

22 Ibid., p. 14.

the right to limit access to personal information about them held by the federal government, as well as the right to correct that information and control its dissemination.[23] The Privacy Act, enacted in 1983, mirrors the Access to Information Act in its structure, setting out rights, exemptions, review process and scope of the protection of privacy.

Within the Privacy Act there are four exemptions which are relevant to archivists' work. Personal information may be disclosed to the Public Archives for archival purposes (appraisal); to private persons for research or statistical purposes where the agency head is satisfied that the identifiable data is necessary for the project's completion, to researchers settling native land claims, and by the Public Archives on a discretionary basis to researchers needing access to files containing such information.[24]

So, the responsibility for the protection of privacy is either with the records creating agency or with the archivists, depending upon where the records containing personal information are preserved. It is not possible to realize all the implications of such a responsibility if we do not consider that the largest part of personal information is computerized. Computers are indeed a powerful tool, capable of collecting, processing, storing, manipulating and linking millions of pieces of personal information from a variety of sources. It is possible to build a profile of an individual, to find out where he lives, his marital status, where he works, how much he earns, what his driving record is, even what he likes to read from the local public library. Such information in the wrong hands could be used against that person in countless different ways. But while all of this is possible, it must also be pointed out that the computer is not an ungovernable monster but a

23 Hayward, "Federal Access Legislation," p. 51.
24 Ibid., p. 52.

tool which, if used carefully and properly with the right safeguards, has distinct advantages over paper records.

The computer room can be secured so that only a few people have direct access to it. Passwords and data checks ensure that only authorized personnel can look at or alter records. Limiting printers linked to the system to one or two controlled installations reduces unauthorized printouts, and the lack of access outside the agency via modem helps prevent tampering.

As tools for a dissemination of personal information which does not violate privacy, computers are exceedingly well suited. They allow unrestricted portions of restricted files to be released to researchers or produce records without fields containing personal identifiers. Compilations and summaries are equally easy to prepare and are often acceptable to most researchers requiring personal information for social science research; they do not particularly care *who* the person really is, only what his habits and actions are.

Concluding the issues of access to information and privacy it is possible to say that the problems they present to archivists are not directly related to the medium on which the information is stored; therefore, the philosophical, ethical principles governing the communication of paper records are valid as well for machine readable records. This is also reflected in the large body of archival literature on the subjects of reference, copyright, freedom of information and privacy. Specialized works do exist for machine readable records, but they tend to deal with the technical aspects of the medium. Once again, it appears that the differences between machine readable and paper records lie in the practical application of a common theory.

B: Transborder Data Flow

Transborder data flow, "the movement across national boundaries of data and information for processing and storage in computer systems,"[25] is not a new phenomenon but one which has not been fully analyzed and evaluated by archivists, being directly linked to the most recent technological developments. Consequently, archivists cannot rely on the guidance of traditional archival theory in dealing with the problems created by a phenomenon which has an enormous influence on the relations between independent nations, making national boundaries and even national laws ineffective.

The information which crosses national boundaries usually consists of technical, economic, scientific and personal data. It can be transmitted either as raw information (text) through the mail, or as machine readable data (tapes, discs) by post, or as electronic impulses transmitted through telecommunications equipment.[26] The institutions which rely most heavily on transborder data flow include multinational corporations sending data from regional to head offices, international unions, service bureaux and credit bureaux. Banks are also relying more and more upon electronic transfer of funds through banking networks, such as those shown by various symbols on many instant banking machines in Canada and the United States. Transborder data flow has an effect on four distinct aspects of a nation's existence: national economy, national sovereignty, individual privacy,

25 "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.

26 J. Freese, "Social and Cultural Implications of Transborder Data Flows," 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. 542.

and cultural identity or sovereignty. 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 its degree of internal control over such things as industrial sabotage, power failures, civil unrest, strikes, unemployment etc. More damaging than this, 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. Where 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 machine readable 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 government, or the subject of the information, are powerless to stop him. Therefore, transborder data flow can only be controlled through the development of specific national and international legislation ensuring the security of information.

Archivists are equally concerned with an aspect of transborder data flow which has not been discussed in much depth in the literature: the issue of cultural sovereignty. 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.[27] 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 mold 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 sovereignty 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 machine readable 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 he 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.

---

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

Transborder data flow is an issue which should concern archivists a great deal, as it affects their ability to make records available to researchers promptly, securely, and fully. Harold Naugler touches on the subject of transborder data flow briefly in his Archival Appraisal of Machine Readable Records, yet he treats it not as a separate issue but as an appraisal criterion and urges the creation of more national and international legislation.[28] Archival theory needs to be developed in this area, because transborder data flow presents theoretical problems related not only to the communication and appraisal of records, but also to their arrangement and description, that is, to their intellectual control.

C: Computerized Information and Printouts as Evidence in Court

The laws governing the suitability of various forms of information as evidence in a civil or criminal court case have been in place in most Western nations long before the computer was invented. In fact, many of the rules used for evaluating evidence are part of common law, a body of unwritten codes developed in England primarily from judicial decisions based on custom and precedent. Common law forms the basis of the statutes of Canada and the United States, and of the other former colonies of the United Kingdom. By focusing on the development of laws and rules governing the admissibility of evidence in Canada and the United States, we can see how judges and lawyers have coped with the advent of the computer and what the consequences are for archivists charged with the care of computer records.

In any legal action, civil or criminal, the parties involved have the right to request that pieces of evidence be accepted by the court as proof of a point they wish

28 Naugler, The Appraisal of Machine Readable Records, p. 88-90.

to establish. The judge assigned to the case, known as the "trier of fact", must examine the evidence and determine if it is acceptable on the basis of statutes, common law and legal precedents. Once the evidence is declared admissible, it is judged for its weight or effectiveness. At certain times throughout this process, the other party in the action may seek to have the evidence declared inadmissible or without weight.

