The Disposal and Appraisal of Machine-Readable Records — From The Literature

David Roberts

This is substantially the text of a paper given at a seminar on machine-readable archives, held by the Sydney Branch of the Australian Society of Archivists on 19 November 1983. The paper consists of a survey of the literature relating to the disposal and appraisal of machine-readable records, followed by some conclusions about procedures and the identification and discussion of four categories of appraisal criteria, as they can be drawn from the literature.

There is now a considerable body of literature dealing with archives and computers. The bulk of this literature, however, has been concerned with the ways in which computers can be used in the management of archives, while rather less has been written about the various aspects of managing machine-readable archives. This paper will examine the treatment of the disposal and appraisal of machine-readable records in this small but important body of literature, first of all by a brief tour of some of the more important contributions, and secondly by drawing out what we can learn from them in the areas of disposal procedures and appraisal criteria.

Meyer Fishbein's paper, "Appraising Information in Machine Language Form", which appeared in *The American Archivist* in 1972,[1] could indeed be described as seminal. It was the first serious discussion of the broad approach that archivists should take in appraising machine-readable records, that is, the extent to which they can be dealt with like any other media, and the extent to which archivists' thinking would have to change. He concluded that there was value in each approach:

> The basic rule applies to both machine language and conventional records: They will be retained for their research value if the information is unique and has enduring value for significant research.[2]

At the same time he enumerated some of the principal developments in the kinds of machine-readable records being produced up to that time and

the ways in which they would be likely to be used by future researchers, as compared with records in conventional formats.

Finally, he offered a set of procedures, including some basic criteria for retention, for the appraisal of machine-readable records.[3] Although aimed specifically at records containing statistical source material, the procedures and criteria had a broader application, as we shall see later in this paper.

In a paper presented in 1974 to an international seminar on "Automatic Data Processing in Archives" sponsored by UNESCO, ICA and the Public Record Office, M.E. Carroll, of the Public Archives of Canada, discussed the PAC's then brief experience with machine-readable archives.[4] The appraisal process for machine-readable records was described as "the application of the traditional criteria", with "a number of new administrative practices which require a different perspective".[5] A key feature of the PAC's approach at this stage was the requirement to keep master files and extract files based on master files, while process or working files were considered to be of little value. A further feature of note was the requirement that files of long term value be kept in the format and condition in which they were received.

Charles Dollar's article, "Appraising Machine-Readable Records" was published in *The American Archivist* in 1978,[6] and was firstly a description of the then current appraisal practices in the Machine Readable Archives Division of the National Archives and Records Service.

As if to echo Fishbein's conclusion noted earlier, the NARS procedures displayed a mixture of conventional practice and of considerations to be found only in the area of machine-readable archives.[7] The criteria for assessing the value of the records being appraised would not be out of place in a traditional manual of archival appraisal. Typical of these criteria are the questions: "Does the record have legal, evidential or informational value?" and "Do similar records exist elsewhere?"[8]

The requirements that the record be readable accompanied by adequate support documentation, the extent of its dependence on particular software, and the likely cost of accessioning and preserving the record, on the other hand, are factors which are not normally encountered in the appraisal of conventional records.

The second part of Dollar's article discusses the problems posed for disposal arrangements typifed by the PAC criteria by the development of data base management systems: an on-line or interactive computer environment in which it can be argued that such concepts as master files and processing files may have little meaning or do not exist at all.

Finally, Dollar suggests that, far from reducing the preservation costs of machine-readable archives, the continual development of high-density formats may increase such costs, because of the possiblity that:

One or two very valuable files may be dispersed throughout the disk in such a

way that it would be most economical to accession all the data, the garbage as well as the treasure, because the expense of having a computer assemble sub-sets of data that comprise the one or two valuable files may be prohibitive.[9]

On the other side of the Atlantic, the British contribution to the literature really starts with Lionel Bell's paper presented to the 8th International Congress on Archives in 1976 and published in *Archivum* in 1979.[10] In the area of appraising machine-readable records, Bell's principal argument is that "with computer files archivists are for the first time dealing on a large scale with information as such rather than as the by-product of individual administrative transactions".[11] The records which should be retained, he argues, are

material which the future user will wish to transmute by further processing into source material for his historical or other work, ... information which might provide evidence about the common and basic features of human life, their inter-relationships, their rate of change, their stability.[12]

Thus, the administrative context in which the material was created loses its significance in relation to these archives, and the extent to which the information can be manipulated becomes a matter of particular importance.

