

Preserving Machine-Readable Archival Records: A Reply to John Mallinson

by **SUE GAVREL**

Dr. John Mallinson discusses the technical difficulties related to the preservation of EDP records elsewhere in the "Counterpoint" section of this issue of *Archivaria*. He summarizes the deliberations of Subcommittee C of the Committee on Preservation of the National Archives and Records Administration (NARA). Dr. Mallinson raises and examines many of the technical problems related to the preservation of these records but fails to consider how other archival issues, such as appraisal and dissemination, enter into the long-term care of machine-readable archives. He holds the view that there is a misunderstanding of the nature of machine-readable records in the archival community, and that there is a naive hope that some unidentified future technology will solve all machine-readable record problems. This indicates that Dr. Mallinson has had little detailed discussion with archivists responsible for EDP records programmes. The technical difficulties associated with these records have always been of high concern to archivists when undertaking the appraisal and preservation of these records. Archivists face technical difficulties in processing, conserving, and servicing these records; they have tried to develop appropriate procedures to contend with the changes in computer hardware and software. Some records have been and will continue to be lost but many have been preserved and are being used by users for current research. There is no naive belief that some future technology will solve the problem. There are, instead, efforts being made to inform EDP personnel of the fact that data managed by computerized systems are official records containing information related to the operations of the organization and the description of society in the seventies and eighties. I would like to raise the archival considerations related to these records and to explain why choosing a "safe" alternative medium on which to conserve the information they contain, such as microfilm, is not the solution to the problem.

Appraisal of machine-readable records involves many of the difficulties encountered in appraisal of more traditional archival material. The medium permits new research possibilities which must be considered in the selection of the records. Appraisal of EDP records consists of a two-step process involving evaluation of the informational content (content analysis) and analysis of the technical aspects of the records (technical analysis). Determination of the informational value of EDP records is based on the information contained in the records. EDP data are collections of information relating to persons, things, or activities. General points to consider in the determination of informational

value are the uniqueness of the information, the form, and the importance. The uniqueness of the information relates to whether or not the information can be found elsewhere and in what form (machine-readable, paper, microform). The form of the records is important as EDP records can contain micro-data or non-aggregated data or the information can already be summarized or aggregated. EDP records containing micro-data have more research potential than aggregated data.

Inherent in the archival value of EDP records is the fact that these records can be analyzed and re-analyzed in their present form. Three special characteristics of these records must be considered during the appraisal process:

1) *The manipulability of EDP records.* There have been many occasions when records in paper form have been too cumbersome to use for statistical purposes and have been assessed as not having archival value. The same information in machine-readable form, however, may present endless potential for research.

2) *Non-aggregate data have more value than aggregate data.* This is a particular feature of EDP records; in appraisal of paper records the opposite is true. The non-aggregated form provides researchers with the basis from which they can do their own analysis.

3) *Linkage of EDP records.* With more standardization of data elements in EDP records, the possibility exists of linking records in one file to records in other files and this increases the informational value of all the records.

EDP records are not appraised solely on the basis of their contents. They must also be examined in terms of their technical requirements. A number of factors are considered: whether the records can be read; the internal arrangement of the data; the size of the machine-readable file; the frequency of updating the data; and the software and hardware that are used to manipulate the data. The archivist must be aware of how the EDP records being appraised came into existence, of other similar records which exist, and the methods and costs that will be used in properly conserving the data. These factors require an in-depth knowledge of the EDP records and related records. A knowledge of data processing techniques and methods used in the field is necessary for the evaluation of the technical aspects of the records.

The conservation of EDP records does require considerable care and an awareness of changes in technology. Conservation is an active process, requiring the archivist to keep abreast of changes in technology. Tapes cannot be stored on a shelf and forgotten. Complete documentation must be acquired to ensure that the layout of the data on the magnetic tape is defined. Magnetic tapes must be stored in environmentally controlled conditions, handled with care, precision rewind and cleaned on a regular basis, and recopied every eight to ten years. Although the medium is vulnerable, the recopying of data ensures its long-term preservation. Although the value of the records is linked to their manipulability, there is no intrinsic value in the medium. New copies of old tapes are in most cases better than the original. As magnetic tapes are recopied, changes to the physical characteristics of the recording of the data can be made to adapt to current changes in hardware. These changes can be made without affecting the value of the records.

Dr. Mallinson paints a pessimistic portrait of the manufacturers of hardware and software and their lack of standards. Standards do exist for the transfer of data from one system to another and work continues to progress. The use of information created in machine-readable form goes far beyond the principal motive of increasing speed of access.

The more tools technology has provided, the more inventive researchers have become in using the resources. How would researchers use human-readable forms of numeric data such as census records, national surveys, registry systems, and measurement data? Computer output microfilm (COM) is an advanced technology. Computer Input Microfilm (CIM) is far from being sufficiently cost effective to be used by researchers to reconvert data from human readable to computer readable. Optical scanners exist to read the information on microfilm but the error rate is extremely high and requires manual review of all records. The user of the archives is no longer only the genealogist or the historian who researches masses of paper. The users of archives represent a wide range of disciplines. Along with the academic researcher, government researchers are now discovering the resources available to them in archives. New surveys are being designed based on existing data collections that have been conserved in machine-readable form. Such new projects link new data collections to older collections, thus providing for analysis of changes that have occurred. The use of Gallup Poll data which dates from the 1940s to the present provides a valuable source of data to study the trends and changes in opinion over a thirty-year period. In human-readable form (that is, on microfilm), this data could not be used.

Fear of technological change does not present sufficient reason to choose the "safest" medium if that medium cannot document the activities of organizational programmes of the period and society as a whole and provide researchers, both contemporary and future, with access to the wealth of information that exists. Governments, industries, and businesses are highly dependent on the creation and use of computerized records and try to resolve these difficulties by ensuring that the archival concerns are understood and reflected at the time the data are created. It is not that EDP personnel do not care about the need to preserve machine-readable records, but that they are totally unaware of the existence of archives.

In conclusion, although a number of recommendations proposed by the Committee on Preservation were accepted by the National Archives and Records Administration, the recommendation by Subcommittee C pertaining to preservation of machine-readable records, which is advanced in *Archivaria* by Dr. Mallinson, was not accepted by NARA.