Anyone trying to introduce evidence of any kind into court 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 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. Over the years, therefore, case law (opinions of the court and judges as opposed to legislation) has developed a number of exemptions to the Hearsay Rule to allow certain kinds of evidence to be introduced.

One of these exemptions, the Business Records Exemption, directly affects machine readable records because it is under the provisions of this exemption that computer records or printouts are accepted as evidence. The exemption states that records made in the "usual and ordinary course of business", produced contemporaneously or as near as possible to the events and actions they describe, and not created in an attempt to misrepresent the facts, are possibly admissible as evidence. In the case of computer printouts, courts recognize that data may be stored for quite a while before producing a printout, so the rule of contemporaneous production is often considered to be a less important aspect.[29]

---

29 Michael C. Gemignani, Law and the Computer (Boston: CBP Publishing Company, 1982), p. 167.

Computer records are also subject to the Best Evidence Rule. This 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. If, for example, an office fire destroyed the original ledger entry, a photocopy of the cancelled cheque from the bank could be considered acceptable. By extension, in the absence of printouts, the magnetic tape is acceptable. Printouts are, however, preferred because of the ease of reading and authentication.

In any case, 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, machine readable records do fit into the Business Records Exemption. In order to do this, the party must lay out foundation evidence which will attest to the fact that the machine readable records were created in the course of regular business proceedings. Foundation evidence, particularly in the United States, is considered absolutely essential in establishing the regularity of record preparation; in Canada, the testimony of a single "expert witness" familiar with the record keeping system is likely to be as acceptable to the judge as foundation evidence.

Foundation evidence consists 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. William A. Fenwick and Gordon K. Davidson, in their 1978 article on computer evidence,[30] give nine points which they considered have to be answered to the judge's satisfaction in order to create proper foundation evidence for computer printouts. Figure 9 shows those points, which have since

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

been taken as the standard preparation for introducing machine readable records as evidence.

Foundation evidence for computer systems requires establishing:

(1) the reliability of the data processing equipment used to keep the records and produce the output;

(2) the manner in which the basic data was initially entered into the system (e.g., cards, teletype, etc.);

(3) that the data was so entered in the regular course of business;

(4) that the data was entered within a reasonable time after the events recorded by persons having personal knowledge of the events;

(5) the measures taken to ensure the accuracy of the data as entered;

(6) the method of storing the data (e.g., magnetic tape) and the safety taken to prevent loss of the data while in storage;

(7) the reliability of the computer programs and formulas used to process the data;

(8) the measures taken to verify the proper operation and accuracy of these programs and formulas;

(9) the time and mode of preparation of the printouts.

Figure 9: Elements necessary for proper foundation evidence of machine readable records in court[31]

In the United States, the onus of proving the reliability of the computer records as evidence lies with those who want to have them declared admissible evidence. If the foundation evidence is deemed inadequate, the records will not be admitted. In Canada, however, the procedure is not so clear.

31 Fenwick and Davidson, "Use of Computerized Business Records," p. 19.

The Canada Evidence Act is very vague on the subject of machine readable 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; according to Kenneth Chasse, the Act creates an extremely low threshold of admissibility. 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 it,[32] 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 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 single expert witness, while 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

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

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 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 was called upon to provide inactive government records in machine readable form to a judge, the government archivist would have to prepare foundation evidence documenting the procedures carried out during the processing and reference stages. The archivist could also be called as an expert witness in any case where the records of a defunct agency held within his 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 not changed.

In this age of concern over freedom of information and protection of privacy, however, archivst's involvement in the legal aspects of records may go well beyond guaranteeing their value as evidence. Archives could be placed in the position of either the defendant or the plaintiff in a court case. If an archival institution were

accused, for instance, of the violation of an individual's privacy, having allowed access to records in contravention of their donor's restrictions, the archivist would have to provide evidence for the archives' defense, including proper foundation evidence of any automated finding aids which might have been the cause of disclosure.

The most important consequence for archivists of the admissibility in court of computerized evidence is that it forces archivists once again to recognize the importance of full, accurate documentation of every phase of the record keeping continuum. Documentation from the design stage onwards not only is necessary in order to back up machine readable records in court but also gives archivists better control over their holdings, and makes processing and reference so much easier. And finally, in a rather indirect fashion, the elevation of archivists to the role of expert witness enhances their role in society: it makes of their function a vital element of the information-based culture developing through the use of computers.

Conclusion

The applicability of the principles of modern archival theory to machine readable records is a complex issue. Trudy Huskamp Peterson attempted to reassure archivists about the continuing relevance of the archival principles of appraisal, arrangement, description, and access in her seminal article "Archival Principles and the Records of the New Technology" (1984)[1], but in doing so, she did not discuss fully all of the theoretical implications of computer records. This study has attempted to answer the question "Does archival theory need to be redefined in the light of automation and of its products, machine readable records?"

As to their nature, there is no difference between a paper record and a machine readable record. Both Schellenberg's definition of records and Jenkinson's definition of archives disregard the physical medium of the record, and focus instead on the circumstances surrounding its creation, use, and preservation. As long as information is collected and recorded by a physical or moral person to carry out a practical activity and subsequently preserved for the use of its creator the result is a legitimate record.