A similar observation provides the starting point for Michael Roper's examination of the area: "The Changing Face of the File: Machine-readable Records and the Archivist", published in *Archives* in 1980.[13] "The main purpose of preserving archival documents is to secure the information which they contain", he argues, or "the medium is *not* the message".[14] Roper emphasises the "when" rather than the "how" of appraising machine-readable records, insisting that the archivist should "make an early approach to those within his organization who might create such records to make them aware of potential historical value"[15] and such records should be earmarked for retention and adequately documented early in their lives.

This is, of course, sound records management practice in relation to records of any format. For machine-readable records, Roper argues, it is crucial, since re-use, physical deterioration and the loss of support documentation can quickly take their toll of such records[16]

Finally, dealing a blow to the concept of preserving the original physical record, Roper suggests the adoption by machine-readable archives of their own standard format, so that records in incompatible formats can be usefully preserved. The feasibility and cost of such reformatting would be an important factor in the appraisal process.[17]

Archivists and Machine-Readable Records, the proceedings of a conference held in 1979 on the archival management of machine-readable records[18] contains a number of useful contributions to the literature of disposal and appraisal.[19]

Allan Bogue discusses, in the context of social research data in machine-readable form, the age-old archival problem of being able to predict future research trends and methods. He calls for unaggregated data to be preserved, rather than aggregated data, in order to permit future researchers to use it for purposes and in ways entirely different from its original use.[20]

Thomas Mills notes the increasing importance of assessing the legal and evidential value of machine-readable records, rather than their informational value, as the use of computers for carrying out government policies grows.[21]

Robert Warner and Francis Blouin emphasise the earliest possible involvement by archivists in ADP systems, not only to ensure that adequate arrangements are made for the preservation of valuable data, but also to assist the archivist to make informed judgements about the data.[22]

Michael Cook's book, *Archives and the Computer*, published in 1980,[23] is chiefly concerned with the application of automated data processing to archives management. A chapter on machine-readable archives, however, includes a section dealing with appraisal. Cook sets out a number of factors which apply to the appraisal of machine-readable records. We have noted the mention of most of these by the earlier writers. Of the others, Cook recommends the consideration of files for appraisal "in the context of other records (both machine-readable and paper), and any relevant published material, including reports and statistics".[24] Cook's solution to the problem of the continually updated data base, raised by Dollar in 1978, is to "record the position at pre-determined times"[25] that is, by regular sampling.

Another useful feature of Cook's treatment of the subject is the reproduction of a flowchart describing the process of appraisal as practised in the National Archives and Records Service and of the "Guidelines for Selection of Machine-readable and Related Records for Permanent Preservation" as used by the Public Record Office.[26]

Finally, a view of the Soviet experience is provided by I.I. Danilenko and V.V. Tsaplin, of the National Archives of the U.S.S.R., in an article published in *ADPA* in 1984.[27] The article describes the processes and criteria used in the selection of machine-readable archives, under the "Statute for the selection, acquisition for archival storage and the hiring of computer-generated documents", approved by the Archives Administration of the U.S.S.R. Council of Ministers in June 1983.[28]

The criteria described by the authors are already familiar to us, including a preference for unaggregated or source data, the importance of the decisions or problems to which the data relate, the avoidance of duplicated data, the likely use of the data — in turn, based on several further criteria — and the validity of the data.[29]

Of more value is the author's description of a two-stage appraisal process

and of the organisational arrangements established by the statute for appraisal and selection work. The appraisal process involves the periodic listing and description of problems or projects to which sets of data relate, followed by the selection of data within each problem according to the above-mentioned criteria.

The responsibility for selection is placed with an examination commission in each computer centre, which includes archival and user representation, while further approval by a state archive's commission is required prior to transfer of custody.[30]

● ● ● ● ● ● ●

What, then, can we learn from the literature about the ways in which we should develop our programmes of disposal and appraisal for machine-readable records? I will make two points about procedures and look at appraisal criteria in detail.

We find general agreement in the literature about the need to start appraisal early in the life of a machine-readable record, and we have already noted the dangers of re-use, deterioration and loss of support documentation if this is not done.

Essentially, therefore, it is necessary to build disposal arrangements into a computer system, preferably at the design stage. This, in turn, requires the co-operation of system designers, who must be convinced of the need to provide for appraisal and retention of important records and whose expertise will need to be harnessed in order to achieve this objective.

A second point of procedure which appears more than once in the literature is the need to include data verification or validation in the disposal process. Data validation involves assessing the accuracy, reliability and validity of the data. Accuracy can be assessed by comparing manually a file's record layout specifications with a partial printout, which should reveal any inconsistencies and the likelihood of missing data. The reliability and validity of the data should be assessed in order to discover unstated biases in the data or the existence of data imputation, that is, cases where estimates have been substituted for missing responses or incorrect figures. Biases and data imputation may have been harmless or even necessary for the use to which the data was originally put, but may have a considerable effect if the data is put to different uses later in its life.