The appraisal process as it applies to machine readable records shows that there is no fundamental difference in evaluating paper or electronic information. It is true that the process is split into two distinct sections, content analysis and technical analysis, the latter being quite specialized and requiring a high amount of computer expertise; but technical analysis is in principle nothing more than an extended version of the conservation analysis which is applied to paper records before they are accepted for accessioning. If we dissociate the practical concerns which refer to

---

[1] Trudy Huskamp Peterson, "Archival Principles and the Records of the New Technology," American Archivist 47 (Fall 1984): 383-93.

the records' medium from the theoretical principles of archival appraisal, we see that paper and computer records are appraised *exactly the same way*. Both are assessed for their primary and secondary values (administrative, legal, evidential and informational), their arrangement and suitability for use, and their relationship with other records already preserved. Clearly archival theory does not require major revisions with respect to appraisal of machine readable records.

On the contrary, the principles governing the records arrangement should be reformulated, not redefined. Since the physical arrangement of a file is irrelevant with computer records, the archivist does not need to be concerned with the original order of the records as long as the description he provides is accurate. An integrated system of finding aids linked to various levels of arrangement within and outside the fonds, as well as standardization of vocabulary and forms between institutions, is an ideal for paper records which becomes a necessity for machine readable records; the theoretical principles of archival description need to be applied to their fullest extent to make records accessible.

The same reflection can be made as to reference service. Whether the archivist is responsible for paper records or computer records, he must adhere to the twin principles of equal access and full access. His duty is to make archival holdings available to users to the best of his abilities and resources and in accord with the principles and regulations governing freedom of information, copyright, restrictions, and protection of personal privacy. As with the case of description, accuracy and attention to detail have a vital role in the communication of machine readable records and the principles must be applied with more care in order to obtain the same result, but their theoretical foundation is unvaried.

Archival theory does not need to be redefined to accommodate machine readable

records: it is valid for and can service both paper and magnetic media. However, it needs to be extended to cover new areas and to solve problems which did not exist before the use of computers Overall, the practical applications of the fundamental archival principles must be analyzed and developed. The major changes brought about by machine readable records are in archival practice, not in archival theory.

Furthermore, the new attitude towards information determined by the advent of computers has tightened the link between records managers and archivists and determined the acceptance of the continuum model of record keeping, much more appropriate for machine readable records than the life cycle model; it ensures not only the preservation of electronic records through early appraisal, but also the smooth flow of records into archives through improved classification and scheduling. The management of machine readable records has also determined a closer relationship between archivists and librarians; library science has much to offer in the way of descriptive techniques and standards to handle machine readable data files, and its methods can be easily adapted to suit archival needs. Therefore, it can be said that the techniques for the treatment of machine readable records constitute the one area of common ground where records managers, librarians, and archivists can meet and share ideas on equal terms as information managers; such cooperation ought to be promoted because it benefits everyone.

The handling of computers in general and machine readable records in particular has shown archivists that to fulfill their mandate in the society of the future they can no longer accept a passive role in record keeping and be able to preserve society's documentary heritage with any sort of accuracy. Archivists are beginning to realize that they need to become actively involved in all the stages of the documentation process; to intervene in their development; to research and study new practices for the application of archival principles; to explore new issues

and phenomena; to participate in and favour the creation of legislation allowing for a better control of information, records, and archives, and providing them with a sure ground for the exercise of their several functions of keepers and communicators of the memory of their country. It is not a simple task, but the archival profession has in the past shown that it can rise to meet a challenge such as that presented by machine readable records and the future should prove to be no different.

BIBLIOGRAPHY

ALLDREDGE, Everett O. "The Rocky Road in Scheduling and Disposition." Records Management Journal 8,2 (Summer 1970): 2-8.

_____. "Inventorying Machine Readable Records." American Archivist 35 (July 1972): 337-46.

ANDERSON, Michael. "The Preservation of Machine-Readable Data for Secondary Analysis." Archives XVII (1985): 79-93.

ATHERTON, Jay. "From Life Cycle to Continuum: Some Thoughts on the Records Management - Archives Relationship." Archivaria 21 (Winter 1995-86): 43-51.

BAILEY, Catherine A. "Computers and Archives: A survey of the Literature." Paper prepared for Archival Studies 500 course, University of British Columbia, 3 April 1987.

BASYE, Paul E. "Computer Printouts as Evidence." Records Management Journal 10,1 (Summer 1972): 20-2.

BELDING, R.N. et al. "Kentucky's Machine-Readable Challenge." Provenance III (Spring 1985): 58-69.

BELL, Lionel. "The Archival Implications of Machine Readable Records." Archivum 26 (1974): 85-92.

BIGELOW, Robert P. "Can you prove it -Legally?" Infosystems 29 (February 1982): 94,96.

BOLES, Frank. "Sampling in Archives." American Archivist 44 (2) (Spring 1981): 125-30.

BRANDHORST, Ted and Martha Williams. "Machine-Readable data bases: copyright status." Bulletin of the American Society for Information Science 4 (4) (April 1978): 31.

BRICHFORD, Maynard J. Archives and Manuscripts: Appraisal and Accessioning. Chicago: Society of American Archivists, 1977.

BOHNEN, Linda S. "A Reply to Mr. Robertson." Archivaria 6 (Summer 1978): 12-5.

BOOMS, Hans. "Society and the Formation of A Documentary Heritage: Issues in the Appraisal of Archival Sources." Translated by Hermina Joldersma and Richard Klumpenhouwer. Archivaria 24 (Summer 1987): 79-107.

BRONSTEIN, Daniel A. and Daniel Engelberg. "A Preliminary Assessment of the Reception of Computer Evidence: Report of the Computer Evidence Survey Project." Jurimetrics Journal (Spring 1981): 329-32.

BROWN, Thomas Elton. "Archives Law and Machine-Readable Data Files: A Look at the United States." No date. (Photocopy.)

_____. "The Society of American Archivists Confronts the Computer." American Archivist 47 (Fall 1984): 366-82.

_____. "Appraisal in the Information Age." Paper presented at the 12th Annual Conference of the Association of Canadian Archivists, Hamilton, Ontario, 1-5 June, 1987.

_____. "The Evolution of an Appraisal Theory for Automated Records." Archival Informatics Newsletter 1 (3) (Fall 1987): 49-51.

BUREAU OF CANADIAN ARCHIVISTS. Towards Descriptive Standards: Report and Recommendations of the Canadian Working Group on Archival Descriptive Standards. Ottawa: Bureau of Canadian Archivists, 1985.

CAMERON, Ross J. "Appraisal Strategies for Machine Readable Case Files." Provenance I (Spring 1983): 49-55.

CARROLL, M.E. "A Perspective of Machine-Readable Archives and Public Service." Automatic Data Processing in Archives 2 (1) (1976): 7-10.

CHALOU, George. "Reference." In A Modern Archives Reader, pp. 257-63. Edited by Maygene F. Daniels and Timothy Walch. Washington, D.C.: National Archives Trust Fund Board, 1984.

CHARMAN, Derek. "Archives and Records Management - an interface?" Journal of the Society of Archivists 6,7 (April 1981): 423-7.

CHASSE, Kenneth. "The Legal Issues Concerning the Admissibility in Court of Computer Printouts and Microfilm." Archivaria 18 (Summer 1984): 166-201.

_____. "A Reply to A.F. Sheppard's Commentary on the Admissibility in Court of Computer Printouts." Archivaria 20 (Summer 1985): 158-61.

CLUBB, Jerome M. "Archival Implications of Technological and Social Change." In Archivists and Machine Readable Records, pp. 233-41. Edited by Carolyn L. Geda, Erik W. Austin, and Francis X. Blouin. Chicago: Society of American Archivists, 1980.

COOK, Michael. Archives and the Computer. London: Butterworths, 1980.

_____. The Management of Information from Archives. Brookfield, Vermont: Gower Publishing Company, Limited, 1986.

COOK, Terry. "The Tyranny of the Medium: A Comment on Total Archives." Archivaria 9 (Winter 1979-80): 141-51.

DeHETRE, J. David. "Data Processing Evidence - Is it Different?" Chicago-Kent Law Review 52(3) (1976): 567-89.

"Description Form: Abstract for Machine Readable Data File." Photocopied handout distributed at the Workshop on Machine Readable Records, School of Library, Archival and Information Studies, Vancouver, B.C., 2-3 March 1987.

DOLLAR, Charles M. "Appraising Machine-Readable Records." American Archivist 41 (October 1978): 423-30.

DODD, Sue A. "Cataloging Machine-Readable Data Files - A First Step?" Drexel Library Quarterly, 13(1) (January 1977): 48-69.

_____. Cataloging Machine-Readable Data Files - An Interpretive Manual. Chicago: American Library Association, 1982.

_____. "Problems in Describing Machine Readable Data Files." Paper presented at the 10th Annual IASSIST Meeting, Ottawa, Ontario, May 1984.

_____. "Cataloging Microcomputer Files According to AACR2 and Guidelines for the Descriptive Cataloging of Microcomputer Files." Paper presented at the IASSIST Workshop, Ottawa, Ontario, May 15 1984.

DUCHEIN, Michel. "Theoretical Principles and Practical Problems of *Respect des Fonds* in Archival Science." Archivaria 16 (Summer 1983): 64-82.

DURANTI, Luciana. "Towards Automated Records Paleography." Paper presented at the 12th Annual Conference of the Association of Canadian Archivists, Hamilton, Ontario, 1-5 June 1987.

EWART, J. Douglas. "Documentary Evidence: The Admissibility at Common Law of Records Made Pursuant to a Business Duty." Canadian Bar Review 59 (1981): 52-75.

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

FIERHELLER, George A. "A Canadian Viewpoint on the Problems of Transborder Data Flows." In Transborder Data Flows and the Protection of Privacy: Proceedings of a Symposium held in Vienna, Austria 20th -23rd September, 1977, pp. 168-73. Paris: Organisation for Economic Co-operation and Development, 1980.

FIRNBERG, D. "Security Standards for Transborder Data Flows." In Transborder Data Flows and the Protection of Privacy: Proceedings of a Symposium held in Vienna, Austria 20th -23rd September, 1977, pp. 65-8. Paris: Organisation for Economic Cooperation and Development, 1980.

FISHBEIN, Meyer H. "Appraisal of Twentieth Century Records for Historical Use." Illinois Libraries 52 (February 1970): 154-62.

_____. "Appraising Information in Machine-Language Form." American Archivist 38 (January 1975): 31-43.

_____. Guidelines for Administering Machine Readable Archives. Washington, D.C.: ADP Committee of the International Council on Archives, 1980.

_____. "The 'Traditional' Archivist and the Appraisal of Machine-Readable Records." In Archivists and Machine Readable Records, pp. 56-64. Edited by Carolyn L. Geda, Erik W. Austin, and Francis X. Blouin. Chicago: Society of American Archivists, 1980.

FOX, Lisa L. "Archival Preservation in the Age of Technology." Provenance III (Spring 1985): 23-9.

FREESE, J. "Social and Cultural Implications of Transborder Data Flows." In Transborder Data Flow Policies: Papers presented at the IBI Conference on Transborder Data Flow Policies, Rome [Italy], 23-27 June 1980, pp. 541-9. New York: UNIPUB, 1980.

GARZON, G. and E. Vilarino. "Information and Privacy Protection in Transborder Data Flows: The Rights Involved." In Transborder Data Flows and the Protection of Privacy: Proceedings of a Symposium held in Vienna, Austria 20th -23rd September, 1977, pp. 302-7. Paris: Organisation for Economic Co-operation and Development, 1980.