From the literature, we can identify four main kinds of criteria for the appraisal of machine-readable records. Briefly, these are:

    1. those that differ little from criteria which would be normally applied to conventional records

    2. those which assess the records according to the role which they play in the computer system's processes

3. practical considerations determined by the media with which we are dealing and

4. the value of information in terms of its capacity for re-use and re-processing independent of its original purpose.

Let us examine each of these groups of criteria in turn.

As noted earlier, the National Archives and Records Service asks what is a record's "legal, evidential or informational value"[31] and whether such value is unquestionably permanent, or of immediate or long-term research use.

The Public Archives of Canada asks whether machine-readable records "were or may be used to support the formulation of policy", or "were created for a study which might be considered seminal either because of the nature of the study or the type of analysis used", or "were created for a study conducted by an individual or group of individuals renowned in that field".[32]

The P.R.O. guidelines, which are designed for use in determining detailed appraisal standards for particular departments of the British Government, include:

- Files of data which have been processed to produce reports which have had a bearing on departmental or government policy,

- Files of data produced for the benefit of Royal Commissions and departmental and inter-departmental committees and working parties,

- Files of data recording rights or obligations of or against the Crown, and

- Files of data relating to issues which were the subject of interest or controversy on the national or international plane.[33]

Delete the phrase "of data" and such criteria could be applied to a government's paper records without a second thought.

Our second category of criteria for appraising machine-readable records comprises those which assess the records according to the role which they play in the computer system's processes. The Public Archives of Canada, as noted earlier, requires the retention of the master file, that is, "the file which is the product of all files creating it",[34] while working, transactional and summary files may be discarded. Equally, the Public Record Office has no interest in "scratch tapes, raw data input tapes, working processing tapes, test tapes or interim master tapes".[35] The Public Archives of Canada also calls for the retention of extract files, where it is the extract files that have actually been used.[36] In such a case, the extract file would be a better indication of the use and usefulness of the information than the master file.

Two points need to be made about this category of criteria. Firstly, it appears that concepts such as master and process files and therefore the

appraisal criteria based on them, are of questionable relevance in the on-line computer environment. Charles Dollar sets the scene well:

> Typically, an analyst would select the data elements needed for a particular study and then instruct the computer to create a temporary data set composed of these elements. His subsequent analysis might be displayed on the video screen with results copied by hand or printed out by the computer. Once the analysis is completed, the temporary data set would be erased. The only evidence of the process would be the analyst's brief report ... In this scenario there are no processing files in the usual sense of the term, and the automatic erasure of the data set means there would be no master file.[37]

Apart from periodic sampling of the main data base, there seems to be no present solution to this problem.

Secondly, we should note that there is a contradiction between the aim of keeping master files, that is, the products of processes, and the aim of selecting data, as Bell puts it, "in as close a form as possible to that of its collection".[38]

The third category of appraisal criteria which we identified concerns practical considerations flowing from the kind of media with which we are dealing.

Thus, the existence of adequate support documentation, whether in hard copy or machine-readable form, is described by more than one writer as essential, since, without its documentation, a computer file cannot be interpreted and used. "It would be a disaster to have preserved a file of outstanding historical importance but to have lost its documentation", says Roper. "In such a case one might as well discard the file".[39]

Perhaps more than is the case with conventional records, the costs of preservation must be considered in the appraisal of machine-readable records. A decision to retain implicitly involves a commitment to provide a clean, controlled storage environment, periodic exercising and cleaning in the case of tape, the creation of security or reference copies, the maintenance of equipment to read the record and/or the conversion of the record to a different format, and periodic re-copying.

Charles Dollar dramatically illustrates some of the costs involved by quoting, in 1977 figures: $400 to $600 per reel of tape to cover accessioning and conversion to standard character codes.[40]

A related criterion is the degree of "software dependence" of a particular item. Where use of a file depends on the use of a particular software package, especially one written for the system, this must be considered. If a software dependent file is to be retained, a choice must be made between retaining it in its original format, along with any necessary programme tapes or disks, and converting the file to a standard or software independent format. The feasibility and cost of either of these courses of action will necessarily be a factor in considering whether or not to retain the file.

Finally, let us look at our fourth category of criteria for appraising machine-readable records, which can be described as the independent informational value of the data. As we have already noted, there is a school of thought which argues that it is information as such, regardless of its original or current use, that future users will wish to exploit, and that, therefore, it is the information and the extent to which it can be re-used and re-processed that should be appraised.