GAVREL, Katherine, and McDonald, John. Appraisal Guidelines in the Machine Readable Archives Division. Ottawa: Public Archives of Canada, 1981.

GAVREL, Sue. "Issues Associated with Accessing Machine-Readable Data." in Archives, Automation and Access: proceedings of an interdisciplinary conference at the University of Victoria, Victoria B.C March 1-2 1985. Victoria: University of British Columbia Press, 1985.

_____. "Preserving Machine-Readable Archival Records: A Reply to John Mallinson." Archivaria 22 (Summer 1986): 153-55.

GEDA, Carolyn L. "Social Science Data Archives." American Archivist 42 (April 1979): 158-66.

GELLER, Sidney B. Care and Handling of Computer Magnetic Storage Media. Washington, D.C: U.S. Government Printing Office, 1983.

GEMIGNANI, Michael. Law and the Computer. Boston: CBP Publishing Company, 1982.

HAYWARD, Robert J. "Federal Access and Privacy Legislation and the Public Archives of Canada." Archivaria 18 (Summer 1984): 47-57.

HEDSTROM, Margaret. Archives and Manuscripts: Machine Readable Records. Chicago: Society of American Archivists, 1983.

HOLBERT, Sue E. Archives and Manuscripts: Reference and Access. Chicago: Society of American Archivists, 1977.

HOPKINS, Mark. "Admissibility: A Review and Update." Archivaria 20 (Summer 1985): 162-65.

HULL, Felix. The Use of Sampling Techniques in the Retention of Records: A RAMP Study with Guidelines. Paris: UNESCO, 1981.

INTNER, Sheila S. "Suggestions for the Cataloging of Machine-Readable Materials." Library Resources & Technical Services 17 (4) (Oct/Dec 1983): 366-70.

JENKINSON, Sir Hilary. A Manual of Archive Administration . 2nd Revised Edition. London: Percy Lund, Humphries & Co. Ltd, 1965.

KEON, Jim. "The Canadian Archivist and Copyright Legislation." Archivaria 18 (Summer 1984): 91-8.

KEPLEY, David R. "Sampling in the Archives: A Review." American Archivist 47 (Summer 1984): 237-42.

KEPLINGER, M. "Information and Documentation Services." In Transborder Data Flows and the Protection of Privacy: Proceedings of a Symposium held in Vienna, Austria 20th -23rd September, 1977, pp. 153-9. Paris: Organisation for Economic Co-operation and Development, 1980.

KIRBY, M.D. "Access to Information and Privacy: The Ten Information Commandments." Archivaria 23 (Winter 1986-87): 4-15.

KNOPPERS, Jake. "The Impact of Privacy, Freedom of Information and Related Legislation and Regulations on the Canadian Federal Government." Automatic Data Processing in Archives 3,2 (1980): 45-9.

_____. "Towards a Canadian Electronic Cultural Heritage." Ottawa: Public Archives of Canada, 1983.

_____. "Integrating Technologies = Integrating Disciplines?" Records Management Quarterly 17,1 (January 1983): 5-7, 26.

KYER, C. Ian. "Computer Records as Courtroom Evidence." Computer Law 1 (8) (August 1984): 57-60.

_____. "Computer Records as Courtroom Evidence." Computer Law (Sept-Oct 1984): 76-80.

_____. "The Federal Access to Information Act: A Statute to be aware of." Canadian Computer Law 2 (9) (July 1985): 167-73.

_____. "The Federal Privacy Act." Canadian Computer Law Reporter 2 (10) (August 1985): 188-91.

_____. "Ontario's proposed Freedom of Information and Protection of Individual Privacy Act: Businessmen beware." Canadian Computer Law Reporter 3 (1) (November 1985): 5-7.

_____. "Information, Computers and the Law." Paper presented at the 12th Annual Conference of the Association of Canadian Archivists, Hamilton, Ontario, 1-5 June 1987.

LA FORCE, Gina. "Archives and Copyright in Canada: An Outsider's View." Archivaria 11 (Winter 1980-81): 37-51.

LEWINSON, Paul. "Archival Sampling." American Archivist 20 (October 1957): 291-312.

LONGSWORTH, M.A. "Transborder Data Flows and the Protection of Privacy." In Transborder Data Flows and the Protection of Privacy: Proceedings of a Symposium held in Vienna, Austria 20th -23rd September, 1977, pp. 74-80. Paris: Organisation for Economic Co-operation and Development, 1980.

Machine Readable Archives Bulletin. 1 (3) (Fall 1983).

MALLINSON, John. "Preserving Machine Readable Archival Records for the Millennia." Archivaria 22 (Summer 1986): 147-52.

McDONALD, John. "An Approach to the Identification and Scheduling of EDP Data." Paper presented at the 46th Annual Conference of the Society of American Archivists, Boston, Mass., 18-22 October, 1982.

McDONALD, J., ed. Scheduling Data in Automated Systems - A Final Report. Ottawa: Public Archives of Canada, 1985.

MITCHELL, Thornton. Norton on Archives. Carbondale, Illinois: Southern Illinois University Press, 1975.

MOTZ, Arlene. "Applying Records Management Principles to Magnetic Media." Records Management Quarterly 20,2 (April 1986): 22-6.

MULLER, S; Feith, J.A.; and Fruin, R. Manual for the Arrangement and Description of Archives. Translated by Arthur H. Leavitt. New York: H.W. Wilson Company, 1968.

NAUGLER, Harold. The Archival Appraisal of Machine Readable Records: A RAMP Study with Guidelines. Paris: UNESCO, 1983.

PAGE, Don. "The Access Dilemma." Archivaria 8 (Summer 1979): 135-8.

PETERSON, Trudy Huskamp. "Archival Principles and the Records of the New Technology." American Archivist 47 (Fall 1984): 383-93.

"Preserving the Canadian Electronic Cultural Heritage." Ottawa: Public Archives of Canada, 1982.

PUBLIC ARCHIVES OF CANADA. Machine Readable Archives Drug Use Files. Ottawa: Public Archives of Canada, 1978.

_____. The Machine Readable Archives: An Overview of its Operations and Procedures. Ottawa: Ministry of Supply and Services, 1980.

_____. Machine Readable Archives Alcohol, Drug and Tobacco Use Files. Ottawa: Public Archives of Canada, 1982.

_____. Machine Readable Archives Recreation and Leisure Data Files. Ottawa: Public Archives of Canada, 1983.

_____. Guidelines for the Scheduling of Data in Automated Systems. Ottawa: Public Archives of Canada, 1986.

PUGH, Mary Jo. "The Illusion of Omniscience: Subject Access and the Reference Archivist." In A Modern Archives Reader, pp. 254-77. Edited by Maygene F. Daniels and Timothy Walch. Washington, D.C.: National Archives Trust Fund Board, 1984.

REITMAN, Alan. "Freedom of Information and Privacy: The Civil Libertarian's Dilemma." American Archivist 38 (1975): 501-08.

RIEGER, Morris. "Archives and Automation." American Archivist 29 (January 1966): 109-11.

ROBBIN, Alice. "The Pre-Acquisition Process: A Strategy for Locating and Acquiring Machine Readable Data." Drexel Library Quarterly 13,1 (1977): 21-42.

_____. "Understanding the Machine Readable Numeric Record." Midwestern Archivist IV (1979): 5-23.

ROBERTS, David. "The Disposal and Appraisal of Machine Readable Records - From the Literature." Archives and Manuscripts 13 (May 1985): 30-8.

ROBERTS, John. Legislation on Public Access to Government Documents. Ottawa: Ministry of Supply and Services, 1977.

ROBERTSON, Gordon. "Confidentiality in Government." Archivaria 6 (Summer 1978): 3-11.

ROPER, Michael. "The Changing Face of the File." Archives XIV (Spring 1980): 145-50.

_____. "Advanced Technical Media: the conservation and storage of audio-visual and machine-readable records." Journal of the Society of Archivists 7 (October 1982): 106-12.

ROSENKRANTZ, Gerald J. "What Should the Role of the Records Manager be in the ADP Shop?" Records Management Journal 9,3 (Autumn 1971): 2-6.

ROTHSTEIN, Samuel. "Across the Reference Desk: 100 Years of Reference Encounters." Canadian Library Journal (October 1977): 391-8.

ROWE, Judith S. "Privacy Legislation: Implications for Archives." In Archivists and Machine Readable Records, pp. 193-203. Edited by Carolyn L. Geda, Erik W. Austin, and Francis X. Blouin. Chicago: Society of American Archivists, 1980.

SCHELLENBERG, Theodore R. Modern Archives: Principles and Techniques. Chicago: University of Chicago Press, 1956; reprint ed., Chicago: Midway Reprint, 1975.

SCHELLENBERG, T.R. "The Appraisal of Modern Public Records." Bulletins of the National Archives 8 (October 1956): 254-59.

SHEPPARD, A.F. "Records and Archives in Court." Archivaria 19 (Winter 1984-85): 196-203.

SOMMERVILLE, I. Software Engineering. Reading, Massachusetts: Addison-Wesley Publishing Company, 1982.

SPROWL, James A. "Evaluating the Credibility of Computer-Generated Evidence." Chicago-Kent Law Review 52 (3) (1976): 547-66.

TASK FORCE ON PRIVACY AND COMPUTERS. Privacy and Computers. Ottawa: Information Canada, 1972; reprinted 1974.

TAYLOR, Hugh A. The Arrangement and Description of Archival Materials. New York: International Council on Archives, 1979.

TENER, Jean. "Accessibility and Archives." Archivaria 6 (Summer 1978): 15-31.

THEXTON, J.E. "Archival Potential of Machine Readable Records in Business." American Archivist 37 (January 1974): 37-42.

THIBODEAU, Kenneth. "Machine Readable Archives and Future History." Computers and Humanities 10 (2) (March/April 1976): 89-92.

"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, pp. 575-680. New York: UNIPUB, 1980.

TREASURY BOARD OF CANADA. Administrative Policy Manual: Chapter 440: Electronic Data Processing. Section 7. Ottawa: December, 1978.

VOGEL, Marguerite A. "Transborder Data Flow: A Canadian Focus." M.A. Thesis, Simon Fraser University, 1984.

WALNE, Peter, ed. Dictionary of Archival Terminology: ICA Handbook Series Volume 3. New York: K.G. Saur, 1984.

WARNER, Robert M. and Francis X. Blouin. "Some Implications of Records in Machine-Readable Form for Traditional Archival Practice." In Archivists and Machine Readable Records, pp. 242-48. Edited by Carolyn L. Geda, Erik W. Austin, and Francis X. Blouin. Chicago: Society of American Archivists, 1980.

WESTIN, Alan F. "The Long Term Implications of Computers for Privacy and the Protection of Public Order." In Computers and Privacy in the Next Decade, pp. 167-81. Edited by Lance J. Hoffman. New York: Academic Press, 1980.

ZEISSET, Paul. "Census Bureau Confidentiality Practices and their implications for Archivists." In Archivists and Machine Readable Records, pp. 204-16. Edited by Carolyn L. Geda, Erik W. Austin, and Francis X. Blouin. Chicago: Society of American Archivists, 1980.

ZINN, Howard. "Secrecy, Archives and the Public Interest." Boston University Journal 19 (1971): 37-44.