It is difficult to do more than generalise about appraisal criteria for this purpose. Roper summarises it as

> those machine-readable records which provide evidence of trends or develop-ments in social and economic history; which contain statistical or financial data over long periods or wide areas, especially where those data are available otherwise only in aggregated form; or which contain information of historical or practical importance relating to aspects of scientific, technical and medical research and development.[41]

In contrast to policies of seeking master and extract files, this school of thought prefers unaggregated data to aggregated data and, as noted earlier, prefers data in as close a form as possible to that of its collection. This approach seeks to ensure that the data can be put to the widest possible range of uses.

In summary, then, it seems clear that the advent of machine-readable records requires some rethinking by archivists of traditional disposal and appraisal practices and standards. The growing body of literature on the subject will help us to do this, as more and more of us come to deal with machine-readable records in our working lives.

FOOTNOTES

1. M.H. Fishbein, "Appraising Information in Machine Language Form", *American Archivist*, Vol. 35, No. 1, January 1972, p. 35-43.

2. Ibid. p. 38.   3. Ibid. p. 40-41.

4. M.E. Carroll, "The Public Archives of Canada's Experience in Establishing a Machine-Readable Archives", in Dr L. Bell and M. Roper, eds., *Proceedings of an International Seminar on Automatic Data Processing in Archives*, Public Record Office, 1975, p. 122-135.

5. Ibid. p. 123.

6. C.M. Dollar, "Appraising Machine-Readable Records", *American Archivist*, Vol. 41, No. 4, October 1978, p. 423-430.

7. Since the publication of Dollar's article, NARS have issued their *General Records Schedule 20 for Machine-Readable Records*. Divided into sections dealing with "master files", "processing files" and documentation, the Schedule adopts a mix of conventional criteria and criteria based on the role played by the file within the particular ADP system.

8. C.M. Dollar, loc. cit., p. 427.   9. Ibid. p. 429.

10. Dr L. Bell, "The Archival Implications of Machine-Readable Records", *Archivum*, XXVI, 1979, p. 85-92.

11. Ibid. p. 91.   12. Ibid.

13. M. Roper, "The Changing Face of the File: Machine-Readable Records and the Archivist", *Archives*, Vol. XIV, No. 63, Spring, 1980.

14. Ibid. p. 145.   15. Ibid. p. 146.   16. Ibid.   17. Ibid. p. 147-148.

18. C.L. Geda, E.W. Austin, and F.X. Blouin, eds., *Archivists and Machine-Readable Records*, Proceedings of a Conference on Archival Management of Machine-Readable Records, Ann Arbor, Michigan, 7-10 February 1979, Society of American Archivists, 1980.

19. Fishbein, Dollar and Cook also contributed papers which touch on disposal and appraisal. Fishbein's main point is the need to assess the validity of survey-type data, in terms of its method of collection. Dollar and Cook largely repeat what they have written elsewhere.

20. A.G. Bogue, "Historical Research and Archival Data", in Geda, Austin and Blouin, eds., op. cit., p. 25.

21. T.E. Mills, "Archival Considerations in the Management of Machine-Readable Records in New York State Government", in Geda, Austin and Blouin, eds., op. cit., p. 104-105.

22. R.M. Warner and F.X. Blouin, "Some Implications of Records in Machine-Readable Form for Traditional Archival Practice", in Geda, Austin and Blouin, eds., op. cit., p. 246.

23. M. Cook, *Archives and the Computer*, Butterworths, 1980.

24. Ibid. p. 110-111.   25. Ibid. p. 112.   26. Ibid. p. 113 and 133.

27. I.I. Danilenko and V.V. Tsaplin, "Archival Storage of Machine-Readable Documents: Soviet Experience", *ADPA*, Vol. 4, No. 3, 1984, p. 33-39.

28. Ibid. p. 34.   29. Ibid. p. 36-37.

30. Ibid. p. 36 and 38.   31. C.M. Dollar, loc. cit., p. 425.

32. M.E. Carroll, loc. cit., p. 131.

33. Public Record Office, *The Selection and Preparation for Transfer of Machine-Readable Records: a Provisional Guide*, Public Record Office, 1975, Appendix B.

34. M.E. Carroll, loc. cit., p. 132.   35. M. Cook, op. cit., p. 110.

36. M.E. Carroll, loc. cit.   37. C.M. Dollar, loc. cit., p. 429.

38. Dr L. Ball, loc. cit., p. 91.   39. M. Roper, loc. cit., p. 147.

40. C.M. Dollar, loc. cit., p. 428.   41.   M. Roper, loc. cit.