Appendix A: Glossary of Terms[1]

BIT

An abbreviation of the term BInary digiT, it is the smallest unit of electronically stored information It corresponds to a number in the binary numbering system (Base Two), either 0 or 1, where 0 usually stands for the absence of an electronic pulse, while 1 represents the presence of a pulse.

BLOCK

A section of information recorded onto a disc or tape, then separated from other blocks by a small section of blank (no-recorded) tape or disc which is known as the inter-record gap (IRG). A block may be composed of many logical records, conversely, or a logical record may stretch over many blocks. *Blocked*, therefore, means data recorded in blocks.

BYTE

A string of adjacent bits, usually 8 in number, which are operated upon as a single storage unit in Electronic Data Processing. A *character* is represented by one byte.

CHARACTER

A single element of a set of letters, digits, or other symbols such as punctuation and spaces used in the representation of *data*. Each character is composed of a unique arrangement of bits. For example, in, EBCDIC, a particular code used mainly with IBM computers,

A is represented by	1100 0001
B is	1100 0010
C is	1100 0011
Z is	1110 1001

CODE

Rules for representing instructions and/or information in a symbolic form. A code can be established by a user to represent his information in a form that can be processed by the computer. It can also refer to the standard codes such as EBCDIC used to represent characters in binary form.

CODEBOOK

A book or manual containing explanations of the various codes used to represent information in abbreviated or numeric form within a datafile.

DATA

Information represented in a formal structure, such as a code or a particular

[1] The majority of the following definitions have been taken from Harold Naugler's The Archival Appraisal of Machine Readable Records: A RAMP Study with Guidelines (1983), Michael Cook, Archives and the Computer (1980), and Margaret L. Hedstrom, Archives and Manuscripts: Machine Readable Records (1983).

format, so that it may be handled by an automatic computer process. It is distinct from the results of the computer process, and the commands that control that process.

DATA BASE (or DATABASE)
A collection of carefully integrated datafiles stored in a central location so that they are available to many different users at the same time in order to carry out a variety of applications. Users may have access to all or only a part of the database's contents.

DATA ELEMENT
Bytes or characters used in combination to refer to a single, separate item of information such as a name, an address, or a telephone number.

DATA FIELD -see FIELD

DENSITY
A term referring to the number of bits in a single linear *track* of magnetic tape or disc. Noted as bits per inch (bpi), characters per inch (cpi), or frames per inch (fpi). Most common densities are 6250 and 1600 bpi; older tapes or discs may be recorded at 556 or 800 bpi.

DISC (or DISK)
A storage device consisting of a circular plate with a magnetic coating on both of its sides, which then rotates at high speeds past several "read/write heads" which move information to and from the computer. Discs may be "hard" or "floppy".

DOCUMENTATION
One or more descriptive documents which explain the physical hardware and software (programs) needed to create, maintain and use a datafile. Also indicates the arrangement, content and possibly the coding of the data contained in the file.

DRUM
Cylindrical magnetic storage device; data is stored by selective magnetization of parts of the curved surface.

ELECTRONIC DATA PROCESSING (EDP)
Manipulation by a computer of any kind of information stored in an electronic form. Synonymous with AUTOMATIC DATA PROCESSING, which is the term used primarily in the United States.

FIELD
A subdivision of a *logical record* which is allocated for a specific category of data, often in the form of a data element. Fields are usually of two kinds: fixed (assigned a set number of positions), or variable (the length may vary from field to field.).

FILE STRUCTURE

Refers to the way a particular file is organized. It is rectangular if each logical record contains information on one observation or unit of analysis; hierarchical if each logical record contains information on more than one unit of analysis.

FREQUENCY (IES)

A statistical procedure for examining the distribution and occurrence of a datafile's variables and their values.

GENERATION

The stage in an update cycle in which a file was created. The file and the most recent update is the current generation, son, or daughter. The previous file or update is the first, father, or mother. Prior to that, they are the second, grandfather, or grandmother generation.

HARDWARE

The actual physical equipment which makes up a computer system as distinct from the *software* or programs used to run it.

INPUT/ OUTPUT

The peripheral equipment (i.e card readers, keyboards, printers, etc.) used to communicate with a computer. May also refer to the data involved in these communications as well as the medium carrying it, or the activities involved.

LOGICAL RECORD

A group of related data elements referring to a single person, thing, place or event and treated as a unit. May have a specified number of characters (fixed length) or may have varying numbers of characters (variable).

OPERATING SYSTEM

Software which controls the execution of a computer's programs. It is linked to a specific type of machine and also provides debugging, compiling and file handling processes.

PARITY

A means for verification of recorded data and detection of errors. One bit per byte is designated as a parity bit and parity is set to 0 or 1 for the whole file. Through binary addition, the total value of each of the bits in a byte must equal the standard value. Totals of 0 are even parity; 1 are odd parity.

PROGRAM

A set of instructions given to the computer in order to solve a particular problem or to carry out a set of operations.

RECORD LAYOUT

A diagram or list of contents of a logical record, describing the information found in each field, the length of the field, and its position in the record. Also known as the *file layout*.

SOFTWARE

Any set of computer programs which solve a problem or carry out any other functions of data processing, along with its related documentation.

TRACK

One of 7 or 9 parallel rows of bits along the length of a magnetic tape. Also a series of concentric circles on a disc or drum where the data is recorded.

Appendix B: Modified Example of a Public Archives of Canada "System Overview"

From the Government of British Columbia, Records Management Branch

---

SCHEDULE REGISTRY NO.: 098-86

RECORDS RETENTION AND DISPOSAL SCHEDULE

INFORMATION SYSTEM DESCRIPTIVE INVENTORY

MINISTRY, CROWN CORPORATION, OR AGENCY:
Ministry of Attorney General

DIVISION, BRANCH, SECTION, OR OFFICE:
Land Title Branch

SYSTEM TITLE:
Automated Land Title Office System (ALTOS)

APPLICATION TITLE:
Automated Land Title Office System (ALTOS)

PURPOSE:
Automation of the manual process of land title registration, documenting the ownership of land and the charges against it.

INFORMATION CONTENT:
ALTOS includes the date on which the application for registration was received, the new registration number and previous title number, legal description of the land, parcel identifier (PID) number, registered owner's name and address, and charges against the land in the form of liens and/or mortgages. It also includes legal notations, such as easements, restrictions on the land's development, and statutory restrictions.
The Power of Attorney Index is included as part of ALTOS but in a separate form from the electronic title. It shows the name of the person holding the power of attorney, the kind of power, its operational dates, and the relevant land registration numbers.

RETENTION AND DISPOSAL SUMMARY:

Cancelled indefeasible titles will be copied to COM after 5 years on the system. One month after the microfilm has been checked, the electronic titles will be deleted from ALTOS.

The Power of Attorney Index will be retained on line permanently.

For Transitory Electronic Data Processing (EDP) Records (Operational), see records schedule recommendation and authority 016-86.

INPUTS/ PROCESSES/ OUTPUTS:

Inputs: ALTOS input is classified in LTO ORCS, section 2, primary 24010:

24010-05	Indefeasible Titles -Paper Working Copies
-05-02	Correction/ Amendment Working Copies
-05-03	Conversion/ Cancellation Working Copies
-06	Short Legal Description (SLD) Forms
-06-02	Subdivided SLD Forms
-06-03	Unsubdivided SLD Forms
-07	Correction/ Amendment Forms

Processes: Each Land Title Office (LTO) maintains its own database of electronic titles. Upon receipt, applications are examined by land title examiners. If the application is registerable, the titles are updated to reflect the information contained in the application. On transfers, some information may be brought forward from the old title, combined with the new update, and then proofread. The old title is cancelled in most cases, but in some cases only a part of it or an interest in the property is being transferred, leaving the old title live but showing the transfer.

The paper documentation which is created the title update or transfer is filed. The electronic title is the official title supported by these filed documents.

Searches may be performed by extracting specific information from a title, i.e.an owner name search, which results in an alphabetic listing of owners names. A specific name may be selected which would result in a display of title and/or charge numbers; a charge search which would display the charges on a title; a pending number search which would display numbers pending.

Through these procedures, the title itself is not changed, although the format in which the information is requested and printed varies.

Outputs: ALTOS output is classified in LTO ORCS, sections 2 and 3:

24010-03	Current Certificates of Indefeasible Title -Manual System (Form 21) and ALTOS
-04	Cancelled and Converted Certificates of Indefeasible Title -Manual System (Form 21) and ALTOS
24020-05	Duplicate Indefeasible Title (Form 21a)
24030-06	Provisional Indefeasible Title Certificates (Form 21a)
25510-04	State of Title Certificates Requested
25570-03	Cancelled Indefeasible Titles Requested

In the future, cancelled titles may be converted to COM fiche or stored off-line. Currently, they are all stored on line.

SOFTWARE ENVIRONMENT:

ALTOS is an in-house information management system designed specifically for the Land Title Branch by BCSC and approved for Land Titles by Terrry Carlow, former land title registrar for New Westminster, and H.T. Kennedy, former Director of the Office of Land Titles. ALTOS automates manual land title registration procedures which have been used in LTOs for the past 60 years. It is written in the PL/I programming language, with an ADF security subsystem.

HARDWARE ENVIRONMENT:

ALTOS runs on an IBM 3092 mainframe computer at the British Columbia Systems Corporation (BCSC).

SYSTEM MILESTONES:

System design approved December 1980. First introduced into Victoria LTO on 8 August 1983. Other dates of introduction to LTOs: New Westminster, April 1984; Kamloops, March 1985; Vancouver, November, 1985; Prince George, March 1986; Prince Rupert, June 1986; and Nelson, November, 1986. Expected completion date for conversion of all hardcopy indefeasible titles: 1990.

Those outside the Land Title Offices (i.e., government agents in regions with no LTO and various government ministries) permitted access to ALTOS in March 1987.

USER CONTACTS (POSITION):

Lois Courtright, Provincial Computer Coordinator, Office of the Director of Land Titles

Registrars of the Individual Land Title Offices:

Kenneth Jacques, Victoria
David Ellis, Vancouver
Jack Raven, New Westminster
Jack Groves, Kamloops
Anne Marion, Nelson
William Gandy, Prince George
Ian Smith, Prince Rupert

TECHNICAL CONTACTS (POSITION):

Del Forbes, Client Services Representative, BCSC
Dianne Glencross, Program Analyst, BCSC

DOCUMENTATION/ MANUALS CROSS REFERENCES:

BCSC, B.C. Land Titles Office Project, 1980.
Operations Manuals - unique to each office.
"Reference Manual for the Land Title Project," 1986/09/15

ADMINISTRATIVE OR OPERATIONAL RECORDS CLASSIFICATION:

Schedule Registry No.:Primary and Secondary Numbers and Titles: LTO ORCS 098-86

24010-03	Current Certificates of Indefeasible Title -Manual System (Form 21) and ALTOS
24010-04	Cancelled and Converted Certificates of Indefeasible Title -Manual System (Form 21) and ALTOS

CLASSIFICATION SYSTEM CROSS REFERENCES (OTHER RELATED RECORDS)

Schedule Registry No.: Primary and Secondary Numbers and Titles: ARCS 001-86

6100-6180 Electronic Data Processing (EDP) Systems Applications Block